МИНИСТЕРСТВО СЕЛЬСКОГО ХОЗЯЙСТВА РОССИЙСКОЙ ФЕДЕРАЦИИ
ФГБОУ ВО «Кубанский государственный аграрный университет имени И. Т. Трубилина»

А. И. Орлов, Е. В. Луценко

АНАЛИЗ ДАННЫХ, ИНФОРМАЦИИ И ЗНАНИЙ В СИСТЕМНОЙ НЕЧЕТКОЙ ИНТЕРВАЛЬНОЙ МАТЕМАТИКЕ

Монография

Краснодар
КубГАУ
2022

УДК 004.8 (075.8)
ББК 32.965
Л86

Р е ц е н з е н т ы :

В. В. Степанов – профессор кафедры информатики и вычислительной техники Кубанского государственного технологического университета, д-р техн. наук, профессор;

Г. А. Аршинов – профессор кафедры компьютерных технологий и систем Кубанского государственного аграрного университета, д-р техн. наук, канд. физ.-мат. наук, профессор

Орлов А.И., Луценко Е.В.

О66 Анализ данных, информации и знаний в системной нечеткой интервальной математике: научная монография / А. И. Орлов, Е. В. Луценко. – Краснодар: КубГАУ, 2022. – 402 с.

ISBN

Монография включает две части. В 1-й части в 10 главах рассматриваются теоретические основы системной нечеткой интервальной математики. Во 2-й части из 5 глав рассматриваются соотношение смыслового содержания понятий «данные», «информация» и «знания», а также и теоретические и математические основы базового, сценарного, спектрального и текстового автоматизированного системно-когнитивного анализа (АСК-анализ). АСК-анализ является одним из вариантов практической реализации системной нечеткой интервальной математики. Приводятся детальные численные примеры применения сценарного и спектрального АСК-анализа для прогнозирования на финансовых рынках и анализа изображений.

Предназначена для обучающихся бакалавриата, магистратуры и аспирантуры, а также преподавателей, исследователей и разработчиков в области высоких статистических технологий и искусственного интеллекта, для всех интересующихся данной проблематикой.

ISBN
DOI:

ОГЛАВЛЕНИЕ

СОДЕРЖАНИЕ

ПРЕДИСЛОВИЕ

В 2014 г. вышла наша книга "Системная нечеткая интервальная математика"[1]. Название было выработано в процессе подготовки этой монографии. Так мы назвали центральное направление наших исследований. В настоящую книгу мы включили основные полученные после 2014 г. научные результаты по методам анализа данных, информации и знаний в системной нечеткой интервальной математике.

Научной общественности была представлена новая парадигма математических методов исследования. Речь шла о новой парадигме прикладной статистики, эконометрики, математической статистики, математических методов экономики, организационно-экономического моделирования в экономике и управления. Считаем необходимым при разработке организационно-экономического, математического и программного обеспечения для решения задач конкретной прикладной области, например, ракетно-космической отрасли, исходить из новой парадигмы математических методов исследования. Аналогичное требование предъявляем к преподаванию соответствующих дисциплин - при разработке учебных планов и рабочих программ необходимо исходить из новой парадигмы математических методов исследования. В главе 1 мы приводим базовую информацию о новой парадигме. Изложение посвящено в основном научной области (специальности) «Математические и инструментальные методы экономики», включающей организационно-экономическое и экономико-математическое моделирование, эконометрику и статистику, а также теорию принятия решений, системный анализ, кибернетику, исследование операций. Обсуждаем основные понятия. Рассказываем о ходе разработки новой парадигмы. Проводим развернутое сравнение старой и новой парадигм математических методов исследования. Даем информацию об учебной литературе, подготовленной в соответствии с новой парадигмой математических методов исследования.

Системная нечеткая интервальная математика тесно переплетена с статистикой нечисловых данных, выделенной как самостоятельная область прикладной статистики в 1979 г.. Первоначально для обозначения этой области математических методов экономики использовался термин "статистика объектов нечисловой природы". Наш базовый учебник называется "Нечисловая статистика". Статистика нечисловых данных - одна из четырех основных областей прикладной статистики (наряду со статистикой чисел, многомерным статистическим анализом, статистикой временных рядов и случайных процессов). Она делится на статистику в

[1] Орлов А.И., Луценко Е.В. Системная нечеткая интервальная математика. Монография (научное издание). – Краснодар, КубГАУ. 2014. – 600 с.

пространствах общей природы и разделы, посвященные конкретным типам нечисловых данных (статистика интервальных данных, статистика нечетких множеств, статистика бинарных отношений и др.). В настоящее время статистика в пространствах общей природы - центральная часть прикладной статистики, а включающая ее статистика нечисловых данных - основная область прикладной статистики. Это утверждение подтверждается, в частности, анализом публикаций в разделе "Математические методы исследования" журнала "Заводская лаборатория. Диагностика материалов" - основном месте публикаций отечественных исследований по прикладной статистике. Глава 2 посвящена анализу основных идей статистики нечисловых данных на фоне развития прикладной статистики с позиций новой парадигмы математических методов исследования. Описаны различные виды нечисловых данных. Проанализирован исторический путь статистической науки. Рассказано о развитии статистики нечисловых данных. Разобраны основные идеи статистики в пространствах общей природы: средние величины, законы больших чисел, экстремальные статистические задачи, непараметрические оценки плотности распределения вероятностей, методы классификации (диагностики и кластер-анализа), статистики интегрального типа. Кратко рассмотрены некоторые статистические методы анализа данных, лежащих в конкретных пространствах нечисловой природы: непараметрическая статистика (реальные распределения обычно существенно отличаются от нормальных), статистика нечетких множеств, теория экспертных оценок (медиана Кемени - это выборочное среднее экспертных упорядочений) и др. Обсуждаются некоторые нерешенные задачи статистики нечисловых данных

Непараметрические оценки плотности распределения вероятностей в пространствах произвольной природы - один из основных инструментов нечисловой статистики. В главе 3 рассмотрены их частные случаи – ядерные оценки плотности в пространствах произвольной природы, гистограммные оценки и оценки типа Фикс-Ходжеса. Цель главы 3 - завершение цикла наших работ, посвященного математическому изучению асимптотических свойств различных видов непараметрических оценок плотности распределения вероятности в пространствах общей природы. Тем самым подводится математический фундамент под применения таких оценок в нечисловой статистике. Начинаем с рассмотрения среднего квадрата ошибки ядерной оценки плотности и - с целью максимизации порядка его убывания - выбор ядерной функции и последовательности показателей размытости. Основные введенные нами понятия - круговая функция распределения и круговая плотность. Порядок сходимости в общем случае тот же, что и при оценивании плотности числовой случайной величины, но основные условия наложены не на плотность случайной величины, а на круговую плотность. Далее рассматриваем

другие виды непараметрических оценок плотности - гистограммные оценки и оценки типа Фикс-Ходжеса. Затем изучаем непараметрические оценки регрессии и их применение для решения задач дискриминантного анализа в пространствах общей природы

В главе 4 рассмотрены основные идеи асимптотической математической статистики интервальных данных, в которой элементы выборки – не числа, а интервалы. Алгоритмы и выводы статистики интервальных данных принципиально отличаются от классических. Приведены результаты, связанные с основополагающими понятиями нотны и рационального объема выборки. Статистика интервальных данных является составной частью системной нечеткой интервальной математики.

Изучаемые в главе 5 коэффициенты корреляции и детерминации широко используются при статистическом анализе данных в рамках системной нечеткой интервальной математики. Согласно теории измерений линейный парный коэффициент корреляции Пирсона применим к переменным, измеренным в шкале интервалов. Его нельзя использовать при анализе порядковых данных. Непараметрические ранговые коэффициенты Спирмена и Кендалла оценивают связь порядковых переменных, Важно, что при проверке значимости отличия коэффициента корреляции от 0 критическое значение зависит от объема выборки. Поэтому использование т.н. "шкалы Чеддока" некорректно. При применении пассивного эксперимента коэффициенты корреляции можно обоснованно использовать для прогнозирования, но не для управления. Для получения предназначенных для управления вероятностно-статистических моделей необходим активный эксперимент. Влияние выбросов на коэффициент корреляции Пирсона весьма велико. При увеличении числа проанализированных наборов предикторов заметно растет максимальный из соответствующих коэффициентов корреляции - показателей качества приближения (эффект «вздувания» коэффициента корреляции). Рассмотрены основные модели регрессионного анализа. Выделены модели метода наименьших квадратов с детерминированной независимой переменной. Рассматриваем произвольное распределение отклонений, при этом для получения предельных распределений оценок параметров и регрессионной зависимости предполагаем выполнение условий центральной предельной теоремы. Второй тип моделей основан на выборке случайных векторов. Зависимость является непараметрической, распределение двумерного вектора - произвольным. Об оценке дисперсии независимой переменной можно говорить только в модели на основе выборки случайных векторов, равно как и о коэффициенте детерминации как критерии качества модели. Обсуждается сглаживание временных рядов. Рассмотрены методы восстановления зависимостей в пространствах общей природы. Показано, что предельное распределение естественной

оценки размерности модели является геометрическим, а построение информативного подмножества признаков наталкивается на эффект "вздувания коэффициентов корреляции". Обсуждаются различные подходы к регрессионному анализ интервальных данных. Анализ многообразия моделей регрессионного анализа приводит к выводу, что не существует единой "стандартной модели"

Вероятностно-статистические модели данных - основа методов прикладной статистики. При анализе статистических данных часто необходимо оценивать две составляющие вероятностно-статистических моделей - структуру моделей и их параметры. Методы расчета состоятельных оценок параметров хорошо известны (например, применяют методы одношаговых оценок, которые пришли на смену методам максимального правдоподобия). Структура модели обычно выбирается исследователем (можно сказать, что используются экспертные методы). Некоторые параметры структуры можно оценивать с помощью математико-статистических методов. Например, степень многочлена в регрессионной зависимости или число слагаемых в модели смеси распределений, используемой для классификации. Для подобных параметров модели используется общий термин - размерность вероятностно-статистической модели. Более общая составляющая модели - информативное подмножество признаков. В главе 6 рассмотрено асимптотическое поведение оценок размерностей ряда моделей. Изучено асимптотическое поведение ряда оценок степени полинома при восстановлении зависимости. Получены состоятельные оценки размерности и структуры модели в регрессии. Рассмотрены подходы к оцениванию числа элементов смеси в задачах классификации. Обсуждаются оценки размерности модели в факторном анализе и многомерном шкалировании. С целью обоснования последовательного выполнения этапов статистического анализа данных анализируются проблемы "стыковки" алгоритмов классификации и регрессии. Полезными оказываются оптимизационные формулировки ряда задач прикладной статистики. Основные результаты касаются состоятельности оценок. Краткие формулировки ряда теорем содержатся в ранее вышедших публикациях. Проблема оценивания размерности вероятностно-статистической модели как самостоятельное направление прикладной статистики впервые рассмотрена здесь. Впервые публикуются доказательства включенных в настоящую главу теорем. Эти теоремы и подробные доказательства и являются основными научными результатами работы

Назрела необходимость навести порядок в математических методах классификации. Это повысит их роль в решении прикладных задач, в частности, при диагностике материалов. Прежде всего следует выработать требования, которым должны удовлетворять методы классификации.

Первоначальная формулировка таких требований - основное содержание главы 7. Математические методы классификации мы рассматриваем как часть методов прикладной статистики. Обсуждаем естественные требования к рассматриваемым методам анализа данных и представлению результатов расчетов, вытекающие из накопленных отечественной вероятностно-статистической научной школой достижений и идей. Даются конкретные рекомендации по ряду вопросов, а также критика отдельных ошибок, встречающихся в научных публикациях. В частности, методы анализа данных должны быть инвариантны относительно допустимых преобразований шкал, в которых измерены данные, т.е. методы должны быть адекватны в смысле теории измерений. Основой конкретного статистического метода анализа данных всегда является та или иная вероятностная модель. Она должна быть явно описана, ее предпосылки обоснованы - либо из теоретических соображений, либо экспериментально. Методы обработки данных, предназначенные для использования в реальных задачах, должны быть исследованы на устойчивость относительно допустимых отклонений исходных данных и предпосылок модели. Должна указываться точность решений, даваемых с помощью используемого метода. При публикации результатов статистического анализа реальных данных необходимо указывать их точность (приводить доверительные интервалы). В качестве оценки прогностической силы алгоритма классификации вместо доли правильных прогнозов рекомендуется использовать прогностическую силу. Математические методы исследования делятся на "разведочный анализ" и "доказательную статистику". Специфические требования к методам обработки данных возникают в связи с их "стыковкой" при последовательном выполнении. Обсуждаются границы применимости вероятностно-статистических методов. Рассматриваются также конкретные постановки задач классификации и типовые ошибки при применении различных методов их решения.

К инструментальным методам экономики относится метод Монте-Карло (синоним - метод статистических испытаний). Он широко используется при разработке, изучении и применении математических методов исследования в эконометрике, прикладной статистике, организационно-экономическом моделировании, при разработке и принятии управленческих решений, является основой имитационного моделирования. Разработанная нами новая парадигма математических методов исследования (см. главу 1) опирается на применение метода Монте-Карло. В математической статистике для многих метолов анализа данных получены предельные теоремы об асимптотическом поведении рассматриваемых величин при безграничном росте объемов выборок. Следующий шаг - изучение свойств этих величин при конечных объемах выборок. Для такого изучения с успехом применяют метод Монте-Карло.

В главе 8 этот метод применяем для изучения свойств статистических критериев проверки однородности двух независимых выборок. Рассмотрены наиболее используемые при анализе реальных данных критерии - Крамера-Уэлча (совпадающий при равенстве объемов выборок с критерием Стьюдента); Лорда, Вилкоксона (Манна-Уитни), Вольфовица, Ван-дер-Вардена, Смирнова, типа омега-квадрат (Лемана-Розенблатта). Метод Монте-Карло позволяет оценить скорости сходимости распределений статистик критериев к пределам, сравнить свойства критериев при конечных объемах выборок. Для применения метода Монте-Карло необходимо выбрать функции распределения элементов двух выборок. Для этого в главе 8 использованы нормальные распределения и распределения Вейбулла - Гнеденко. Получена рекомендация: для проверки гипотезы совпадения функций распределения двух выборок целесообразно использовать критерий Лемана - Розенблатта типа омега-квадрат. Если есть основания предполагать, что распределения отличаются в основном сдвигом, то можно использовать также критерии Вилкоксона и Ван-дер-Вардена. Однако даже в этом случае критерий типа омега-квадрат может оказаться более мощным. В общем случае, кроме критерия Лемана - Розенблатта, допустимо применение критерия Смирнова, хотя для этого критерия реальный уровень значимости может значительно отличаться от номинального. Оценены частоты расхождений статистических выводов по разным критериям.

В современных условиях эконометрика как научная, практическая и учебная дисциплина становится всё более востребованной. Современная эконометрика - неотъемлемая составляющая научного обеспечения искусственного интеллекта и цифровой экономики. Методы эконометрики составляют значительную часть инструментов контроллинга. При ее преподавании весьма важно преодолеть оковы устаревших взглядов XX в., излагая современную эконометрику. Полезным является опыт двадцатилетней реализации авторской программы по эконометрике на факультете "Инженерный бизнес и менеджмент" МГТУ им. Н.Э. Баумана. Основные составляющие современной эконометрики представлены в разработанном нами учебном курсе, которому и посвящена глава 9. В ядро современной эконометрики включаем следующие базовые разделы: выборочные исследования; метод наименьших квадратов; эконометрический анализ инфляции; методы экспертных оценок; теория измерений и средние величины; введение в теорию риска; основы статистики нечисловых данных; непосредственный анализ статистических данных; статистический контроль; эконометрический анализ связанных выборок; основы теории нечетких множеств; элементы статистики интервальных данных; основы теории классификации; элементы теории рейтингов; эконометрика как научная дисциплина. Приведен перечень контрольных работ и формулировки домашних заданий. Обширный

список литературных источников показывает, что авторский курс эконометрики в соответствии с принципом "образование - через науку" основан на недавних научных исследованиях, многие из которых опубликованы в "Научном журнале КубГАУ". Представленный в главе 9 курс разработан в соответствии с положениями отечественной научной школы в области эконометрики на основе современной парадигмы организационно-экономического моделирования, эконометрики и статистики. Основные составляющие современной эконометрики представлены в разработанном нами учебном курсе. Целесообразно именно его преподавать во многих университетах и вузах другого профиля, оставив в прошлом устаревшие учебники, в которых .из всех базовых тем современной эконометрики рассматривается лишь одна - метод наименьших квадратов.

Как показано в главе 10, системная нечеткая интервальная математика - основа математики XXI в. Определения математики как науки менялись со временем. В XIX в. ее определяли как науку о числах и фигурах (телах). В XXI в. математика - наука о формальных структурах. Следовательно, ее нельзя относить к естественным наукам. Математика изучает мысленные конструкции. В практике математических исследований аксиоматические теории - это, как правило, недостижимый идеал. Есть два направления деятельности математиков. Исследования в первом из них нацелены на построение и изучение моделей реальности, на получение научных результатов, которые - прямо или опосредованно - позволяют решать практические задачи. Представители второго направления занимаются решением конкретных трудных задач. Примеры - "великая теорема Ферма", задача пяти красок и т.п. Именно они готовят новых математиков, руководят профессиональными объединениями. В результате первое направление оказывается ущемленным. С точки зрения представителей первого направления наиболее важные области математики - это математический анализ, алгебра (линейная, высшая и др.) и геометрия (многомерная, начертательная, топология и др.). Для решения прикладных задач в XX в. наиболее важными оказались теория вероятностей и математическая статистика, теория оптимизации, дифференциальные и разностные уравнения. Начиная со второй половины XX в. появились новые области математики - статистика нечисловых данных, теория нечетких множеств, автоматизированный системно-когнитивный анализ, интервальная математика. Объединяющую их системную нечеткую интервальную математику рассматриваем в главе 10 как основу математики XXI века. Основная часть областей математики, разработанных представителями второго направления, в применении к решению прикладных задач оказалась, увы, бесплодной. Необходимо различать математические, прагматические и компьютерные числа. Разработан ряд подходов к моделированию связей математических и

прагматических чисел - на основе группировки, интервального анализа, нечетких множеств, автоматизированного системно-когнитивного анализа. В конце главы 10 кратко рассказано о многообразии литературных источников по тематике этой главы.

В 2014 г. вышла монография авторов "Системная нечеткая интервальная математика"[2]. Во 2-ю часть настоящей монографии вошли основные полученные после 2014 года результаты развития автоматизированного системно-когнитивного анализа (АСК-анализ) и его программного инструментария – интеллектуальной системы «Эйдос». Это развитие касается, прежде всего, сценарного и спектрального АСК-анализа, а также еприменения АСК-анализа для интелекутального анализа текстсов.

Авторы считают, что АСК-анализ является одним из вариантов практической реализации системной нечеткой интервальной математики.

Во 2-й части рассматриваются соотношение смыслового содержания понятий «данные», «информация» и «знания», а также и теоретические и математические основы базового, сценарного, спектрального и текстового автоматизированного системно-когнитивного анализа (АСК-анализ).

Приводятся детальные численные примеры применения сценарного и спектрального АСК-анализа для прогнозирования на финансовых рынках и анализа изображений.

Сценарный АСК-анализ развит на основе одного предложенного автором частного случая теоремы А.Н. Колмогорова (1957). По своей сути замечательная теорема А.Н. Колмогорова (1957) (точнее этот ее частный случай), является теоретической основой всей математической теории разложения функций в ряды, т.е. так называемой теории рядов.

В математике разработано много различных конкретных вариантов разложений функций в ряды.

Однако, к сожалению, определение вида базисных функций и весовых коэффициентов для данной конкретной функции представляет собой математическую проблему, для которой пока не найдено общего математически строго решения.

При этом для частных случаев, т.е. конкретных видов базисных функций, таких решений найдено довольно много.

В данной работе предлагается рассматривать математическую модель АСК-анализа как вариант общего и универсального практического решения проблемы разработки базисных функций и весовых коэффициентов для разложения в ряд по ним произвольной функции

[2] Орлов А.И., Луценко Е.В. Системная нечеткая интервальная математика. Монография (научное издание). – Краснодар, КубГАУ. 2014. – 600 с.

состояния идентифицируемого объекта. Прослеживается сопоставление смысла понятий АСК-анализа и теоремы А.Н.Колмогорова.

Приводятся численные примеры технического, фундаментального и техно-фундаментального сценарного АСК-анализа.

В этих численных примерах на основе анализа ретроспективных исходных данных выявляются фактически наблюдавшиеся прошлые и будущие сценарии развития событий.

Путем их обобщения формируются образы будущих сценариев развития событий, которые рассматриваются как базисные функции классов.

Будущие сценарии обуславливаются прошлыми сценариями развития событий (значениями факторов).

При прогнозировании текущая ситуация сравнивается с этими обобщенными образами и разлагается в ряд по ним (прямое преобразование, объектный анализ).

Средневзвешенный прогноз формируется путем обратного преобразования образов классов с их весами, т.е. как их взвешенная суперпозиция.

При этом в качестве базисных функций используются обобщенные образы прогнозируемых сценариев того что будет и того что не будет с их весами, в качестве которых используется достоверность прогноза

Автоматизированный системно-когнитивный анализ (АСК-анализ) изображений обеспечивает автоматическое выявление признаков конкретных изображений из цветов пикселей и контуров изображений, синтез обобщенных образов изображений (классов), выявление наиболее характерных и нехарактерных для классов признаков изображений, определение ценности признаков изображений для их различения, удаление из модели малоценных признаков (абстрагирование), решение задач количественного сравнения конкретных изображений с обобщенными образами классов и обобщенных образов классов друг с другом, а также задачи исследования моделируемой предметной области путем исследования ее модели.

В данной работе рассматриваются новые возможности АСК-анализа и реализующей его интеллектуальной системы «Эйдос», обеспечивающие выявление признаков изображений путем их ***спектрального анализа,*** формирования обобщенных спектров классов, решение задач сравнения изображений конкретных объектов с классами и классов друг с другом по их спектрам.

Впервые стало возможным формировать обобщенные спектры классов с весами цветов по степени их характерности и нехарактерности для классов, причем это не интенсивность цвета в спектре, а количество информации в цвете о принадлежности объекта с этим цветом к данному классу.

По сути, речь идет об обобщении спектрального анализа путем применении интеллектуальных когнитивных технологий и теории информации в спектральном анализе.

Во-первых, все говорят о том, что в спектральных линиях содержится информация о том, какой элемент или вещество входят в состав объекта, но никто не удосужился посчитать какое же это конкретно количество этой информации, а затем использовать его для определения состава объекта методы распознавания образов, основанные на использовании этой информации.

Во-вторых, спектральный анализ традиционно используется для определения элементарного и молекулярного состава объекта, а мы предлагаем использовать его не только для этого, но и для идентификации любых изображений. Приводится численный пример.

Применение АСК-анализа для интеллектуального ***анализа текстов*** позволяет решать следующие задачи:

- формировать обобщенные лингвистические образы классов (семантические ядра) на основе фрагментов или примеров относящихся к ним текстов на любом языке;'
- количественно сравнивать лингвистический образ конкретного человека, или описание объекта, процесса с обобщенными лингвистическими образами групп (классов);
- сравнивать обобщенные лингвистические образы классов друг с другом и создавать их кластеры и конструкты;
- исследовать моделируемую предметную область путем исследования ее лингвистической системно-когнитивной модели;
- проводить интеллектуальную атрибуцию текстов, т.е. определять вероятное авторство анонимных и псевдонимных текстов, датировку, жанр и смысловую направленность содержания текстов;
- все это можно делать для любого естественного или искусственного языка или системы кодирования.

Ссылки на работы автора по теккстовому АСК-анализу размещены здесь: http://lc.kubagro.ru/aidos/Works_on_ASK-analysis_of_texts.htm.

Авторы:
Орлов Александр Иванович, профессор,доктор экономических наук,доктор технических наук, кандидат физико-математических наук,
https://orlovs.pp.ru/

Луценко Евгений Вениаминович, профессор, доктор экономических наук, кандидат технических наук,
http://lc.kubagro.ru/
https://www.researchgate.net/profile/Eugene-Lutsenko

ЧАСТЬ 1-Я. ТЕОРЕТИЧЕСКИЕ ОСНОВЫ СИСТЕМНОЙ НЕЧЕТКОЙ ИНТЕРВАЛЬНОЙ МАТЕМАТИКИ

ГЛАВА 1. О НОВОЙ ПАРАДИГМЕ МАТЕМАТИЧЕСКИХ МЕТОДОВ ИССЛЕДОВАНИЯ

В 2011 - 2021 гг. в серии статей в научных журналах и докладов на международных, зарубежных и всероссийских научных конференциях была представлена научной общественности новая парадигма математических методов исследования [1] в области организационно-экономического моделирования, эконометрики и статистики [2 - 5]. Речь шла о новой парадигме прикладной статистики [6, 7], математической статистики [8, 9], математических методов экономики [10], анализа статистических и экспертных данных в задачах экономики и управления [11, 12]. Новая парадигма основана на системной нечеткой интервальной математике.

Считаем необходимым при разработке организационно-экономического обеспечения для решения задач конкретной прикладной области, например, ракетно-космической отрасли, исходить из новой парадигмы математических методов исследования. Аналогичное требование предъявляем к преподаванию соответствующих дисциплин. При разработке учебных планов и рабочих программ необходимо исходить из новой парадигмы математических методов исследования.

В настоящей главе приведем базовую информацию о новой парадигме математических методов исследования.

1.1. Краткая формулировка новой парадигмы

Математические методы исследования используются для решения практических задач с давних времен. В Ветхом Завете рассказано о весьма квалифицированно проведенной переписи военнообязанных (Четвертая книга Моисеева "Числа"). В первой половине XX в. была разработана классическая парадигма методов обработки данных, полученных в результате измерений (наблюдений, испытаний, анализов, опытов). Математические методы исследования, соответствующие классической парадигме, широко используются. Со стороны может показаться, что в этой области основное давно сделано, современные работы направлены на мелкие усовершенствования. Однако это совсем не так. Новая парадигма математических методов исследования принципиально меняет прежние представления. Она зародилась в 1980-х гг., но была развита в серии наших монографий и учебников уже в XXI в.

Типовые исходные данные в новой парадигме – объекты нечисловой природы (элементы нелинейных пространств, которые нельзя складывать и

умножать на число, например, множества, бинарные отношения), а в старой – числа, конечномерные векторы, функции. Ранее (в классической старой парадигме) для расчетов использовались разнообразные суммы, однако объекты нечисловой природы нельзя складывать, поэтому в новой парадигме применяется другой математический аппарат, основанный на расстояниях между объектами нечисловой природы и решении задач оптимизации.

Изменились постановки задач анализа данных. Старая парадигма исходит из идей начала XX в., когда К. Пирсон предложил четырехпараметрическое семейство распределений для описания распределений реальных данных. В это семейство как частные случаи входят, в частности, подсемейства нормальных, экспоненциальных, Вейбулла-Гнеденко, гамма-распределений. Сразу было ясно, что распределения реальных данных, как правило, не входят в семейство распределений Пирсона (об этом говорил, например, академик С.Н. Бернштейн в 1927 г. в докладе на Всероссийском съезде математиков). Однако математическая теория параметрических семейств распределений (методы оценивание параметров и проверки гипотез) оказалась достаточно интересной, и именно на ней до сих пор основано преподавание во многих вузах. Итак, в старой парадигме основной подход к описанию данных - распределения из параметрических семейств, а оцениваемые величины – их параметры, в новой парадигме рассматривают произвольные распределения, а оценивают - характеристики и плотности распределений, зависимости, правила диагностики и др. Центральная часть теории – уже не статистика числовых случайных величин, а статистика в пространствах произвольной природы.

В старой парадигме источники постановок новых задач - традиции, сформировавшиеся к середине XX века, а в новой - современные потребности математического моделирования и анализа данных (XXI век), т.е. запросы практики. Конкретизируем это общее различие. В старой парадигме типовые результаты - предельные теоремы, в новой - рекомендации для конкретных значений параметров, в частности, объемов выборок. Изменилась роль информационных технологий – ранее они использовались в основном для расчета таблиц (в частности, информатика находилась вне математической статистики), теперь же они - инструменты получения выводов (имитационное моделирование, датчики псевдослучайных чисел, методы размножение выборок, в т.ч. бутстреп, и др.). Вид постановок задач приблизился к потребностям практики – при анализе данных от отдельных задач оценивания и проверки гипотез перешли к статистическим технологиям (технологическим процессам анализа данных). Выявилась важность проблемы «стыковки алгоритмов» - влияния выполнения предыдущих алгоритмов в технологической цепочке

на условия применимости последующих алгоритмов. В старой парадигме эта проблема не рассматривалась, для новой – весьма важна.

Если в старой парадигме вопросы методологии моделирования практически не обсуждались, достаточными признавались схемы начала XX в., то в новой парадигме роль методологии (учения об организации деятельности) является основополагающей. Резко повысилась роль моделирования – от отдельных систем аксиом произошел переход к системам моделей. Сама возможность применения вероятностного подхода теперь – не «наличие повторяющегося комплекса условий» (реликт физического определения вероятности, использовавшегося до аксиоматизации теории вероятностей А.Н. Колмогоровым в 1930-х гг.), а наличие обоснованной вероятностно-статистической модели. Если раньше данные считались полностью известными, то для новой парадигмы характерен учет свойств данных, в частности, интервальных и нечетких. Изменилось отношение к вопросам устойчивости выводов – в старой парадигме практически отсутствовал интерес к этой тематике, в новой разработана развитая теория устойчивости (робастности) выводов по отношению к допустимым отклонениям исходных данных и предпосылок моделей.

1.2. Новая парадигма в области математических и инструментальных методов экономики

Изложение в этой главе посвящено в основном научной области «Математические и инструментальные методы экономики», включающей организационно-экономическое и экономико-математическое моделирование, эконометрику и статистику, а также теорию принятия решений, системный анализ, кибернетику, исследование операций. Рассмотрим основное содержание новой парадигмы этой научно-практической области, разработанной в 80-х гг. в процессе создания Всесоюзной статистической ассоциации. Новая парадигма сопоставляем со старой (соответствующей середине XX века). Дадим сводку монографий, учебников и учебных пособий, подготовленных в XXI в. в соответствии с новой парадигмой.

Математические и инструментальные методы экономики – одна из специальностей научных работников, относящаяся к экономическим наукам. Она посвящена разработке интеллектуальных инструментов для решения задач теории и практики экономического анализа.

Так, конкретные модели и методы экономики предприятия и организации производства основаны, в частности, на научных результатах таких научных областей, как организационно-экономическое и экономико-математическое моделирование, эконометрика и статистика. Эти научные области относятся к математическим методам экономики. Они предоставляют интеллектуальные инструменты для решения различных

задач стратегического планирования и развития предприятий, организации производства и управления хозяйствующими субъектами, конструкторской и технологической подготовки производства. В монографии [13] на с. 395-424 выделено 195 групп задач управления промышленными предприятиями и для них указаны базовые группы экономико-математических методов и моделей.

Развитие математических методов экономики привело к формированию новой парадигмы в этой области, существенно отличающейся от послевоенной парадигмы, созданной в 1950-1970 гг. и используемой многими преподавателями и научными работниками и в настоящее время. Настоящая глава посвящена основным идеям новой парадигмы математических методов экономики.

1.3. Основные понятия

Целесообразно начать с определений используемых понятий.

Термин «***парадигма***» происходит от греческого «paradeigma» — пример, образец и означает совокупность явных и неявных (и часто не осознаваемых) предпосылок, определяющих научные исследования и признанных на определенном этапе развития науки [14].

Организационно-экономическое моделирование – научная, практическая и учебная дисциплина, посвященная разработке, изучению и применению математических и статистических методов и моделей в экономике и управлении народным хозяйством, прежде всего промышленными предприятиями и их объединениями [15].

Экономико-математическое моделирование — описание экономических процессов и явлений в виде экономико-математических моделей. При этом экономико-математическая модель — математическое описание экономического процесса или объекта, произведенное в целях их исследования и управления ими: математическая запись решаемой экономической задачи (поэтому часто термины «модель» и «задача» употребляются как синонимы). В самой общей форме модель — условный образ объекта исследования, сконструированный для упрощения этого исследования. При построении модели предполагается, что ее непосредственное изучение дает новые знания о моделируемом объекте, которые позволят разработать и обосновать адекватные управленческие воздействия [16].

Эконометрика – это наука, изучающая конкретные количественные и качественные взаимосвязи экономических объектов и процессов с помощью математических и статистических методов и моделей [17]. Обычно используют несколько более узкое определение: ***эконометрика*** – это статистические методы в экономике [18].

Статистика исходит прежде всего из опыта; недаром ее зачастую определяют как науку об общих способах обработки результатов

эксперимента [19]. ***Прикладная статистика*** – это наука о том, как обрабатывать данные [20].

Специалисту очевидна близость, переплетение, зачастую совпадение всех научных, практических и учебных дисциплин, рассмотренных выше. К ним можно прибавить еще несколько: теорию принятия решений, системный анализ, кибернетику, исследование операций… Исходя из нашего профессионального опыта, попытки искусственно ввести границы между этими дисциплинами не являются плодотворными, хотя и позволяют организовать долгие дискуссии.

На международной научной конференции по организации производства "Вторые Чарновские чтения" [21] работала секция «Организационно-экономическое и экономико-математическое моделирование, эконометрика и статистика». Это название было получено путем объединения названий учебных дисциплин «Организационно-экономическое моделирование», «Эконометрика», «Прикладная статистика», «Статистика», которые изучаются студентами Научно-учебного комплекса «Инженерный бизнес и менеджмент», а также названия Лаборатории экономико-математических методов в контроллинге Научно-образовательного центра «Контроллинг и управленческие инновации» Московского государственного технического университета им. Н.Э. Баумана. На заседании секции была проведена дискуссия по выбору наиболее адекватного названия научной области, к которой относились представленные работы. Приведенное выше название признано слишком длинным. Название «Организационно-математическое моделирование» отклонено как малоизвестное и сужающее рассматриваемую тематику. Одобрено название «Математическое моделирование в организации производства», а при проведении конференций по более широкой тематике – «Математическое моделирование экономики и управления». Заметная доля исследований в этой области относятся к научной специальности «Математические и инструментальные методы экономики», практически все используют те или иные математические методы экономики.

1.4. Разработка новой парадигмы

Организационно-экономическое и экономико-математическое моделирование, эконометрика и статистика предоставляют интеллектуальные инструменты для решения различных задач организации производства и управления предприятиями и организациями. Например, в учебнике по организации и планированию машиностроительного производства (производственному менеджменту) [22] более 20 раз используются эконометрические (если угодно, математические и статистические) методы и модели, как это подробно продемонстрировано, например, в [23].

Рассматриваемые методы широко используются для решения различных задач теории и практики экономического анализа. В частности, проводится когнитивное моделирование [24] развития наукоемкой промышленности (на примере оборонно-промышленного комплекса) и систем налогообложения [25, 26], модельное обоснование инновационного развития наукоемкого сектора российской экономики [27]. Моделируют организационные изменения [28], применяют информационные технологии [29]. Все шире используются экспертные оценки [30 - 32], в том числе для построения обобщенных показателей (рейтингов) [33 - 41].

Во второй половине 1980-х гг. в нашей стране развернулось общественное движение по созданию профессионального объединения специалистов в области организационно-экономического и экономико-математического моделирования, эконометрики и статистики (кратко – статистиков). Аналоги такого объединения - британское Королевское статистическое общество (основано в 1834 г.) и Американская статистическая ассоциация (создана в 1839 г.). К сожалению, деятельность учрежденной в 1990 г. Всесоюзной статистической ассоциации (ВСА) [42] оказалась парализованной в результате развала СССР.

В ходе организации ВСА проанализировано состояние и перспективы развития рассматриваемой области научно-прикладных исследований и осознаны основы уже сложившейся к концу 1980-х гг. ***новой парадигмы организационно-экономического моделирования, эконометрики и статистики.***

В течение следующих лет новая парадигма развивалась и к настоящему времени оформлена в виде серии монографий и учебников для вузов, состоящей более чем из 10 книг.

1.5. Сравнение старой и новой парадигм

Проведем развернутое сравнение старой и новой парадигм математических методов исследования. При этом опираемся на материалы раздела "Математические методы исследования" научно-технического журнала "Заводская лаборатория. Диагностика материалов". С момента основания раздела в 1961 г. в нем опубликовано более тысячи статей.

Типовые исходные данные в новой парадигме – объекты нечисловой природы (элементы нелинейных пространств, которые нельзя складывать и умножать на число, например, множества, бинарные отношения), а в старой – числа, конечномерные векторы, функции. Ранее (в старой парадигме) для расчетов использовались разнообразные суммы, однако объекты нечисловой природы нельзя складывать, поэтому в новой парадигме применяется другой математический аппарат, основанный на расстояниях между объектами нечисловой природы и решении задач оптимизации.

Изменились постановки задач анализа данных и экономико-математического моделирования. Старая парадигма математической статистики исходит из идей начала XX в., когда К. Пирсон предложил четырехпараметрическое семейство распределений для описания распределений реальных данных. В это семейство как частные случаи входят, в частности, подсемейства нормальных, экспоненциальных, Вейбулла-Гнеденко, гамма-распределений. Сразу было ясно, что распределения реальных данных, как правило, не входят в семейство распределений Пирсона (об этом говорил, например, академик С.Н. Бернштейн в 1927 г. в докладе на Всероссийском съезде математиков [43]; см. также [44]). Однако математическая теория параметрических семейств распределений (методы оценивание параметров и проверки гипотез) оказалась достаточно интересной с теоретической точки зрения (в ее рамках был доказан ряд трудных теорем), и именно на ней до сих пор основано преподавание во многих вузах. Итак, в старой парадигме основной подход к описанию данных - распределения из параметрических семейств, а оцениваемые величины – их параметры, в новой парадигме рассматривают произвольные распределения, а оценивают - характеристики и плотности распределений, зависимости, правила диагностики и др. Центральная часть теории – уже не статистика числовых случайных величин, а статистика в пространствах произвольной природы, т.е. нечисловая статистика [15, 45].

В старой парадигме источники постановок новых задач - традиции, сформировавшиеся к середине XX века, а в новой - современные потребности математического моделирования и анализа данных (XXI век), т.е. запросы практики. Конкретизируем это общее различие. В старой парадигме типовые результаты - предельные теоремы, в новой - рекомендации для конкретных значений параметров, в частности, объемов выборок. Изменилась роль информационных технологий – ранее они использовались в основном для расчета таблиц (в частности, информатика находилась вне математической статистики), теперь же они - инструменты получения выводов (имитационное моделирование, датчики псевдослучайных чисел, методы размножение выборок, в т.ч. бутстреп, и др.). Вид постановок задач приблизился к потребностям практики – при анализе данных от отдельных задач оценивания и проверки гипотез перешли к статистическим технологиям (технологическим процессам анализа данных). Выявилась важность проблемы «стыковки алгоритмов» - влияния выполнения предыдущих алгоритмов в технологической цепочке на условия применимости последующих алгоритмов. В старой парадигме эта проблема не рассматривалась, для новой – весьма важна.

Если в старой парадигме вопросы методологии моделирования практически не обсуждались, достаточными признавались схемы начала XX в., то в новой парадигме роль методологии (учения об организации

деятельности) [46] является основополагающей. Резко повысилась роль моделирования – от отдельных систем аксиом произошел переход к системам моделей. Сама возможность применения вероятностного подхода теперь – не «наличие повторяющегося комплекса условий» (реликт физического определения вероятности (по Мизесу), использовавшегося до аксиоматизации теории вероятностей А.Н. Колмогоровым в 1930-х гг.), а наличие обоснованной вероятностно-статистической модели. Если раньше данные считались полностью известными, то для новой парадигмы характерен учет свойств данных, в частности, интервальных и нечетких [47]. Изменилось отношение к вопросам устойчивости выводов – в старой парадигме практически отсутствовал интерес к этой тематике, в новой разработана развитая теория устойчивости (робастности) выводов по отношению к допустимым отклонениям исходных данных и предпосылок моделей [13, 48].

Результаты сравнения парадигм удобно представить в виде табл. 1.

Таблица 1. **Сравнение основных характеристик старой и новой парадигм**

№	***Характеристика***	***Старая парадигма***	***Новая парадигма***
1	Типовые исходные данные	Числа, конечномерные вектора, функции	Объекты нечисловой природы [15, 45]
2	Основной подход к моделированию данных	Распределения из параметрических семейств	Произвольные функции распределения
3	Основной математический аппарат	Суммы и функции от сумм	Расстояния и алгоритмы оптимизации [[15, 45]]
4	Источники постановок новых задач	Традиции, сформировавшиеся к середине XX века	Современные прикладные потребности анализа данных (XXI век)
5	Отношение к вопросам устойчивости выводов	Практически отсутствует интерес к устойчивости выводов	Развитая теория устойчивости (робастности) выводов [13, 48]
6	Оцениваемые величины	Параметры распределений	Характеристики, функции и плотности распределений, зависимости, правила диагностики и др.
7	Возможность применения	Наличие повторяющегося комплекса условий	Наличие обоснованной вероятностно-статистической модели
8	Центральная часть теории	Статистика числовых случайных величин	Нечисловая статистика [15, 45]
9	Роль информационных технологий	Только для расчета таблиц (информатика находится вне статистики)	Инструменты получения выводов (датчики псевдослучайных чисел, размножение выборок, в

			т.ч. бутстреп, и др.) [49, 50]
10	Точность данных	Данные полностью известны	Учет неопределенности данных, в частности, интервальности и нечеткости [47]
11	Типовые результаты	Предельные теоремы (при росте объемов выборок)	Рекомендации для конкретных объемов выборок
12	Вид постановок задач	Отдельные задачи оценивания параметров и проверки гипотез	Высокие статистические технологии (технологические процессы анализа данных) [51]
13	Стыковка алгоритмов	Не рассматривается	Весьма важна при разработке процессов анализа данных
14	Роль моделирования	Мала (отдельные системы аксиом)	Системы моделей – основа анализа данных
15	Анализ экспертных оценок	Отдельные алгоритмы	Прикладное «зеркало» общей теории [31, 32]
16	Роль методологии	Практически отсутствует	Основополагающая [13, 52, 53]

1.6. Учебная литература, подготовленная в соответствии с новой парадигмой

В 1992 г. на базе секции статистических методов Всесоюзной статистической ассоциации была организована Российская ассоциация статистических методов, а в 1996 г. – Российская академия статистических методов. В соответствии с новой парадигмой проводились научные исследования, публиковались статьи, по этой тематике были организованы семинары и конференции. Однако в соответствии с ситуацией 90-х годов размах работ сокращался, как и число участвующих в них исследователей. Поэтому на рубеже тысячелетий нами было принято решение сосредоточить усилия на подготовке учебной литературы, соответствующей новой парадигме.

Первым был выпущенный в 2002 г. учебник по эконометрике [18], переизданный в 2003 г. и в 2004 г. Четвертое издание «Эконометрики» [54] существенно переработано. Оно соответствует первому семестру курса, в отличие от первых трех изданий, содержащих материалы для годового курса. В четвертое издание [54] включены новые разделы, полностью обновлена глава про индекс инфляции, добавлено методическое обеспечение.

В нашем фундаментальном курсе 2006 г. по прикладной статистике [20] в рамках новой парадигмы рассмотрены как нечисловая статистика, так и классические разделы прикладной статистики, посвященные методам

обработки элементов линейных пространств - чисел, векторов и функций (временных рядов).

В том же 2006-м году в рамках новой парадигмы был выпущен курс теории принятия решений [26]. Его сокращенный (в 1,5 раза) вариант вышел годом раньше [55].

В соответствии с потребностями практики в России в 2005 г. введена новая учебная специальность 220701 «Менеджмент высоких технологий», относящаяся к тогда же введенному направлению подготовки 220700 «Организация и управление наукоемкими производствами», предназначенному для обеспечения инженерами-менеджерами высокотехнологичных предприятий. Большинство студентов научно-учебного комплекса «Инженерный бизнес и менеджмент» МГТУ им. Н.Э. Баумана обучаются по этой специальности. Общий взгляд на нее представлен в учебнике [56].

Государственным образовательным стандартом по специальности «Менеджмент высоких технологий» предусмотрено изучение дисциплины «Организационно-экономическое моделирование». Одноименный учебник выпущен в трех частях (томах). Первая из них [15] посвящена сердцевине новой парадигмы – нечисловой статистике. Ее прикладное «зеркало» - вторая часть [31], современный учебник по экспертным оценкам. В третьей части [57] наряду с основными постановками задач анализа данных (чисел, векторов, временных радов) и конкретными статистическими методами анализа данных классических видов (чисел, векторов, временных рядов) рассмотрены вероятностно-статистические модели в технических и экономических исследованиях, медицине, социологии, истории, демографии, а также метод когнитивных карт (статистические модели динамики).

В названиях еще двух учебников есть термин «организационно-экономическое моделирование». Это вводная книга по менеджменту [58] и современный учебник по теории принятия решений [59], в которых содержание соответствует новой парадигме, в частности, подходам трехтомника по организационно-экономическому моделированию [15, 31, 57]. Отметим, что, в учебнике [59] значительно большее внимание по сравнению с более ранним учебником по теории принятия решений [26] уделено теории и практике экспертных оценок, в то время как общие проблемы менеджмента выделены для обсуждения в отдельное издание [58].

К рассмотренному выше корпусу учебников примыкают справочник по минимально необходимым для восприятия рассматриваемых курсов понятиям теории вероятностей и прикладной математической статистики [60] и книги по промышленной и экологической безопасности [61] и [62], в которых большое место занимает изложение научных результатов в соответствии с новой парадигмой, в частности, активно используются

современные статистические и экспертные методы, математическое моделирование. Опубликовано еще несколько изданий, например, [63], но от их рассмотрения здесь воздержимся.

Основные книги А.И. Орлова были доработаны и переизданы в 2020 - 2022 гг. [64 - 74].

Публикация учебной литературы на основе новой парадигмы шла непросто. Зачастую издать определенную книгу удавалось с третьего-четвертого раза. Неоценима поддержка Научно-учебного комплекса «Инженерный бизнес и менеджмент» и МГТУ им. Н.Э. Баумана в целом, Учебно-методического объединения вузов по университетскому политехническому образованию.

Все перечисленные монографии, учебники, учебные пособия имеются в Интернете в свободном доступе. Соответствующие ссылки приведены на персональной странице А.И. Орлова на сайте МГТУ им. Н.Э. Баумана http://www.bmstu.ru/ps/~orlov/ и на аналогичной странице форума http://forum.orlovs.pp.ru/viewtopic.php?f=1&t=1370, однако иногда различны названия и выходные данные книг в бумажном и электронном вариантах.

Информация о новой парадигме появилась в печати недавно – в 2011 г. (см. [1 - 12]), когда публикация книг с изложением научных подходов и результатов на основе новой парадигмы математических методов исследования была уже практически закончена. Разработчики новой парадигмы не без оснований опасались, что им могут помешать довести работу до конца. В своей тактике публикаций они во многом следовали Гауссу, который воздерживался от публикации работ по неевклидовой геометрии, опасаясь «криков беотийцев» [65, с.91].

Опасения, увы, имели основания. Так, в июне 2015 г. была сделана попытка удалить из Википедии статью "Орлов Александр Иванович (учёный)". Выставивший статью на удаление некий Булатов ("номинатор") написал: "Значимость учёного возможна, но подобный торжественно-помпезный стиль совершенно неприемлем для Википедии. Статья требует полного переписывания в нейтральном стиле с привлечением независимых источников. — Bulatov 18:18, 9 июня 2015 (UTC)". Нетрудно понять причины поведения номинатора. Как нетрудно установить, номинатор - Булатов Александр Вячеславович - работал (судя по http://www.ipu.ru/node/116) в Институте проблем управления (ИПУ) РАН в Лаборатории № 45 под названием «Математические методы исследования оптимальных управляемых систем». Поэтому он так резко отреагировал на фразу «Разработана новая парадигма математических методов исследования», в которой есть значительное совпадение с названием научного подразделения, в котором он числится. Следовало бы ожидать, что к.ф.-м.н. Булатов А.В. познакомится с новой парадигмой, которой посвящено достаточно много публикаций (например, указанных в. РИНЦ). Или обратится к руководству своего института за разъяснениями.

Например, к зам. директора ИПУ РАН член-корр. РАН Д.А. Новикову, соавтору проф. А.И. Орлова по ряду работ. Вместо этого к.ф.-м.н. Булатов А.В. потребовал удаления статьи из-за одной фразы. Очевидно, обсуждению на научном уровне вопроса о новой парадигме в Википедии не место. Обсуждать его надо на научных собраниях, в научной печати. Если к.ф.-м.н. Булатов А.В. не согласен с проф. А.И. Орловым, он может написать об этом статью или выступить на конференции. [Этот абзац, разъясняющий суть дела, был сразу же кем-то удален из обсуждения в Википедии.]

Ярлык "К удалению" висел до октября 2015 г. С полным текстом обсуждения можно познакомиться на Интернет-ресурсе "Википедия:К удалению/9 июня 2015". В обсуждении наряду со здравыми мнениями во всей красе показали себя лица, ничего не понимающие в научной деятельности. Отстоять само существование статьи "Орлов Александр Иванович (учёный)" удалось, лишь затратив десятки квалифицированных трудочасов (объем обсуждения превышает объем настоящей главы). При этом статья в Википедии была испорчена большим числом безграмотных поправок.

Новая парадигма математических методов исследования с интересом встречена научной общественностью, обсуждалась на многочисленных международных и всероссийских научных конференциях. К ранее указанным публикациям [1 - 12] добавим литературные ссылки [76 - 86].

На основе сказанного выше можно констатировать, что к настоящему моменту рекомендация Учредительного съезда Всесоюзной статистической ассоциации (1990) по созданию комплекта учебной литературы на основе новой парадигмы математических методов исследования выполнена. Предстоит большая работа по внедрению новой парадигмы организационно-экономического моделирования, эконометрики и статистики в научные исследования (теоретические и прикладные), прикладные научно-исследовательские работы и преподавание.

ГЛАВА 2. СТАТИСТИКА НЕЧИСЛОВЫХ ДАННЫХ - ЦЕНТРАЛЬНАЯ ЧАСТЬ СОВРЕМЕННОЙ ПРИКЛАДНОЙ СТАТИСТИКИ

Статистика нечисловых данных основана на системной нечеткой интервальной математике. В настоящее время статистика нечисловых данных - одна из четырех основных областей прикладной статистики. Остальные три - статистика чисел, многомерный статистический анализ, статистика временных рядов и случайных процессов. В свою очередь статистика нечисловых данных делится на статистику в (нелинейных)

пространствах общей природы и разделы, посвященные конкретным типам нечисловых данных, такие, как статистика интервальных данных, статистика нечетких множеств, статистика бинарных отношений и др. Естественно, что научные результаты, полученные в рамках статистики в пространствах общей природы, могут быть использованы для конкретных видов данных (например, теория непараметрических оценок плотности распределения вероятностей). Следовательно, статистика в пространствах общей природы - центральная часть прикладной статистики, а включающая ее статистика нечисловых данных - основная область прикладной статистики. Это утверждение подтверждается, в частности, анализом публикаций в разделе "Математические методы исследования" журнала "Заводская лаборатория. Диагностика материалов" - основного издания по прикладной статистике в России. По данным [1], на первое место по числу публикаций вышла именно статистика нечисловых данных. Так, за десять лет (2006 - 2015) ей посвящены 27,6% всех публикаций раздела "Математические методы исследования", т.е. 63,0% статей по прикладной статистике [1].

Первоначально для новой области прикладной статистики использовался термин "статистика объектов нечисловой природы". Он впервые появился в 1979 г. в монографии [2] для обозначения совокупности некоторых полученных в ней научных результатов. В том же году в статье [3] нами была развернута программа построения этой новой области статистических методов, приведены первоначальные формулировки большинства основных теорем. Через год в «Заводской лаборатории» (так тогда назывался этот журнал) появилась обобщающая статья [4] пяти наиболее активных авторов среди занимавшихся различными аспектами статистики нечисловых данных. Итоги первых десяти лет развития новой области прикладной статистики были подведены в нашем обстоятельном обзоре [5] (120 литературных ссылок). Дальнейшее развитие было не менее плодотворным. Обзор [6] за тридцать лет содержал 150 литературных ссылок. К тридцатилетию вышел и первый учебник по статистике нечисловых данных [7]. В названии учебника использован термин "нечисловая статистика". Он представляется слишком кратким, в то время как исходный термин "статистика объектов нечисловой природы" - слишком тяжеловесным. Далее будем называть рассматриваемую область прикладной статистики "статистикой нечисловых данных". Такое название в наилучшей степени отражает ее содержание. Все три термина (статистика объектов нечисловой природы, статистика нечисловых данных, нечисловая статистика) - синонимы.

В настоящей главе обсудим содержание, развитие и основные идеи статистики нечисловых данных. Появление и развитие статистики нечисловых данных соответствуют переходу к новой парадигме математических методов исследования (см. главу 1).

2.1. Различные виды нечисловых данных

Типичный исходный объект в прикладной статистике - это выборка, т.е. совокупность независимых одинаково распределенных случайных элементов. Какова природа этих элементов? В классической математической статистике элементы выборки - это числа. В многомерном статистическом анализе - вектора. А в статистике нечисловых данных элементы выборки - это объекты нечисловой природы, которые нельзя складывать и умножать на числа. Другими словами, объекты нечисловой природы лежат в пространствах, не имеющих векторной (линейной) структуры.

Примерами объектов нечисловой природы являются:

- значения качественных признаков (измеренных в шкалах наименований и порядковых), в том числе результаты кодировки объектов с помощью заданного перечня категорий (градаций);
- упорядочения (ранжировки) экспертами объектов экспертизы - образцов продукции (при оценке её технического уровня, качества, конкурентоспособности)), ее характеристик; заявок на проведение научных работ (при проведении конкурсов на выделение грантов) и т.п.;
- классификации, т.е. разбиения объектов на группы сходных между собой (кластеры);
- толерантности, т.е. бинарные отношения, описывающие сходство объектов между собой, например, сходства тематики научных работ, оцениваемого экспертами с целью рационального формирования экспертных советов внутри определенной области науки;
- результаты парных сравнений или контроля качества продукции по альтернативному признаку («годен» - «брак»), т.е. последовательности из 0 и 1;
- множества (обычные или нечеткие), например, зоны, пораженные коррозией, или перечни возможных причин аварии, составленные экспертами независимо друг от друга;
- слова, предложения, тексты;
- графы;
- вектора, координаты которых - совокупность значений разнотипных признаков, например, результат составления статистического отчета о научно-технической деятельности организации или анкета эксперта, в которой ответы на часть вопросов носят качественный характер, а на часть - количественный;
- ответы на вопросы экспертной, медицинской, маркетинговой или социологической анкеты, часть из которых носит количественный характер (возможно, интервальный), часть сводится к выбору одной из нескольких подсказок, а часть представляет собой тексты; и т.д.

Все средства измерения имеют погрешности. Однако до недавнего времени это очевидное обстоятельство никак не учитывалось в статистических процедурах. Только с конца 1970-х годов начала развиваться статистика интервальных данных, в которой предполагается, что исходные данные (элементы выборки) - это не числа, а интервалы. Статистику интервальных данных можно рассматривать как часть интервальной математики. Выводы в ней часто принципиально отличны от классических.

Различным подходам к статистическому анализу интервальных данных посвящена принципиально важная дискуссия [12]. Работают две основные научные школы - А.П. Вощинина и наша. В первой из них изучены проблемы регрессионного анализа, планирования эксперимента, сравнения альтернатив и принятия решений в условиях интервальной неопределенности К этой школе относится недавняя работа Н.В. Скибицкого [13]. В разработанной нами асимптотической статистике интервальных данных на значения случайных величин наложены малые интервальные неопределенности. Основные результаты этого направления подробно изложены в обширных главах учебников [7, 14, 15], монографии [16], в недавнем обзоре [17].

Интервальные данные можно рассматривать как частный случай нечетких множеств. Действительно, если характеристическая функция нечеткого множества равна 1 на некотором интервале и равна 0 вне этого интервала, то задание такого нечеткого множества эквивалентно заданию интервала. С методологической точки зрения важно, что *теория нечетких множеств в определенном смысле сводится к теории случайных множеств*. Цикл соответствующих теорем приведен в монографии [2], а также в учебниках [7, 14, 15, 18], монографии [16], недавней статье [19]. Казалось бы, много публикаций. Но приходится констатировать, что отнюдь не все специалисты (как математики, так и экономисты, и специалисты по управлению) знакомы с теоремами о сведении теории нечетких множеств к теории вероятностей. Применение результатов о сведении теории нечеткости к теории случайных множеств при решении прикладных задач - дело будущего.

2.2. Об истории и структуре статистической науки

Развитие статистических методов в нашей стране проанализировано в главе 2 монографии [20]. Дадим здесь краткую сводку, позволяющую выявить роль статистики нечисловых данных.

К 60-м годам XX в. в стране и мире сформировалась научно-практическая дисциплина, которую называем классической математической статистикой. Новое поколение отечественных исследователей училось теории по фундаментальной монографии шведского математика Г. Крамера [21], написанной в военные годы и

впервые изданной на русском языке в 1948 г. Из прикладных руководств назовем учебник [22] и таблицы с комментариями [23].

Затем внимание многих специалистов сосредоточилось на изучении математических конструкций, используемых в статистике. Примером таких работ является монография [24]. В ней получены продвинутые математические результаты, но трудно выделить рекомендации для статистика, анализирующего конкретные данные.

Как реакция на уход в математику выделилась новая научная дисциплина - прикладная статистика. В базовом учебнике по прикладной статистике [14] в качестве рубежа, когда это стало очевидным, указан 1981 г. – дату выхода массовым тиражом (33 940 экз.) сборника [25], в названии которого использован термин «прикладная статистика». С этого времени линии развития математической статистики и прикладной статистики разошлись. Первая из этих дисциплин полностью ушла в математику, перестав интересоваться практическими делами. Вторая позиционировала себя в качестве науки об обработке данных – результатов наблюдений, измерений, испытаний, анализов, опытов, обследований [14].

Вполне естественно, что в прикладной статистике стали создаваться свои математические методы и модели. Необходимость их развития вытекает из потребностей конкретных прикладных исследований. Это математизированное ядро прикладной статистики назовем *теоретической статистикой*. Тогда под собственно *прикладной статистикой* следует понимать обширную промежуточную область между теоретической статистикой и применением статистических методов в конкретных областях. Таким образом, общая схема современной статистической науки выглядит следующим образом (от абстрактного к конкретному):

1. Математическая статистика (математические методы в статистике) – часть математики, изучающая статистические структуры. Сама по себе не дает рецептов анализа статистических данных, однако разрабатывает методы, полезные для использования в теоретической статистике.

2. Теоретическая статистика – наука, посвященная моделям и методам анализа конкретных статистических данных.

3. Прикладная статистика (в узком смысле) посвящена статистическим технологиям сбора и обработки данных. В нее входят, в частности, вопросы формирования вероятностно-статистических моделей и выбора конкретных методов анализа данных (т.е. методология прикладной статистики и других статистических методов), проблемы разработки и применения информационных статистических технологий, организации выборочных исследований, сбора данных и использования статистических программных продуктов.

4. Применение статистических методов в конкретных областях (в экономике и управлении (менеджменте) – эконометрика, в биологии – биометрика, в химии – хемометрия, в технических исследованиях –

технометрика, в геологии, демографии, социологии, медицине, психологии, истории, и т.д.).

Часто позиции 2 и 3 вместе называют прикладной статистикой. Иногда позицию 1 именуют теоретической статистикой. Эти терминологические расхождения связаны с тем, что описанное выше развитие рассматриваемой научно-прикладной области не сразу, не полностью и не всегда адекватно отражается в сознании специалистов. Так, до сих пор выпускают учебники, соответствующие старой парадигме - уровню представлений середины XX века.

Примечание. Здесь мы уточнили схему внутреннего деления статистической теории, предложенную ранее в [26]. Естественный смысл приобрели термины «теоретическая статистика» и «прикладная статистика» (в узком смысле). Однако необходимо иметь в виду, что в базовом учебнике [14] прикладная статистика понимается в широком смысле, т.е. как объединение позиций 2 и 3. К сожалению, в настоящее время невозможно отождествить теоретическую статистику с математической, поскольку последняя (как часть математики - научной специальности «теория вероятностей и математическая статистика») заметно оторвалась от задач практики.

Отметим, что математическая статистика, как и теоретическая с прикладной, заметно отличается от ведомственной науки органов официальной государственной статистики. ЦСУ, Госкомстат, Росстат применяли и применяют лишь проверенные временем приемы XIX века. Возможно, нам следовало бы от этого ведомства полностью отмежеваться и сменить название научной дисциплины, например, на «Анализ данных». В настоящее время компромиссным самоназванием является термин «статистические методы».

Во второй половине 1980-х годов в нашей стране развернулось общественное движение, имеющее целью создание профессионального объединения статистиков. Аналогами являются британское Королевское статистическое общество (основано в 1834 г.) и Американская статистическая ассоциация (создана в 1839 г.). К сожалению, деятельность учрежденной в 1990 г. Всесоюзной статистической ассоциации оказалась парализованной в результате развала СССР. Среди стран СНГ наибольшую активность в настоящее время проявляют узбекские исследователи, регулярно проводящие на высоком научном уровне представительные конференции по статистике и ее применениям.

2.3. О развитии статистики нечисловых данных

С 70-х годов XX в. в основном на основе запросов теории экспертных оценок (а также технических исследований, экономики, социологии и медицины) развивались различные направления статистики нечисловых данных. Были установлены основные связи между

конкретными видами таких объектов, разработаны для них базовые вероятностные модели [27]. Сводка полученных результатов дана в монографии [2], обзоре [4]. Это - предыстория статистики нечисловых данных. А история начинается с осмысления созданного, констатации [2, 3] в 1979 г. появления новой области прикладной статистики.

Следующий этап (1980-е годы) - развитие статистики нечисловых данных в качестве самостоятельного научного направления в рамках математических методов исследования, ядром которого являются методы статистического анализа данных произвольной природы. Для работ этого периода характерна сосредоточенность на внутренних (внутриматематических) проблемах статистики нечисловых данных. Проводились всесоюзные конференции, выпускались монографии, сборники трудов, защищались диссертации (Орлов А.И., Пярна К.А., Рыданова Г.В., Сатаров Г.А., Трофимов В.А., Шер А.П., Шмерлинг Д.С. и др.). Наиболее представительным является сборник [28], подготовленный совместно комиссией «Статистика объектов нечисловой природы» Научного Совета АН СССР по комплексной проблеме «Кибернетика» и Институтом социологических исследований АН СССР. Конкретная информация по работам 80-х годов имеется в обзорах [5, 6].

В настоящее время в связи с активным использованием наукометрических показателей разнообразными администраторами в области научной деятельности распространилась преувеличенная оценка роли журналов в развитии науки. Опыт статистики нечисловых данных показывает, что естественная цепочка развития научного результата такова: тезисы доклада — тематический сборник — монография — учебник — широкое использование [29]. Для развития нового направления публикации в научных журналах, вообще говоря, не обязательны. Ясно, что издание собственных журналов или завоевание позиций в уже существующих возможно лишь на этапе зрелости нового направления, но не на этапе его создания.

К 1990-м годам статистика нечисловых данных с теоретической точки зрения была достаточно хорошо развита, основные идеи, подходы и методы были разработаны и изучены математически, в частности, доказано достаточно много теорем. Однако она оставалась недостаточно апробированной на практике. И в 90-е годы наступило время перейти от теоретико-статистических исследований к применению полученных результатов на практике, а также включить их в учебный процесс, что и было сделано. В 90-е годы в «Заводской лаборатории» опубликованы обзоры [5, 11, 27] по статистике объектов нечисловой природы и многочисленные конкретные исследования, рассмотренные позже в [6]. Серия работ была выполнена по статистике интервальных данных.

В 2000-е годы наиболее заметное явление - развернутые изложения основных результатов статистики нечисловых данных в учебниках по

прикладной статистике, теории принятия решений, эконометрике [14, 15, 18]. Выпущен первый учебник по статистике нечисловых данных [7].

В 2010-е годы представленная научной общественности новая парадигма математических методов исследования закрепила положение статистики нечисловых данных как центральной быстро растущей части современной прикладной статистики (ср. обзор [1]). Опубликована серия работ по непараметрическим оценкам плотности распределения.

2.4. Основные идеи статистики в пространствах общей природы

В чем принципиальная новизна статистики нечисловых данных? Для классической математической статистики характерна операция сложения. При расчете выборочных характеристик распределения (выборочное среднее арифметическое, выборочная дисперсия и др.), в регрессионном анализе и других областях этой научной дисциплины постоянно используются суммы. Математический аппарат - законы больших чисел, Центральная предельная теорема и другие теоремы - нацелены на изучение сумм. Принципиально важно, что в статистике нечисловых данных нельзя использовать операцию сложения, поскольку элементы выборки лежат в пространствах, где нет операции сложения. Методы обработки нечисловых данных основаны на принципиально ином математическом аппарате - на применении различных расстояний в пространствах объектов нечисловой природы и решении задач оптимизации.

Следует отметить, что в статистике нечисловых данных одна и та же математическая схема может с успехом применяться во многих прикладных областях, для анализа данных различных типов, а потому ее лучше всего формулировать и изучать в наиболее общем виде, для объектов произвольной природы.

Кратко рассмотрим несколько идей, развиваемых в статистике нечисловых данных для элементов выборок, лежащих в пространствах произвольного вида. Они нацелены на решение классических задач описания данных, оценивания, проверки гипотез - но для неклассических данных, а потому неклассическими методами.

Первой обсудим проблему определения средних величин. В рамках теории измерений удается указать вид средних величин, соответствующих тем или иным шкалам измерения [30]. Теория измерений [31, 32], в середине XX в. рассматривавшаяся как часть математического обеспечения психологии, к настоящему времени признана общенаучной дисциплиной. Проблемы теории измерений постоянно рассматриваются в разделе "Математические методы исследования" журнала "Заводская лаборатория. Диагностика материалов" (см. соответствующий раздел [33]).

В классической математической статистике средние величины вводят с помощью операций сложения (выборочное среднее

арифметическое, математическое ожидание) или упорядочения (выборочная и теоретическая медианы). В пространствах произвольной природы средние значения нельзя определить с помощью операций сложения или упорядочения. Теоретические и эмпирические средние приходится вводить как решения экстремальных задач. Теоретическое среднее определяется как решение задачи минимизации математического ожидания (в классическом смысле) расстояния от случайного элемента со значениями в рассматриваемом пространстве до фиксированной точки этого пространства (минимизируется указанная функция от этой точки). Для получения эмпирического среднего математическое ожидание берется по эмпирическому распределению, т.е. берется сумма расстояний от некоторой точки до элементов выборки и затем минимизируется по этой точке (примером является медиана Кемени [34], методам нахождения которой посвящены недавние работы М.С. Жукова и его диссертация [35]). При этом как эмпирическое, так и теоретическое средние как решения экстремальных задач могут быть не единственными элементами рассматриваемого пространства, а являться некоторыми множествами таких элементов. Они могут оказаться и пустыми. Тем не менее удалось сформулировать и доказать законы больших чисел для та определенных средних величин, т.е. установить сходимость (в специально определенном смысле) эмпирических средних к теоретическим [7, 36].

Оказалось, что методы доказательства законов больших чисел допускают существенно более широкую область применения, чем та, для которой они были разработаны. А именно, удалось изучить асимптотику решений экстремальных статистических задач [7, 37], к которым, как известно, сводится большинство постановок прикладной статистики. В частности, дополнительно к законам больших чисел установлена состоятельность оценок минимального контраста, в том числе оценок максимального правдоподобия и робастных оценок. К настоящему времени подобные оценки изучены также и в статистике интервальных данных. Полученные результаты относительно асимптотики решений экстремальных статистических задач применяются в ряде работ.

В статистической теории в пространствах общей природы большую роль играют непараметрические оценки плотности распределения вероятностей, используемые, в частности, в различных алгоритмах регрессионного, дискриминантного, кластерного анализов. В статистике нечисловых данных предложен и изучен ряд типов непараметрических оценок плотности в пространствах произвольной природы, в том числе в дискретных пространствах [38]. В частности, доказана их состоятельность, изучена скорость сходимости и установлен (для ядерных оценок плотности) примечательный факт совпадения наилучшей скорости сходимости в произвольном пространстве с той, которая имеет быть в классической теории для числовых случайных величин [39].

Введем обозначения. Пусть (Z, A) – измеримое пространство, p и q – сигма-конечные меры на (Z, A), причем p абсолютно непрерывна относительно q, т.е. из $q(B) = 0$ следует $p(B) = 0$ для любого множества B из сигма-алгебры A. В этом случае на (Z, A) существует неотрицательная измеримая функция $f(x)$ такая, что

$$q(C) = \int_C f(x)\, p(dx) \tag{1}$$

для любого множества C из сигма-алгебры измеримых множеств A. Функция $f(x)$ называется производной Радона - Никодима меры q по мере p, а в случае, когда q - вероятностная мера, также плотностью вероятности q по отношению к мере p [40, с.460].

Пусть X_1, X_2 ,…, X_n – независимые одинаково распределенные случайные элементы (величины), распределение которых задается вероятностной мерой q. Нами введено несколько видов непараметрических оценок плотности вероятности q по выборке X_1, X_2 ,…, X_n. Подробнее изучены линейные оценки. Подробнее рассмотрены их частные случаи – ядерные оценки плотности в пространствах произвольной природы. Асимптотическая теория ядерных оценок плотности развита, прежде всего, для нужд статистики конкретных видов объектов нечисловой природы, в которой основной интерес представляют конечные пространства Z. Мера p при этом не непрерывная, а дискретная, например, считающая. Таким образом, в рамках единого подхода удается рассмотреть оценки плотностей и оценки вероятностей.

В предположении непрерывности неизвестной плотности $f(x)$ представляется целесообразным «размазать» каждый атом эмпирической меры, т.е. рассмотреть линейные оценки, введенные в нашей первой работе по статистике нечисловых данных [3, с.24]:

$$f_n(x) = \frac{1}{n} \sum_{1 \le i \le n} g_n(x, X_i), \quad g_n : Z^2 \to R^1, \tag{2}$$

в которых действительнозначные функции g_n удовлетворяют некоторым условиям регулярности.

Пусть d – показатель различия (синоним - мера близости) на Z [7] (в наиболее важных частных случаях – метрика (расстояние) на Z). Нами введены ядерные оценки плотности – оценки вида (2) с

$$g_n(x, X_i) = \frac{1}{b(h_n, x)} K\left(\frac{d(x, X_i)}{h_n} \right), \quad K : [0, +\infty) \to R^1, \tag{3}$$

где $K = K(u)$ – ядро (ядерная функция), h_n – последовательность положительных чисел (показателей размытости), $b(h_n, x)$ – нормировочный множитель. Линейные оценки (2) с функциями g_n из (3) названы нами «обобщенными оценками типа Парзена-Розенблатта», т.к. в частном случае $Z = R^1$, $d(x, X_i) = | x - X_i |$, $b(h_n, x) = h_n$ они переходят в известные оценки, введенные Розенблаттом и Парзеном.

Цель статей [38, 41] - завершение цикла работ, посвященного математическому изучению асимптотических свойств различных видов непараметрических оценок плотности распределения вероятности в пространствах общей природы. Изучен средний квадрат ошибки ядерной оценки плотности. С целью максимизации порядка его убывания обоснован выбор ядерной функции и последовательности показателей размытости. Основные понятия - круговая функция распределения и круговая плотность. Порядок сходимости в общем случае тот же, что и при оценивании плотности числовой случайной величины [39], но основные условия наложены не на плотность случайной величины, а на круговую плотность. Далее рассматриваем другие виды непараметрических оценок плотности - гистограммные оценки и оценки типа Фикс-Ходжеса. Затем изучаем непараметрические оценки регрессии и их применение для решения задач дискриминантного анализа в пространстве общей природы

Дискриминантный, кластерный, регрессионный анализы в пространствах произвольной природы основаны либо на параметрической теории - и тогда применяется подход, связанный с асимптотикой решения экстремальных статистических задач - либо на непараметрической теории - и тогда используются алгоритмы на основе непараметрических оценок плотности.

Для анализа нечисловых, в частности, экспертных данных весьма важны методы классификации [42]. Интересно движение мысли в обратном направлении - наиболее естественно ставить и решать задачи классификации, основанные на использовании расстояний или показателей различия, именно в рамках статистики объектов нечисловой природы (а не, скажем, многомерного статистического анализа). Это касается как распознавания образов с учителем (другими словами, дискриминантного анализа), так и распознавания образов без учителя (т.е. кластерного анализа). Аналогичным образом задачи многомерного шкалирования, т.е. визуализации данных, также естественно отнести к статистике объектов нечисловой природы. Важны методы оценки истинной размерности признакового пространства [43].

Отметим несколько конкретных научных результатов математической теории классификации. В задачах диагностики (дискриминантного анализа), как следует из леммы Неймана-Пирсона, целесообразно строить алгоритмы на основе отношения непараметрических оценок плотностей распределения вероятностей, соответствующих классам. Установлено, что наилучшим показателем качества алгоритма диагностики является прогностическая сила [44]. Устойчивость классификации относительно выбора метода кластер-анализа обосновывает вывод о реальности кластеров. И т.д. (см. соответствующий раздел в обзоре [1]).

Для проверки гипотез в пространствах нечисловой природы могут быть использованы статистики интегрального типа [3, 45], в частности, типа омега-квадрат. Отметим, что предельная теория таких статистик, построенная первоначально в классической постановке, приобрела естественный (завершенный, изящный) вид именно для пространств произвольного вида [46], поскольку при этом удалось провести рассуждения, опираясь на базовые математические соотношения, а не на те частные (с общей точки зрения), что были связаны с конечномерным пространством.

2.5. О некоторых областях статистики конкретных нечисловых данных

Кратко рассмотрим некоторые статистические методы анализа данных, лежащих в конкретных пространствах нечисловой природы.

Непараметрическая статистика – это прежде всего ранговая статистика, т.е. основанная на рангах – номерах элементов выборок в вариационных рядах. Ранги измерены в порядковых шкалах, а значения ранговых статистик инвариантны относительно любых строго возрастающих преобразований - допустимых преобразований в таких шкалах. Непараметрическая статистика позволяет делать статистические выводы, оценивать характеристики и плотность распределения, проверять статистические гипотезы без слабо обоснованных предположений о том, что функция распределения элементов выборки входит в то или иное параметрическое семейство. Например, широко распространена вера в то, что статистические данные часто подчиняются нормальному распределению. Математики думают, что это - экспериментальный факт, установленный в прикладных исследованиях. Прикладники уверены, что математики доказали нормальность результатов наблюдений. Между тем анализ конкретных результатов наблюдений, в частности, погрешностей измерений, приводит всегда к одному и тому же выводу - в подавляющем большинстве случаев реальные распределения существенно отличаются от нормальных. На этот объективный факт обращал внимание В.В. Налимов в своей классической монографии [47]. Научная школа метролога П.В. Новицкого многочисленными экспериментами подтвердила отсутствие нормальности погрешностей измерений [48]. В сводке [49], в частности, установлено, что по выборкам объемов 6 - 50, как правило, не удается отличить нормальное распределение от других видов распределений.

Некритическое использование гипотезы нормальности часто приводит к значительным ошибкам, например, при отбраковке резко выделяющихся результатов наблюдений (выбросов), при статистическом контроле качества и в других случаях [14]. Поэтому целесообразно использовать непараметрические методы, в которых на функции распределения результатов наблюдений наложены лишь весьма слабые

требования. Обычно предполагается лишь их непрерывность. К настоящему времени с помощью непараметрических методов можно решать практически тот же круг задач, что ранее решался параметрическими методами. Примеры - оценивание характеристик распределения и проверка гипотезы однородности для независимых и связанных выборок [14]. Однако эта информация еще не вошла в массовое сознание. До сих пор тупиковой тематике параметрической статистики посвящены обширные разделы учебников и программных продуктов. Современное состояние непараметрической статистики проанализировано в [50]. Эта область исследований продолжает активно развиваться.

Представляют практический интерес результаты, связанные с конкретными областями статистики объектов нечисловой природы, в частности, со статистикой нечетких множеств [51] и со статистикой случайных множеств (напомним, что теория нечетких множеств в определенном смысле сводится к теории случайных множеств [19]), с непараметрической теорией парных сравнений и люсианов (бернуллиевских бинарных векторов) [52], с аксиоматическим введением метрик в конкретных пространствах объектов нечисловой природы [7], а также с рядом других конкретных постановок.

Результаты контроля штучной продукции по альтернативному признаку представляют собой последовательности из 0 и 1, т.е. объекты нечисловой природы (люсианы), а потому теорию статистического контроля относят к статистике нечисловых данных [5, 6]. Постоянно публикуем работы по этой тематике, предназначенные для специалистов по статистическим методам управления качеством продукции (см. [15, 53] и др.).

Статистика нечисловых данных порождена потребностями практики, прежде всего в области экспертных оценок. Можно констатировать, что анализ экспертных оценок [54] - это прикладное «зеркало» общей теории. Решения задач теории экспертных оценок обобщались в статистике нечисловых данных. При движении мысли в обратном направлении результаты статистики в пространствах общей природы интерпретировались для анализа экспертных оценок. Как и для статистики нечисловых данных в целом, публикации шли по траектории: тезисы доклада — тематический сборник — монография — учебник — широкое использование [29]. Вполне естественно, что названия сборников трудов неформального научного коллектива, развивающего статистику нечисловых данных, начинались со слов «Экспертные оценки» [55 - 58]. Отметим, что публикации в журналах не сыграли значительной роли в развитии рассматриваемых научных направлений. Обзор развития экспертных технологий в нашей стране дан в [59].

Вопросы внедрения математических методов исследования всегда были в центре внимания специалистов по статистике нечисловых данных

[60]. Подчеркивалось большое теоретическое и прикладное значение статистики нечисловых данных, необходимость перехода от отдельных методов анализа данных к разработке высоких статистических технологий [61] и использования современных систем внедрения математических методов, таких как система «Шесть сигм» и ее аналоги. Обсуждались проблемы программного обеспечения [35, 62]. Однако приходится констатировать, что создание линейки современных программных продуктов по статистике нечисловых данных – пока дело будущего.

2.6. Некоторые нерешенные задачи статистики нечисловых данных

Начнем с обсуждения влияния отклонений от традиционных предпосылок. В вероятностной теории статистических методов выборка обычно моделируется как конечная последовательность независимых одинаково распределенных случайных величин или векторов. В устаревшей парадигме середины XX в. часто предполагают, что эти величины (вектора) имеют нормальное распределение.

При внимательном взгляде совершенно ясна нереалистичность приведенных классических предпосылок. Независимость результатов измерений обычно принимается «из общих предположений», между тем во многих случаях очевидна их коррелированность. Одинаковая распределенность также вызывает сомнения из-за изменения во времени свойств измеряемых образцов, средств измерения и психофизического состояния специалистов, проводящих измерения (испытания, анализы, опыты). Даже обоснованность самого применения вероятностных моделей иногда вызывает сомнения, например, при моделировании уникальных измерений (согласно классическим воззрениям, теорию вероятностей обычно привлекают при изучении массовых явлений). И уж совсем редко распределения результатов измерений можно считать нормальными.

Итак, методы классической математической статистики обычно используют вне сферы их обоснованной применимости. Какова влияние отклонений от традиционных предпосылок на статистические выводы? В настоящее время об этом имеются лишь отрывочные сведения. Так, три примера в статье [6] показывают весь спектр возможных свойств классических расчетных методов в случае отклонения от нормальности. Так, методы построения доверительного интервала для математического ожидания оказываются вполне пригодными при таких отклонениях. Методы проверки однородности двух независимых выборок с помощью двухвыборочного критерия Стьюдента пригодны в некоторых случаях. В задаче отбраковки (исключения) резко выделяющихся наблюдений (выбросов) расчетные методы, основанные на нормальности, оказались полностью непригодными.

Очевидно, имеется ***необходимость изучения свойств расчетных методов классической математической статистики, опирающихся на предположение нормальности, в ситуациях, когда это предположение не выполнено.*** Аппаратом для такого изучения наряду с методом Монте-Карло могут послужить предельные теоремы теории вероятностей, прежде всего Центральная Предельная Теорема, поскольку интересующие нас расчетные методы обычно используют разнообразные суммы. Пока подобное изучение не проведено, остается неясной научная ценность, например, применения основанного на предположении многомерной нормальности факторного анализа к векторам из переменных, принимающих небольшое число градаций и к тому же измеренных в порядковой шкале.

Нерешенным проблемам статистики посвящены статьи [132, 133]. Одна из важных проблем - использование асимптотических результатов при конечных объемах выборок. Конечно, естественно изучить свойства алгоритма с помощью метода Монте-Карло. Однако из какого конкретного распределения брать выборки при моделировании? От выбора распределения зависит результат. Кроме того, датчики псевдослучайных чисел лишь имитируют случайность. До сих пор неизвестно, каким датчиком целесообразно пользоваться в случае возможного безграничного роста размерности пространства.

Другая проблема – обоснование выбор одного из многих критериев для проверки конкретной гипотезы. Например, для проверки однородности двух независимых выборок можно использовать критерии Стьюдента, Крамера-Уэлча, Лорда, хи-квадрат, Вилкоксона (Манна-Уитни), Ван-дер-Вардена, Сэвиджа, Н.В. Смирнова, типа омега-квадрат (Лемана-Розенблатта), Реньи, Г.В.Мартынова и др. Какой выбрать?

Критерии однородности проанализированы в [65]. Естественных подходов к сравнению критериев несколько - на основе асимптотической относительной эффективности по Бахадуру, Ходжесу-Леману, Питмену. И каждый из перечисленных критериев является оптимальным при соответствующей альтернативе или подходящем распределении на множестве альтернатив. При этом математические выкладки обычно используют альтернативу сдвига, сравнительно редко встречающуюся в практике анализа реальных статистических данных. Итог печален - блестящая математическая техника, продемонстрированная в [65], не позволяет дать рекомендации для выбора критерия проверки однородности при анализе реальных данных.

Проблемы разработки высоких статистических технологий поставлены в [61] (см. также одноименный сайт http://orlovs.pp.ru). Используемые при обработке реальных данных статистические технологии состоят из последовательности операций, каждая из которых, как правило, хорошо изучена, поскольку сводится к оцениванию (параметров,

характеристик, распределений) или проверке той или иной гипотезы. Однако статистические свойства результатов обработки, полученных в результате последовательного применения таких операций, мало изучены [66]. Необходима теория, позволяющая изучать свойства статистических технологий и так их конструировать, чтобы обеспечить высокое качество обработки данных.

В заключение отметим, что развернутое описание статистики нечисловых данных дано в монографиях [7, 14, 18]. Современное состояние отражено в обзорах [67, 68]. При дальнейшем развитии исследований важно опираться на современную методологию статистических методов [69].

ГЛАВА 3. АСИМПТОТИКА ОЦЕНОК ПЛОТНОСТИ РАСПРЕДЕЛЕНИЯ ВЕРОЯТНОСТЕЙ

Согласно новой парадигме прикладной математической статистики [1], сердцевиной этой научной области является статистическая теория в пространствах произвольной природы. Эти пространства не предполагаются линейными. Подходы и результаты статистики в пространствах произвольной природы могут применяться могут применяться как при анализе числовых данных, так и нечисловых (бинарных отношений, множеств и др.). Статистическая теория в пространствах произвольной природы и статистические методы анализа конкретных нечисловых данных выделены в 1979 г. как самостоятельная область прикладной математической статистики [2, 3]. Она была названа статистикой объектов нечисловой природы. Позже ее стали называть статистикой нечисловых данных или нечисловой статистикой [4, 5].

Непараметрические оценки плотности распределения вероятностей в пространствах произвольной природы - один из основных инструментов нечисловой статистики. Систематическое изложение теории таких оценок начато в статьях [6 - 11], непосредственным продолжением которых является настоящая статья. Регулярно используются ссылки на условия и утверждения из статей [7, 9, 10].

Введем обозначения. Пусть (Z, A) – измеримое пространство, p и q – сигма-конечные меры на (Z, A), причем p абсолютно непрерывна относительно q, т.е. из $q(B) = 0$ следует $p(B) = 0$ для любого множества B из сигма-алгебры A. В этом случае на (Z, A) существует неотрицательная измеримая функция $f(x)$ такая, что

$$q(C) = \int_C f(x)\, p(dx) \tag{1}$$

для любого множества C из сигма-алгебры измеримых множеств A. Функция $f(x)$ называется производной Радона - Никодима меры q по мере

p, а в случае, когда q - вероятностная мера, также плотностью вероятности q по отношению к мере p [12, с.460].

Пусть X_1, X_2 ,…, X_n – независимые одинаково распределенные случайные элементы (величины), распределение которых задается вероятностной мерой q. В статьях [6, 7] введено несколько видов непараметрических оценок плотности вероятности q по выборке X_1, X_2 ,…, X_n. Подробнее изучены линейные оценки. В статьях [8, 9] рассмотрены их частные случаи – ядерные оценки плотности в пространствах произвольной природы. В статьях [10, 11] асимптотическая теория ядерных оценок плотности развита, прежде всего, для нужд статистики конкретных видов объектов нечисловой природы, в которой основной интерес представляют конечные пространства Z. Мера p при этом не непрерывная, а дискретная, например, считающая. Таким образом, в рамках единого подхода удается рассмотреть оценки плотностей и оценки вероятностей.

В предположении непрерывности неизвестной плотности $f(x)$ представляется целесообразным «размазать» каждый атом эмпирической меры, т.е. рассмотреть линейные оценки, введенные в нашей первой работе по нечисловой статистике [3, с.24]:

$$f_n(x) = \frac{1}{n} \sum_{1 \le i \le n} g_n(x, X_i), \quad g_n : Z^2 \to R^1, \tag{2}$$

в которых действительнозначные функции g_n удовлетворяют некоторым условиям регулярности.

Пусть d – показатель различия (синоним - мера близости) на Z [4] (в наиболее важных частных случаях – метрика на Z). В [13] нами введены ядерные оценки плотности – оценки вида (2) с

$$g_n(x, X_i) = \frac{1}{b(h_n, x)} K\left(\frac{d(x, X_i)}{h_n} \right), \quad K : [0, +\infty) \to R^1, \tag{3}$$

где $K = K(u)$ – ядро (ядерная функция), h_n – последовательность положительных чисел (показателей размытости), $b(h_n, x)$ – нормировочный множитель. В [14] линейные оценки (2) с функциями g_n из (3) названы нами «обобщенными оценками типа Парзена-Розенблатта», т.к. в частном случае $Z = R^1$, $d(x, X_i) = | x - X_i |$, $b(h_n, x) = h_n$ они переходят в известные оценки, введенные Розенблаттом [15] и Парзеном [16].

Цель настоящей главы - завершение цикла работ, начатого статьями [6 - 11], посвященного математическому изучению асимптотических свойств различных видов непараметрических оценок плотности распределения вероятности в пространствах общей природы. Тем самым подводится математический фундамент под применения таких оценок в нечисловой статистике (см. [3 - 5, 13, 14] и другие работы, в которых доказательства предельных свойств рассматриваемых оценок были опущены).

Начинаем с рассмотрения среднего квадрата ошибки ядерной оценки плотности

$$\alpha_n = M(f_n(x) - f(x))^2 = (Mf_n(x) - f(x))^2 + Df_n(x) \quad (4)$$

и выбора последовательности h_n с целью максимизации порядка убывания α_n при $n \to \infty$ (здесь и далее M - символ математического ожидания, D - символ дисперсии). Затем рассматриваем другие виды непараметрических оценок плотности - гистограммные оценки и оценки типа Фикс-Ходжеса. Затем изучаем непараметрические оценки регрессии и их применение для решения задач дискриминантного анализа в пространстве общей природы.

3.1. Круговая функция распределения

Пусть справедливы следующие условия регулярности:

R1) плотность $f(x)$ непрерывна в точке x, в которой оцениваем плотность;

R2) мера p согласована с показателем различия d, т.е. мера шара радиуса t равна t,

$$p\{y: d(x, y) < t\} = t, \qquad 0 \le t \le t_1 = p(Z), \quad (5)$$

другими словами, показатель различия является предпочтительным [7],

R3) ядро $K(u)$ - непрерывная финитная функция, $K(u) = 0$ при $u > E$, такая, что

$$\int_0^E K(u)du = \int_0^\infty K(u)du = 1. \quad (6)$$

Введем в рассмотрение шары $L_t(x) = \{y: d(x, y) < t\}$ радиуса t и аналог функции распределения случайной величины X со значениями в Z с плотностью $f(x)$:

$$G(x,t) = P\{X \in L_t(x)\} = \int_{L_t(x)} f(y)p(dy). \quad (7)$$

Назовем $G(x, t)$ круговой функцией распределения в точке x. Все нужные в дальнейшем свойства вероятностной модели, как будет показано, выражаются с помощью круговой функции распределения.

Из непрерывности плотности в точке x и равенства (5) следует, что круговая функция распределения $G(x, t)$ дифференцируема по t при $t = 0$ и

$$G'_t(x,0) = f(x). \quad (8)$$

Изучим смещение оценки $f_n(x)$. Имеем цепочку равенств

$$Mf_n(x) = \frac{1}{h_n}\int_Z K\left(\frac{d(x,y)}{h_n}\right) f(y)p(dy) = \frac{1}{h_n}\int_0^\infty K\left(\frac{t}{h_n}\right)dG(x,t) = \int_0^E K(u)\frac{dG(x,h_n u)}{h_n}. \quad (9)$$

Пусть справедливо следующее условие.

*R*4) В некоторой окрестности точки $t = 0$ (т.е. при $0 \le t \le t_0$ при некотором t_0) существует производная по t круговой функцией распределения $G(x, t)$, т.е. $G'_t(x,t) = g(x,t)$.

Тогда при $h_n E < t_0$ имеем

$$Mf_n(x) = \int_0^E K(u) g(x, h_n u) du \quad . \tag{10}$$

Основная идея дальнейших рассуждений состоит в том, чтобы для изучения скорости сходимости ядерных оценок применить разложение $g(x, t)$ в ряд по степеням t в окрестности $t = 0$ (в предположении существования указанного разложения).

R5) Пусть справедливо разложение

$$g(x,t) = g(x,0) + t g'_t(x,0) + \frac{t^2}{2} g''_{tt}(x,0) + o(h_n^2), \ 0 \le t \le h_n E \tag{11}$$

(согласно (8) $g(x, 0) = f(x)$). Подставим это разложение в (10), получим с учетом (6) и непрерывности функции $K(u)$:

$$Mf_n(x) = f(x) + h_n g'_t(x,0) \int_0^E uK(u) du + h_n^2 g''_{tt}(x,0) \int_0^E u^2 K(u) du + o(h_n^2) \quad . \tag{12}$$

3.2. Первые оценки скорости сходимости

Согласно теореме 7 статьи [9] при справедливости рассматриваемых условий

$$Df_n(x) = \frac{f(x)}{nh_n} \int_0^E K^2(u) du + o\left(\frac{1}{nh_n}\right) . \tag{13}$$

Если

$$a = g'_t(x,0) \int_0^E uK(u) du \neq 0 \quad , \tag{14}$$

то средний квадрат ошибки ядерной оценки плотности (4) согласно (12) и (13) равен

$$\alpha_n = h_n^2 a^2 + \frac{f(x)}{nh_n} \int_0^E K^2(u) du + o\left(h_n^2 + \frac{1}{nh_n}\right) . \tag{15}$$

Следовательно, при $f(x) \neq 0$ оптимальное по порядку скорости сходимости значение h_n определяется из условия "уравнивания погрешностей" [2]

$$h_n^2 = \frac{1}{nh_n}, \ h_n = n^{-1/3} \quad . \tag{16}$$

Тогда, как легко видеть, средний квадрат ошибки ядерной оценки плотности имеет порядок "n в степени (-2/3)":

$$\alpha_n \cong C n^{-2/3} \tag{17}$$

при некоторой константе C.

Обоснуем сказанное более подробно. С точностью до бесконечно малых более высокого порядка средний квадрат ошибки ядерной оценки плотности равен

$$B(h) = a^2h^2 + \frac{F}{nh},\ F = f(x)\int_0^E K^2(u)du,\ h = h_n \quad . \tag{18}$$

С целью решения задачи оптимизации

$$B(h) \to \max$$

вычислим производную $B(h)$ по h и приравняем ее 0:

$$B'(h) = 2a^2h - \frac{F}{nh^2} = 0 \quad . \tag{19}$$

Решая уравнение (19) относительно h, получаем, что

$$h = h_n = \left(\frac{F}{2a^2n}\right)^{\frac{1}{3}} = \left(\frac{F}{2a^2}\right)^{\frac{1}{3}} n^{-1/3} \quad . \tag{20}$$

Соотношение (20) уточняет ранее полученное соотношение (17).

Для частного случая $Z = R^1$, т.е. для оценок Парзена-Розенблатта, соотношения (16) - (17) известны (см. [17, с.315]).

Если соотношение (14) не выполнено, т.е. $a = 0$, то согласно (12) заключаем, что

$$\alpha_n = [g''_{tt}(x,0)]^2\left[\int_0^E u^2K(u)du\right]^2 + \frac{f(x)}{nh_n}\int_0^E K^2(u)du + o\left(h_n^4 + \frac{1}{nh_n}\right) \quad . \tag{21}$$

Оптимальное по порядку сходимости h_n определяется из условия

$$h_n^4 = \frac{1}{nh_n},\ h_n = n^{-1/5} \quad , \tag{22}$$

и при этом

$$\alpha_n \cong C_1n^{-4/5} \quad . \tag{23}$$

при некоторой константе C_1.

Соотношения (22) - (23) совпадают с известными результатами для весьма частного случая $Z = R^1$, т.е. для оценок Парзена-Розенблатта (см. [17, с.316]).

3.3. Примеры ядерных оценок

Из сравнения формул (17) и (23) ясно, что сходимость убыстряется при $a = 0$, где a определено в (14). Поскольку a - произведение двух сомножителей, то $a = 0$ тогда и только тогда, когда $g'_t(x,0) = 0$ или

$$\int_0^E uK(u)du = 0 \quad . \tag{24}$$

Роль сомножителей разная: первый определяется свойствами пространства с мерой, показателя различия (другими словами, меры близости) и

плотности распределения случайной величины, а второй - свойствами ядра.

С целью выявления свойств первого сомножителя рассмотрим два примера.

Пример 1. Рассмотрим множество действительных чисел $Z = R^1$. Пусть p - мера Лебега, d - расстояние Евклида (показатель различия), $F(x)$ - функция распределения случайной величины X, причем ее плотность f (по мере Лебега, т.е. в обычном смысле) дважды непрерывно дифференцируема. Тогда

$$G(x,t) = P\left\{X \in \left(x - \frac{t}{2}, x + \frac{t}{2}\right)\right\} = F\left(x + \frac{t}{2}\right) - F\left(x - \frac{t}{2}\right), \tag{25}$$

а потому

$$g(x,t) = \frac{1}{2}\left\{f\left(x + \frac{t}{2}\right) + f\left(x - \frac{t}{2}\right)\right\}. \tag{26}$$

Продифференцируем (26) по t:

$$g'_t(x,t) = \frac{1}{4}\left\{f'\left(x + \frac{t}{2}\right) - f'\left(x - \frac{t}{2}\right)\right\}. \tag{27}$$

При $t = 0$ при всех x

$$g'_t(x,0) = 0. \tag{28}$$

Пример 2. Рассмотрим Z = [0, $+\infty$). Пусть мера p и расстояние d получены из рассмотренных в примере 1 сужением на [0, $+\infty$). Пусть функция распределения и плотность обладают теми же свойствами, что и в примере 1. Тогда

$$G(0,t) = F(t),\ g(0,t) = f(t),\ g'_t(0,t) = f'(t). \tag{29}$$

Следовательно, первый сомножитель в (14), вообще говоря, отличен от 0.

Можно показать, что для конечномерного пространства $Z = R^k$, меры Лебега p, евклидова расстояния d и дважды дифференцируемой плотности f первый сомножитель в (14) обращается в 0, а для мер p, отличных от Лебеговой, вообще говоря, не обращается (при сохранении прочих перечисленных в примерах 1 и 2 условий).

Из сказанного следует, что для введенных нами оценок (2) - (3) в случае конечномерного пространства $Z = R^k$, меры Лебега p, евклидова расстояния d и дважды дифференцируемой плотности f скорость сходимости задается формулой (23), а для классических оценок Парзена-Розенблатта - формулой (17) (см. также [17, с.315-316]), т.е. введенные нами оценки сходятся гораздо быстрее, чем оценки Парзена-Розенблатта.

3.4. Улучшение скорости сходимости ядерных оценок

Поскольку статистик может сам выбирать ядро, то для повышения скорости сходимости целесообразно принять условие (24). Однако

скорость сходимости можно еще более повысить за счет выбора ядра из более узкого класса.

При более высокой гладкости круговой плотности $g(x, t)$ можно получить более высокую скорость сходимости среднего квадрата ошибки ядерной оценки плотности α_n, соответствующим образом выбирая ядро $K(u)$. Предположим, что круговая плотность $g(x,t)$ допускает разложение

$$g(x,t) = f(x) + tg'_t(x,0) + \frac{t^2}{2} g''_{tt}(x,0) + \frac{t^3}{3!} g'''_{ttt}(x,0) + ... + \frac{t^k}{k!} g^{(k)}_{t^{(k)}}(x,0) + o(h_n^k), \quad (30)$$

причем остаточный член равномерно ограничен на отрезке $[0, h_nE]$. Тогда

$$Mf_n(x) = f(x) + h_n g'_t(x,0)\int_0^E uK(u)du + \frac{h_n^2}{2} g''_{tt}(x,0)\int_0^E u^2K(u)du +$$

$$+ \frac{h_n^3}{3!} g'''_{ttt}(x,0)\int_0^E u^3K(u)du + ... + \frac{h_n^k}{k!} g^{(k)}_{t^{(k)}}(x,0)\int_0^E u^kK(u)du + \theta_n h_n^k, \quad (31)$$

где $\theta_n \to 0$ при $n \to \infty$.

Пусть теперь

$$\int_0^E u^iK(u)du = 0, \quad i = 1,2,...,k-1. \quad (32)$$

Тогда

$$Mf_n(x) - f(x) = \frac{h_n^k}{k!} g^{(k)}_{t^{(k)}}(x,0)\int_0^E u^kK(u)du + o(h_n^k) \quad (33)$$

Следовательно,

$$\alpha_n = h_n^{2k}\left(\frac{1}{k!} g^{(k)}_{t^{(k)}}(x,0)\int_0^E u^kK(u)du\right)^2 + \frac{f(x)}{nh_n}\int_0^E K^2(u)du + o\left(h_n^{2k} + \frac{1}{nh_n}\right) \cong Ah_n^{2k} + \frac{B}{nh_n} \quad (34)$$

при соответствующих A и B. Оптимальная по порядку скорость сходимости будет при

$$h_n^{2k} = \frac{1}{nh_n}, \quad h_n = n^{-1/2k+1} \quad (35)$$

(в предположении $A \neq 0,\ B \neq 0$). При этом

$$\alpha_n \cong n^{-2k/2k+1} = n^{\left(-1+1/2k+1\right)}. \quad (36)$$

Этот результат - продвинутое обобщение теоремы 4.1 в книге И.А. Ибрагимова и Р.З. Хасьминского [17, с.316], относящейся к оцениванию плотности одномерной случайной величины. Порядок сходимости в общем случае тот же, что и в указанной теореме, но условия наложены не на плотность $f(x)$, а на круговую плотность $g(x, t)$. Это существенно в случае $Z \neq R^k$, т.к. в R^k имеются традиционно выделенные мера (Лебега) и расстояние (Евклида).

С прикладной точки зрения предположение (30) о гладкости круговой плотности представляется достаточно естественным. Напомним,

что вплоть до XIX в. математики практически не делали различия между непрерывными, дифференцируемыми и аналитическими функциями. Инженеры не делают такого различия и сейчас. Отсюда методологическое предложение: в прикладных задачах допустимо использовать математические модели с той степенью гладкости рассматриваемых функций, которая позволяет наиболее легко обосновать алгоритмы расчетов, разумеется, если эта степень гладкости не противоречит фактам соответствующей предметной области. Другой пример подобного методологического подхода: шкала любого прибора конечна, поэтому результаты первичных измерений целесообразно моделировать с помощью финитных случайных величин; такие величины имеют все моменты, а это существенно облегчает получение для них предельных теорем. Методологическим вопросам посвящены наши работы [18, 19].

Из соотношения (36) следует, что для любого $\varepsilon > 0$ существует ядро $K(u)$ такое, что для соответствующей оценки

$$\alpha_n = O\left(n^{-1+\varepsilon}\right), \tag{37}$$

а также

$$\lim_{n\to\infty} n^{+1/2-\varepsilon}\,(f_n(x) = f(x)) = 0 \tag{38}$$

по вероятности. Хотя при любом k множество ядер $K(u)$, удовлетворяющих соотношениям (32), бесконечно, не существует ядра, удовлетворяющего этим соотношениям сразу при всех k.

Перенос полученных результатов на случай конечных пространств объектов нечисловой природы осуществляется тем же путем, что и в статьях [10, 11]. Грубо говоря, необходимо, чтобы

$$\alpha_{mn}(g) = o(h_n^k), \tag{39}$$

где m - параметр дискретности (см. [10]). Поскольку принципиальных трудностей, как ясно из рассуждений статьи [10], в указанном переносе нет, мы его здесь не приводим.

3.5. Гистограммные оценки

Развитие теории гистограммных оценок облегчается тем, что гистограмме в произвольном пространстве можно поставить в соответствие гистограмму одномерной случайной величины, соотнеся область в произвольном пространстве Z и отрезок той же меры и той же вероятности попадания в него. Так, теорема 1 из основополагающей статьи Н.В. Смирнова [20] может быть перенесена на случай произвольного пространства Z следующим образом (теоремы 1 - 2).

Теорема 1. *Пусть* α, β *и* γ *- положительные константы. При каждом* $n = 1, 2, \ldots$ *рассмотрим* $k(n)$ *положительных чисел* $p_1(n), p_2(n), \ldots, p_{k(n)}(n)$ *таких, что*

$$p_1(n) + p_2(n) + \ldots + p_{k(n)}(n) = 1 - \alpha < 1, \tag{40}$$

$$\frac{\beta}{k(n)} \le p_i(n) \le \frac{\gamma}{k(n)},\ i = 1,2,...,k(n) \tag{41}$$

Пусть проводится n независимых мультиномиальных испытаний с k(n)+1 исходами, имеющими вероятности $p_1(n)$, $p_2(n)$, ..., $p_{k(n)}(n)$, $p_{k(n)+1}(n) = \alpha$ *соответственно. Обозначим* $m_1(n)$, $m_2(n)$, ..., $m_{k(n)}(n)$, $m_{k(n)+1}(n)$ *количество осуществлений каждого из исходов. Положим*

$$M_n = \max\left(\frac{|m_i(n) - np_i(n)}{\sqrt{np_i(n)}},\ i = 1,2,...,k(n) \right). \tag{42}$$

Пусть $k(n) \to \infty$ *и*

$$\overline{\lim_{n\to\infty}}\left(\frac{k^3(n)(\ln k(n))^3}{n} \right) < \infty . \tag{43}$$

Тогда при любом действительном λ

$$\lim_{n\leftarrow\infty} P\left\{ M_n < m(k(n)) + \frac{\lambda}{m(k(n))} \right\} = \exp\{-2\exp(-\lambda)\}, \tag{44}$$

где $m(k(n))$ - *решение уравнения*

$$\frac{1}{\sqrt{2\pi}} \int_{m(k(n))}^{\infty} \exp\left\{ -\frac{1}{2}x^2 \right\} dx = \frac{1}{k(n)} . \tag{45}$$

В постановке Н.В. Смирнова каждый из упомянутых в теореме 1 исходов состоит в попадании в один из интервалов равной длины, на которые разбит отрезок для построения гистограммы. Условие (40) - это условие (В) Н.В. Смирнова, левое ограничение в (41) - условие (А) Н.В. Смирнова, правое - вытекает из непрерывности оцениваемой плотности. Конечно, сказанное не дает гарантии, что любой последовательности мультиномиальных распределений, удовлетворяющей перечисленным в теореме 1 условиям, можно поставить в соответствие задачу оценивания плотности с помощью гистограммы, т.к. последняя предполагает вполне определенную согласованность указанных мультиномиальных распределений между собой. Однако анализ рассуждений Н.В. Смирнова показывает, что им фактически доказана именно сформулированная выше теорема 1, а результаты об оценивании одномерной плотности с помощью гистограммы можно рассматривать как следствия из этой теоремы.

В начале 70-х годов Э.А. Надарая показал [21, 22], что условие (40) можно отбросить (для дальнейшего удобнее принять, что в (40) можно положить $\alpha = 0$), а условие (43) заменить на более слабое

$$\frac{k(n)(\ln k(n))^3}{n} \to 0 \tag{46}$$

(см. также монографию Г.М. Мания [23, с.88]). Известно асимптотическое разложение левой части (44) по степеням $ln(k(n))$, найдены другие аппроксимирующие выражения, более быстро сходящиеся, оценена скорость сходимости (см. статью В.Д. Конакова [24]).

Теорема 2. *Пусть* $p(Z) = 1$. *Пусть существуют* $0 < \chi < \nu$ *такие, что для плотности* $f(x)$ *случайного элемента имеем при всех* $x \in Z$

$$\chi \le f(x) \le \nu . \tag{47}$$

Пусть для построения гистограммных оценок используются $k(n)$ *областей равной меры* $X_1^n, X_2^n, ..., X_{k(n)}^n$, *т.е.*

$$X_1^n \cup X_2^n \cup ... \cup X_{k(n)}^n = Z,\ X_i^n \cap X_j^n = \varnothing,\ i \ne j,\ p(X_i^n) = \frac{1}{k(n)} . \tag{48}$$

Положим

$$f^*(x) = \frac{1}{k(n)} \int\limits_{X_i^n} f(y)\, p(dy),\ \ x \in X_i^n\ i = 1,2,...,k(n) . \tag{49}$$

Пусть $k(n) \to \infty$ *и выполнено* (46). *Тогда при любом* λ

$$P\left\{ \max_x \left| \frac{f_n(x) - f^*(x)}{\sqrt{f^*(x)}} \right| < \sqrt{\frac{k(n)}{n}} \left(m(k(n)) + \frac{\lambda}{m(k(n)} \right) \right\} \to \exp\left(-2e^{-\lambda} \right) \tag{50}$$

при $n \to \infty$, *где гистограммная оценка определяется как частный случай линейной оценки*

$$f_n(x) = \frac{1}{n} \sum_{1 \le i \le n} g_n(x, X_i) .$$

(*Гистограммные оценки определяются с помощью последовательности* T_n *разбиений* Z *и функций*

$$g_n(x, X) = \begin{cases} \dfrac{1}{p(A(x))},\ x \in A(x), \\ 0,\ x \notin A(x), \end{cases}$$

где $A(x)$ - *элемент разбиения* T_n, *которому принадлежит* x.)

Теорема 2 непосредственно вытекает из теоремы 1 и цитированных выше результатов Э.А. Надарая.

Теорема 3. *В условиях теоремы 2 для любого* $\varepsilon > 0$

$$\lim_{n \to \infty} P\left\{ \sup_{x \in Z} \left| f_n(x) - f^*(x) \right| > \varepsilon \right\} = 0 . \tag{51}$$

Доказательство. Для корня $m(k) = m(k(n))$, $k = k(n)$ уравнения (45) справедливо асимптотическое равенство (см., например, [20, ф-ла (95)])

$$m(k) = \sqrt{2 \ln k} - \frac{\ln\ln k + \ln(4\pi)}{2\sqrt{2 \ln k}} + O\left(\frac{1}{\ln k} \right) . \tag{52}$$

Из (46) и (52) следует, что

$$\sqrt{\frac{k(n)}{n}} \left(m(k(n)) + \frac{\lambda}{m(k(n))} \right) < (\ln k + \lambda) \sqrt{\frac{k}{n}} < \frac{1}{\sqrt{\ln k}} + \frac{\lambda}{\ln^{3/2} k} \tag{53}$$

при достаточно большом n. Тогда из (47), (49), (50) и (53) следует, что при достаточно большом n и любом фиксированном λ

$$P\left\{\sup_x |f_n(x)-f^*(x)| > \sqrt{\nu}\left[(\ln k)^{-1/2}+\lambda(\ln k)^{-3/2}\right]\right\} < 1-\exp\{-2\exp(-\lambda)\}, \quad (54)$$

откуда и следует (51).

Замечание. Ясно, что можно отказаться от рассмотрения областей равной меры, однако при этом аналоги теорем 2 и 3 будут формулироваться несколько более громоздко. Поскольку принципиально нового при этом не появляется, мы ограничивается сказанным выше.

Чтобы установить равномерную сходимость гистограммных оценок, в силу теоремы 3 достаточно указать условия, при которых

$$\sup_{x\in Z} |f(x)-f^*(x)| \to 0 \quad (55)$$

при измельчении разбиения. Понадобятся некоторые дополнительные результаты.

Гистограммной в общем случае называют функцию $f^*: Z \to R^1$ такую, что

$$f^*(x) = f(x_i^n),\ x \in X_i^n,\ i = 1,2,...,k(n), \quad (56)$$

для некоторых $x_i^n \in X_i^n,\ i = 1,2,...,k(n)$. Если плотность f непрерывна, а X_i^n - связные бикомпакты, то формула (49) дает частный случай (56).

Теорема 4. *Пусть Z счетно компактно, топология в Z порождена естественной мерой близости* [9]. *Тогда Z бикомпактно.*

Пусть f - непрерывная функция, Z - счетно-компактно, топология в Z порождена естественной мерой близости, Тогда условие (55) выполнено.

Теорема 5. *Пусть Z - счетно-компактно, топология в Z порождена естественной мерой близости, мера p безатомная, плотность f непрерывна и положительна. Тогда существует последовательность разбиений такая, что для любого* $\varepsilon > 0$

$$\lim_{n\to\infty} P\left\{\sup_{x\in Z} |f(x)-f^*(x)| > \varepsilon\right\} = 0, \quad (57)$$

где $f_n(x)$ - *гистограммная оценка плотности, причем указанная последовательность разбиений не зависит от плотности f* (*одна и та же для всех положительных непрерывных плотностей f*).

При этом мера p называется безатомной, если для любого измеримого подмножества A с $p(A) > 0$ и любого числа $0 < \delta < p(A)$ найдется измеримое подмножество $B \subset A$ такое, что $p(B) = \delta$.

Замечание. Теорема 5 относится к случаю бикомпактного Z. Если Z не таково, то равномерная сходимость имеет место для бикомпактных подмножеств Z, что устанавливается с помощью теоремы 1 (без модификации Э.А. Надарая).

Доказательство теоремы 5. Их теоремы 4 следует, что выполнено следующее условие:

Условие А. Существует последовательность разбиений, измельчающаяся и такая, что для любой точки $x \in Z$ и любой ее окрестности $U(x)$ найдется такое разбиение из этой последовательности, что его область, содержащая x, полностью содержится в $U(x)$.

Будем рассматривать последовательность разбиений, указанную в этом условии. Ясно, что в силу безатомности p можно считать, что отношения мер одного разбиения равномерно отделены от 0 и ∞. В силу теоремы 4 Z бикомпактно, а потому плотность f достигает своих минимального и максимального значений. Поскольку f всюду положительна, то минимальное значение также положительно, т.е. выполнены неравенства (8). Число областей в разбиениях искомой последовательности будем регулировать так, чтобы выполнялось (46) (для этого достаточно нужное число раз повторить одно и то же разбиение из исходной последовательности, рассмотренной выше в связи с условием А). Тогда по теореме 3 справедливо (51) (в связи с заменой условия равной меры областей разбиений на условие отделенности отношений мер областей от 0 и ∞ необходимо использовать замечание после теоремы 3). По теореме 4 справедливо (55). Из (51) и (55) вытекает (57). Теорема 5 доказана.

3.6. Оценки типа Фикс-Ходжеса

Обобщенные оценки типа Фикс-Ходжеса [7] определяются с помощью расширяющейся последовательности множеств $U(x, r)$. Ограничимся частным случаем

$$U(x, r) = L_r(x),\ p(L_r(x)) = r,\ 0 \le r \le r_0. \tag{58}$$

Некоторое обобщение дают ядерные оценки со случайными $h_n = h_n(x, \omega)$, имеющие вид

$$f_n(x) = \frac{1}{n h_n} \sum_{1 \le i \le n} K\left(\frac{d(x, X_i)}{h_n(x; X_1, X_2, ..., X_n)} \right), \tag{59}$$

где

$$h_n = h_n(x; X_1, X_2, ..., X_n) = \inf\left\{ r : \sum_{1 \le i \le n} \chi(X_i \in L_r(x)) \ge k_n \right\}, \tag{60}$$

где $\chi(C) = 1$, если условие С выполнено, и $\chi(C) = 0$ в противном случае.

Если

$$K(u) = \begin{cases} 1, & 0 \le u \le 1, \\ 0, & u > 1, \end{cases} \tag{61}$$

то оценка, задаваемая соотношениями (2) - (3), является обобщенной оценкой типа Фикс-Ходжеса [7]. Таким образом, ядерные оценки и оценки типа Фикс-Ходжеса и оценки типа Фикс-Ходжеса можно рассматривать в рамках одной и той же схемы. Однако в силу (60) оценка (59) не является

суммой независимых одинаково распределенных случайных величин, что затрудняет ее изучение.

Рассмотрим распределение случайной величины $h_n(\omega)$ из (60). Ясно, что h_n является k_n-ой порядковой статистикой совокупности $\{d(x, X_i), i = 1, 2, ..., n\}$, а потому при естественных предположениях имеет асимптотически нормальное распределение. Укажем эти предположения.

Функцией распределения случайных величин $\eta_i = d(x, X_i),\ i = 1,2,...$ является круговая функция распределения

$$G(x,t) = P\{X \in L_t(x)\} = \int_{L_t(x)} f(y)\,p(dy) \quad .$$

Предположим, что она непрерывна и строго возрастает. Имеем

$$P\{G(x,h_n) \le y\} = P\{h_n \le G^{-1}(x,y)\} = P\left\{\sum_{1\le i\le n} \chi(\eta_i \in [0; G^{-1}(x,e)] \ge k_n\right\} . \qquad (62)$$

Справа в (5) стоит вероятность того, что не менее k_n успехов имело быть в n испытаниях Бернулли с вероятностью успеха $p = G(x, G^{-1}(x, y)) = y$ в каждом.

Рассмотрим последовательность $y = y_n$, $n = 1, 2, ...,$ такую, что

$$ny_n \to \infty . \qquad (63)$$

Тогда к правой части (62) можно применить центральную предельную теорему (см., например, [25, с.121]), т.е.

$$\lim_{n\to\infty} \max_k \left| P\left\{\sum_{1\le i\le n} \chi(\eta_i \in [0; G^{-1}(x,y_n)] \ge k\right\} - \Phi\left(\frac{ny_n - k}{\sqrt{ny_n(1-y_n)}}\right)\right| = 0 . \qquad (64)$$

Дополнительно к (63) предположим, что y_n меняется так, что при некоторых (произвольных, но фиксированных) a и b

$$a < \frac{ny_n - k_n}{\sqrt{ny_n(1-y_n)}} < b . \qquad (65)$$

Ясно, что с учетом (63) для этого необходимо выполнения условия

(I) $k_n \to \infty$ при $n \to \infty$.

Это условие соответствует условию "$nh_n \to \infty$ при $n \to \infty$" для ядерных оценок [9].

Поскольку из (65) вытекает, что

$$\left| y_n - \frac{k_n}{n} \right| \le \frac{\max(|a|, |b|)}{2\sqrt{n}} , \qquad (66)$$

то в (64) можно заменить $ny_n(1 - y_n)$ на $k_n(1 - k_n/n)$, т.е. можно переписать (64) в виде

$$\lim_{n\to\infty}\sup_{w\in[a;b]}\left|P\left\{\frac{nG(x,h_n)-k_n}{\sqrt{k_n\left(1-\frac{k_n}{n}\right)}}\le w\right\}-\Phi(w)\right|=0 \quad . \tag{67}$$

Таким образом. величина $G(x,h_n(\omega))$ является асимптотически нормальной случайной величиной с параметрами

$$MG(x,h_n(\omega))\approx\frac{k_n}{n},\quad DG(x,h_n(\omega))\approx\frac{1}{n}\frac{k_n}{n}\left(1-\frac{k_n}{n}\right). \tag{68}$$

Теорема 6. *Пусть мера p и метрика (мера близости, показатель различия) d связаны соотношением* (58). *Пусть плотность f(x) непрерывна и ограничена на Z. Пусть ядро K(u) таково, что*

$$\int_0^\infty K(u)du=1,\quad \int_0^\infty |K(u)|\,du<\infty \quad . \tag{69}$$

Пусть последовательность k_n удовлетворяет условиям

$$k_n\to\infty,\ \frac{k_n}{n}\to 0 \tag{70}$$

при $n\to\infty$. *Тогда обобщенная оценка типа Фикс-Ходжеса* (59) - (60) *является асимптотически несмещенной оценкой плотности f(x).*

Пусть, кроме того,

$$\lim_{a\to\infty}\int_0^a K^2(u)du<\infty \quad . \tag{71}$$

Тогда

$$\lim_{n\to\infty}Df_n(x)=0 \tag{72}$$

и $f_n(x)$ из (59) - (60) *является состоятельной оценкой плотности f*(x).

Доказательство вытекает фактически из того, что в силу приведенных выше рассуждений $h_n(x,\omega)\to 0$ (по вероятности) при $n\to\infty$. Можно фиксировать последовательность {h_n, n = 1, 2, ...} и при этом условии повторить рассуждения, проведенные при доказательстве соответствующих теорем для ядерных оценок [9], поскольку эти рассуждения - аналитические, а не вероятностные. Для получения (72) необходимо учесть зависимость слагаемых в (59), что делается стандартным образом.

3.7. Непараметрические оценки регрессии

Начнем с рассмотрения условных плотностей. Пусть пространство Z есть прямое произведение двух пространств Z_1 и Z_2, т.е. $Z=Z_1\times Z_2$, а мера p в Z есть прямое произведение мер p_1 в Z_1 и p_2 в Z_2, т.е. $p=p_1\times p_2$. Тогда элемент $x\in Z$ будем записывать в виде x = (x_1, x_2), где $x_1\in Z_1$ и $x_2\in Z_2$.

Пусть $X = (X_1, X_2)$ - случайная величина со значениями в Z, где $X_1 \in Z_1$ и $X_2 \in Z_2$, а (X_{11}, X_{12}), (X_{21}, X_{22}), ..., (X_{n1}, X_{n2}) - выборка из распределения, соответствующего $X = (X_1, X_2)$.

Как известно [26, с.145-146], условная плотность распределения X_1 при фиксированном значении $X_2 = x_2$ имеет вид

$$f(x_1 \mid X_2 = x_2) = f(x_1 \mid x_2) = \frac{f(x_1, x_2)}{\int\limits_{Z_1} f(x_1, x_2) p_1(dx_1)}, \quad x_1 \in Z_1, x_2 \in Z_2 \quad , \tag{73}$$

где $f(x_1, x_2)$ - плотность случайного элемента (X_1, X_2) в пространстве $Z = Z_1 \times Z_2$, а знаменатель в (73) отличен от 0.

Для оценки условной плотности (73) представляется естественным заменить совместную плотность $f(x_1, x_2)$ в (73) на ее непараметрическую оценку, т.е. в качестве оценки условной плотности $f(x_1|x_2)$ применять

$$f_n(x_1 \mid x_2) = \frac{f_n(x_1, x_2)}{\int\limits_{Z_1} f_n(x_1, x_2) p_1(dx_1)} \quad , \tag{74}$$

где $f_n(x_1, x_2)$ - оценка совместной плотности $f(x_1, x_2)$.

В [7, 9, 10] и выше в настоящей главе приведен ряд условий, при которых различные оценки плотности вероятности являются состоятельными, т.е. при $n \to \infty$

$$f_n(x_1, x_2) \to f(x_1, x_2) \tag{75}$$

(сходимость по вероятности). Когда оценка условной плотности является состоятельной? Из (73) и (74) ясно, что при справедливости (75) для состоятельности $f_n(x_1|x_2)$ достаточно, чтобы при $n \to \infty$ по вероятности

$$\alpha = \int\limits_{Z_1} (f_n(x_1, x_2) - f(x_1, x_2)) p_1(dx_1) \to 0 \quad . \tag{76}$$

Теорема 7. Пусть $p_1(Z_1) < \infty$,

$$\limsup_{n \to \infty} \sup_{x_1 \in Z_1} | Mf_n(x_1, x_2) - f(x_1, x_2) | = 0 \quad , \tag{77}$$

$$\limsup_{n \to \infty} \sup_{x_1 \in Z_1} Df_n(x_1, x_2) = 0 \quad . \tag{78}$$

Тогда выполнено (*76*).

Доказательство. Имеем

$$\alpha = \int\limits_{Z_1} (f_n(x_1, x_2) - Mf_n(x_1, x_2)) p_1(dx_1) + \int\limits_{Z_1} (Mf_n(x_1, x_2) - f(x_1, x_2)) p_1(dx_1) \quad . \tag{79}$$

В силу (77) и условия $p_1(Z_1) < \infty$ второе слагаемое в (79) стремится к 0 при $n \to \infty$. Вычислим дисперсию α. Положим

$$t(x_1) = f_n(x_1, x_2) - M(f_n(x_1, x_2)) \quad . \tag{80}$$

Воспользуемся соотношением

$$\left[\int_{Z_1} t(x_1)\,p_1(dx_1)\right]^2 = \int_{Z_1} t(x_1)\,p_1(dx_1)\int_{Z_1} t(z)\,p_1(dz) = \int_{Z_1\times Z_1} t(x_1)t(z)\,p_1\times p_1(dx_1\times dz) \quad . \qquad (81)$$

Имеем в силу теоремы Фубини

$$D\alpha = M\left(\int_{Z_1} t(x_1)\,p_1(dx_1)\right)^2 = \int_{Z_1\times Z_1} Mt(x_1)t(z)\,p_1\times p_1(dx_1\times dz) \quad . \qquad (82)$$

По неравенству Коши-Буняковского

$$|\,Mt(x_1)t(z)\,| \le \left\{Df_n(x_1,x_2)Df_n(z,x_2)^{1/2}\right\}. \qquad (83)$$

Из (76) и (83) следует, что при $n\to\infty$

$$D\alpha \to 0. \qquad (84)$$

Из неравенства Чебышёва вытекает, что первое слагаемое в (79) также стремится к 0 при $n\to\infty$ (по вероятности). Следовательно, имеет место соотношение (76), теорема 7 доказана.

Замечание. Соотношения (77) и (78) доказаны для ядерных оценок в [9, 10]. Справедливость этих соотношений для гистограммных оценок вытекает из теорем 2 - 4 раздела 6 "Гистограммные оценки" настоящей статьи. Общие формулировки для линейных оценок вытекают из рассмотрений [7] в предположении, что используемые в неравенствах оценки являются равномерными.

Перейдем к рассмотрению регрессии, т.е. условного математического ожидания. Пусть $h(x, a)$ - мера близости на Z_1. В соответствии с оптимизационным подходом к определению средних величин [4, 27, 28] условным математическим ожиданием (регрессией X_1 на X_2) называется решение оптимизационной задачи

$$M\left(X_1 \mid X_2 = x_2,\ h\right) = Arg\min_{a\in Z_1}\int_{Z_1} h(x_1,a)\,f(x_1\mid x_2)\,p_1(dx_1). \qquad (85)$$

Подставив в (85) вместо плотности ее непараметрическую оценку, получим эмпирическую регрессию (непараметрическую оценку регрессии)

$$M_n\left(X_1 \mid X_2 = x_2,\ h\right) = Arg\min_{a\in Z_1}\int_{Z_1} h(x_1,a)\,f_n(x_1\mid x_2)\,p_1(dx_1). \qquad (86)$$

(В (85) и (86) предполагаем, что знаменатель в (73) отличен от 0.)

Установим, что при $n\to\infty$ эмпирическая регрессия (86) сходится к теоретической регрессии (85):

$$M_n\left(X_1 \mid X_2 = x_2,\ h\right) \to M\left(X_1 \mid X_2 = x_2,\ h\right) \qquad (87)$$

(сходимость понимается так, как при формулировке законов больших чисел [28] и изучении асимптотического поведения решения экстремальных статистических задач [4, 29], поскольку в (85) и (86) определены, вообще говоря, не элементы, а множества).

Как следует из предельной теории решений экстремальных статистических задач, для доказательства состоятельности эмпирической

регрессии как оценки теоретической, т.е. для доказательства (87), достаточно установить равностепенную непрерывность на Z_1 функций

$$q_n(a) = \int_{Z_1} h(x_1, a) f_n(x_1 \mid x_2) p_1(dx_1) \tag{88}$$

и то, что при любом $a \in Z_1$

$$q_n(a) \to q(a) = \int_{Z_1} h(x_1, a) f(x_1 \mid x_2) p_1(dx_1) \tag{89}$$

по вероятности при $n \to \infty$.

Теорема 8. *Пусть* Z_1 *- бикомпакт,* $h(x, a)$ *- непрерывная функция на* $Z_1 \times Z_1$ *и* $f_n(x_1, x_2) \geq 0$ *с вероятностью* 1. *Тогда последовательность функций* $q_n(a)$ *равностепенно непрерывна на* Z_1.

Доказательство. Из условия теоремы 8 следует, что для любого $\varepsilon > 0$ существует разбиение пространства Z_1 такое, что для любых точек a и a' из одного и того же элемента этого разбиения имеем

$$\sup_{x_1 \in Z_1} | h(x_1, a) - h(x_1, a') | < \varepsilon, \tag{90}$$

а тогда

$$| q_n(a) - q_n(a') | = \left| \int_{Z_1} (h(x_1, a) - h(x_1, a')) f_n(x_1 \mid x_2) p_1(dx_1) \right| \leq \varepsilon \int_{Z_1} f_n(x_1 \mid x_2) p_1(dx_1) = \varepsilon, \tag{91}$$

что и доказывает теорему 8.

Теорема 9. *Пусть выполнены условия теоремы* 7, *измеримая мера близости* $h(x, a)$ *ограничена, знаменатель в* (73) *отличен от* 0. *Тогда справедливо* (89).

Доказательство. Достаточно получить (89) для

$$s_n(a) = q_n(a) \int_{Z_1} f_n(x_1, x_2) p_1(dx_1) = \int_{Z_1} h(x_1, a) f_n(x_1, x_2) p_1(dx_1). \tag{92}$$

Доказательство проводится как в теореме 7, отличие только в том, что в аналогах (79) и (82) добавляются множители $h(x_1, a)$ и $h(x_1, a)\, h(z, a)$ соответственно, являющиеся неотрицательными (т.к. $h(x_1, a)$ - мера близости) и ограниченными (в силу условия теоремы).

Соединив условия теорем 8 и 9, получим условия сходимости эмпирической регрессии к теоретической.

Теорема 10. *Пусть* Z_1 *- бикомпакт,* $h(x, a)$ *- непрерывная неотрицательная функция на* $Z_1 \times Z_1$, *для неотрицательной с вероятностью* 1 *оценки плотности* $f_n(x_1, x_2)$ *выполнены условия теоремы* 7 *для всех* $x_2 \in Z_2$ *и знаменатель в* (73) *отличен от* 0 *для всех* $x_2 \in Z_2$. *Тогда для всех* $x_2 \in Z_2$ *справедливо* (87) (*в смысле указанных ранее работ* [4, 28, 29]).

Замечание 1. Условия теоремы 10 могут быть ослаблены. Так, из доказательства теоремы 8 видно, что условие неотрицательности $f_n(x_1, x_2)$ может быть заменено на условие

$$\overline{\lim_{n\to\infty}} \int_{Z_1} | f_n(x_1 | x_2) | p_1(dx_1) < \infty \quad . \tag{93}$$

Это существенно, в частности, для ядерных оценок, поскольку согласно результатам раздела 1 настоящей статьи отказ от неотрицательности ядерных оценок плотности позволяет ускорить их сходимость.

Замечание 2. Теоремы 7 - 10 основаны на равномерной сходимости моментов (77) - (78). Другой ряд теорем с аналогичными заключениями может быть получен в предположении равномерной сходимости оценок плотности. Из равномерной сходимости сходимость моментов, вообще говоря, не следует, хотя для гистограммных оценок имеет место и то, и другое.

3.8. Дискриминантный анализ в пространстве общей природы

Рассмотрим постановку с двумя классами, заданными плотностями $f(x)$ и $g(x)$ соответственно, $x \in Z$. Если f и g известны, то по лемме Неймана-Пирсона область отнесения к первому классу задается неравенством

$$\frac{f(x)}{g(x)} > A \tag{94}$$

где A - некоторая константа (ср. [30]). Если плотности f и g неизвестны, но имеются их состоятельные оценки $f_n(x)$ и $g_m(x)$ по обучающим выборкам объемов n и m соответственно, то область

$$\frac{f_n(x)}{g_m(x)} > A \quad (n \to \infty, m \to \infty) \tag{95}$$

является состоятельной (в соответствующем смысле) оценкой теоретической области (94).

В случае k классов известные постановки ([30], [31]) задачи минимизации средних потерь от принятия ошибочных решений приводят к решениям в терминах плотностей, описывающих классы, априорных вероятностей классов и функции потерь. При этом результаты наблюдений могут лежать в произвольном пространстве (последнее обстоятельство обычно не осознается авторами публикаций по этой тематике), поэтому нам нет необходимости развивать самостоятельную теорию (впрочем, в [2, с.221-223] теория расписана для конечных случайных множеств). При использовании обучающих выборок теоретические плотности заменяются их непараметрическими оценками, рассмотренными выше. Из-за отсутствия принципиально новых моментов развернутую теорию дискриминантного анализа в пространствах общей природы здесь не приводим, ограничившись данными выше замечаниями.

ГЛАВА 4. ОСНОВНЫЕ ИДЕИ СТАТИСТИКИ ИНТЕРВАЛЬНЫХ ДАННЫХ

В статистике интервальных данных элементы выборки — не числа, а интервалы. Это приводит к алгоритмам и выводам, принципиально отличающимся от классических. Настоящая глава посвящена основным идеям и подходам асимптотической статистики интервальных данных. Приведены результаты, связанные с основополагающими в рассматриваемой области прикладной математической статистики понятиями нотны и рационального объема выборки.

4.1. Развитие статистики интервальных данных

Перспективная и быстро развивающаяся область статистических исследований последних десятилетий — математическая статистика интервальных данных. Речь идет о развитии методов прикладной математической статистики в ситуации, когда статистические данные — не числа, а интервалы, в частности, порожденные наложением ошибок измерения на значения случайных величин. Полученные результаты отражены в выступлениях на проведенной в «Заводской лаборатории» дискуссии [1] и в докладах Международной конференции ИНТЕРВАЛ-92 [2]. Приведем основные идеи весьма перспективного для вероятностно-статистических методов и моделей принятия решений асимптотического направления в статистике интервальных данных.

В настоящее время признается необходимым изучение устойчивости (робастности) оценок параметров к малым отклонениям исходных данных и предпосылок модели. Однако популярная среди теоретиков модель засорения (Тьюки-Хьюбера) представляется не вполне адекватной. Эта модель нацелена на изучение влияния больших «выбросов». Поскольку любые реальные измерения лежат в некотором фиксированном диапазоне, а именно, заданном в техническом паспорте средства измерения, то зачастую выбросы не могут быть слишком большими. Поэтому представляются полезными иные, более общие схемы устойчивости, введенные в монографии [3], в которых, например, учитываются отклонения распределений результатов наблюдений от предположений модели.

В одной из таких схем изучается влияние интервальности исходных данных на статистические выводы. Необходимость такого изучения стала очевидной следующим образом. В государственных стандартах СССР по прикладной статистике в обязательном порядке давалось справочное приложение «Примеры применения правил стандарта». При подготовке ГОСТ 11.011-83 [4] разработчикам стандарта были переданы для анализа реальные данные о наработке резцов до предельного состояния (в часах). Оказалось, что все эти данные представляли собой либо целые числа, либо

полуцелые (т.е. после умножения на 2 становящиеся целыми). Ясно, что исходная длительность наработок искажена. Необходимо учесть в статистических процедурах наличие такого искажения исходных данных. Как это сделать?

Первое, что приходит в голову — модель группировки данных, согласно которой для истинного значения X проводится замена на ближайшее число из множества $\{0{,}5n,\ n = 1, 2, 3, ...\}$. Однако эту модель целесообразно подвергнуть сомнению, а также рассмотреть иные модели. Так, возможно, что X надо приводить к ближайшему сверху элементу указанного множества — если проверка качества поставленных на испытание резцов проводилась раз в полчаса. Другой вариант: если расстояния от X до двух ближайших элементов множества $\{0{,}5n,\ n = 1, 2, 3, ...\}$ примерно равны, то естественно ввести рандомизацию при выборе заменяющего числа, и т.д.

Целесообразно построить новую математико-статистическую модель, согласно которой **результаты наблюдений — не числа, а интервалы**. Например, если в таблице приведено значение 53,5, то это значит, что реальное значение — какое-то число от 53,0 до 54,0, т.е. какое-то число в интервале [53,5 – 0,5; 53,5 + 0,5], где 0,5 — максимально возможная погрешность. Принимая эту модель, мы попадаем в новую научную область — статистику интервальных данных [5, 6]. Статистика интервальных данных идейно связана с интервальной математикой, в которой в роли чисел выступают интервалы (см., например, монографию [7]). Это направление математики является дальнейшим развитием всем известных правил приближенных вычислений, посвященных выражению погрешностей суммы, разности, произведения, частного через погрешности тех чисел, над которыми осуществляются перечисленные операции.

В интервальной математике сумма двух интервальных чисел $[a, b]$ и $[c, d]$ имеет вид $[a, b] + [c, d] = [a + c, b + d]$, а разность определяется по формуле $[a, b] - [c, d] = [a - d, b - c]$. Для положительных a, b, c, d произведение определяется формулой $[a, b] \cdot [c, d] = [ac, bd]$, а частное имеет вид $[a, b]/[c, d] = = [a/d, b/c]$. Эти формулы получены при решении соответствующих оптимизационных задач. Пусть x лежит в отрезке $[a, b]$, а y — в отрезке $[c, d]$. Каково минимальное и максимальное значение для $x + y$? Очевидно, $a + c$ и $b + d$ соответственно. Минимальные и максимальные значения для $x - y$, xy, x/y указывают нижние и верхние границы для интервальных чисел, задающих результаты арифметических операций. А от арифметических операций можно перейти ко всем остальным математическим алгоритмам. Так строится интервальная математика.

Как видно из сборника трудов Международной конференции [2], исследователям удалось решить ряд задач теории интервальных дифференциальных уравнений, в которых коэффициенты, начальные

условия и решения описываются с помощью интервалов. По мнению ряда специалистов, статистика интервальных данных является частью интервальной математики [7]. Впрочем, распространена и другая точка зрения, согласно которой такое включение нецелесообразно, поскольку статистика интервальных данных использует несколько иные подходы к алгоритмам анализа реальных данных, чем сложившиеся в интервальной математике (подробнее см. ниже).

В настоящей главе развиваем асимптотические методы статистического анализа интервальных данных при больших объемах выборок и малых погрешностях измерений. В отличие от классической математической статистики, сначала устремляется к бесконечности объем выборки и только потом — уменьшаются до нуля погрешности (в классической математической статистике предельные переходы осуществляются в обратном порядке – сначала уменьшаются до нуля погрешности измерений, и только затем - устремляется к бесконечности объем выборки). В частности, еще в начале 1980-х годов с помощью такой асимптотики сформулированы правила выбора метода оценивания в ГОСТ 11.011-83 [4].

Нами разработана [8] общая схема исследования, включающая расчет нотны (максимально возможного отклонения статистики, вызванного интервальностью исходных данных) и рационального объема выборки (превышение которого не дает существенного повышения точности оценивания). Она применена к оцениванию математического ожидания и дисперсии [1], медианы и коэффициента вариации [9], параметров гамма-распределения [4, 10] и характеристик аддитивных статистик [8], при проверке гипотез о параметрах нормального распределения, в т.ч. с помощью критерия Стьюдента, а также гипотезы однородности с помощью критерия Смирнова [9]. Изучено асимптотическое поведение оценок метода моментов и оценок максимального правдоподобия (а также более общих — оценок минимального контраста), проведено асимптотическое сравнение этих методов в случае интервальных данных, найдены общие условия, при которых, в отличие от классической математической статистики, метод моментов дает более точные оценки, чем метод максимального правдоподобия [11].

Разработаны подходы к рассмотрению интервальных данных в основных постановках регрессионного, дискриминантного и кластерного анализов [12]. Изучено влияние погрешностей измерений и наблюдений на свойства алгоритмов регрессионного анализа, разработаны способы расчета нотн и рациональных объемов выборок, введены и исследованы новые понятия многомерных и асимптотических нотн, доказаны соответствующие предельные теоремы [12, 13]. Проведена первоначальная разработка интервального дискриминантного анализа, рассмотрено влияние интервальности данных на показатель качества классификации

[12, 14]. Основные идеи и результаты рассматриваемого направления в статистике интервальных данных приведены в публикациях обзорного характера [5, 6].

Как показала Международная конференция ИНТЕРВАЛ-92, в области асимптотической математической статистики интервальных данных мы имеем мировой приоритет. По нашему мнению, со временем во все виды статистического программного обеспечения должны быть включены алгоритмы интервальной статистики, «параллельные» обычно используемым алгоритмам прикладной математической статистики. Это позволит в явном виде учесть наличие погрешностей у результатов наблюдений, сблизить позиции метрологов и статистиков.

Многие из утверждений статистики интервальных данных весьма отличаются от аналогов из классической математической статистики. В частности, не существует состоятельных оценок; средний квадрат ошибки оценки, как правило, асимптотически равен сумме дисперсии оценки, рассчитанной согласно классической теории, и некоторого положительного числа (равного квадрату т.н. нотны — максимально возможного отклонения значения статистики из-за погрешностей исходных данных) — в результате, метод моментов оказывается иногда точнее метода максимального правдоподобия [11]; нецелесообразно увеличивать объем выборки сверх некоторого предела (называемого рациональным объемом выборки) — вопреки классической теории, согласно которой чем больше объем выборки, тем точнее выводы.

В стандарт [4] включен раздел 5, посвященный выбору метода оценивания при неизвестных параметрах формы и масштаба и известном параметре сдвига и основанный на концепциях статистики интервальных данных. Теоретическое обоснование этого раздела стандарта опубликовано лишь через 5 лет в статье [10].

В 1982 г. при разработке стандарта [4] сформулированы основные идеи статистики интервальных данных. Однако из-за недостатка времени они не были полностью реализованы в ГОСТ 11.011-83, и этот стандарт написан в основном в классической манере. Развитие идей статистики интервальных данных продолжается уже в течение 25 лет, и еще многое необходимо сделать! Большое значение статистики интервальных данных для современной прикладной статистики обосновано в [15, 16].

Вторая ведущая научная школа в области статистики интервальных данных, дополняющая нашу — это школа проф. А.П. Вощинина (1937 - 2008), активно работающая с конца 70-х годов. Полученные результаты отражены в ряде монографий (см., прежде всего, [17, 18, 19]), статей [1, 20, 21], докладов, в частности, в трудах [2] Международной конференции ИНТЕРВАЛ-92, диссертациях [22, 23]. Изучены проблемы регрессионного анализа, планирования эксперимента, сравнения альтернатив и принятия решений в условиях интервальной неопределенности.

Рассматриваемое ниже наше научное направление отличается нацеленностью на асимптотические результаты, полученные при больших объемах выборок и малых погрешностях измерений, поэтому его полное название таково: асимптотическая математическая статистика интервальных данных.

4.2. Основные идеи статистики интервальных данных

Сформулируем сначала основные идеи асимптотической математической статистики интервальных данных, а затем рассмотрим реализацию этих идей на перечисленных выше примерах. Основные идеи достаточно просты, в то время как их проработка в конкретных ситуациях зачастую оказывается достаточно трудоемкой.

Пусть существо реального явления описывается выборкой x_1, x_2, ..., x_n. В вероятностной теории математической статистики, из которой мы исходим (см. справочник [24]), выборка — это набор независимых в совокупности одинаково распределенных случайных величин. Однако беспристрастный и тщательный анализ подавляющего большинства реальных задач показывает, что статистику известна отнюдь не выборка x_1, x_2, ..., x_n, а величины

$$y_j = x_j + \varepsilon_j, j = 1, 2, ..., n,$$

где ε_1, ε_2, …, ε_n — некоторые погрешности измерений, наблюдений, анализов, опытов, исследований (например, инструментальные ошибки).

Одна из причин появления погрешностей — запись результатов наблюдений с конечным числом значащих цифр. Дело в том, что для случайных величин с непрерывными функциями распределения событие, состоящее в попадании хотя бы одного элемента выборки в множество рациональных чисел, согласно правилам теории вероятностей имеет вероятность 0, а такими событиями в теории вероятностей принято пренебрегать. Поэтому при рассуждениях о выборках из нормального, логарифмически нормального, экспоненциального, равномерного, гамма-распределений, распределения Вейбулла-Гнеденко и др. приходится принимать, что эти распределения имеют элементы исходной выборки x_1, x_2, ..., x_n, в то время как статистической обработке доступны лишь искаженные значения $y_j = x_j + \varepsilon_j$.

Введем обозначения

$$x = (x_1, x_2, ..., x_n), y = (y_1, y_2, ..., y_n), \varepsilon = (\varepsilon_1 + \varepsilon_2 + \ldots + \varepsilon_n).$$

Пусть статистические выводы основываются на статистике $f : R^n \to R^1$, используемой для оценивания параметров и характеристик распределения, проверки гипотез и решения иных статистических задач. Принципиально важная для статистики интервальных данных идея такова: СТАТИСТИК ЗНАЕТ ТОЛЬКО $f(y)$, НО НЕ $f(x)$.

Очевидно, в статистических выводах необходимо отразить различие между $f(y)$ и $f(x)$. Одним из двух основных понятий статистики интервальных данных является понятие нотны.

Определение. Величину максимально возможного (по абсолютной величине) отклонения, вызванного погрешностями наблюдений ε, известного статистику значения $f(y)$ от истинного значения $f(x)$, т.е.

$$N_f(x) = \sup | f(y) - f(x) |,$$

где супремум берется по множеству возможных значений вектора погрешностей ε (см. ниже), будем называть НОТНОЙ.

Если функция f имеет частные производные второго порядка, а ограничения на погрешности имеют вид

$$| \varepsilon_i | \le \Delta, \; i = 1, 2, \ldots, n, \qquad (1)$$

причем Δ мало, то приращение функции f с точностью до бесконечно малых более высокого порядка описывается главным линейным членом, т.е.

$$f(y) - f(x) = \sum_{1 \le i \le n} \frac{\partial f(x)}{\partial x_i} \varepsilon_i + O(\Delta^2).$$

Чтобы получить асимптотическое (при $\Delta \to 0$) выражение для нотны, достаточно найти максимум и минимум линейной функции (главного линейного члена) на кубе, заданном неравенствами (1). Легко видеть, что максимум достигается, если положить

$$\varepsilon_i = \begin{cases} \Delta, & \dfrac{\partial f(x)}{\partial x_i} \ge 0, \\ -\Delta, & \dfrac{\partial f(x)}{\partial x_i} < 0, \end{cases}$$

а минимум, отличающийся от максимума только знаком, достигается при $\varepsilon_i' = -\varepsilon_i$. Следовательно, *нотна* с точностью до бесконечно малых более высокого порядка имеет вид

$$N_f(x) = \left(\sum_{1 \le i \le n} \left| \frac{\partial f(x)}{\partial x_i} \right| \right) \Delta.$$

Это выражение назовем *асимптотической нотной*.

Условие (1) означает, что исходные данные представляются статистику в виде интервалов $[y_i - \Delta; \; y_i + \Delta]$, $i = 1, 2, \ldots, n$ (отсюда и название этого научного направления). Ограничения на погрешности могут задаваться разными способами — кроме абсолютных ошибок используются относительные или иные показатели различия между x и y.

Если задана не предельная абсолютная погрешность Δ, а предельная относительная погрешность δ, т.е. ограничения на погрешности вошедших в выборку результатов измерений имеют вид

$$| \varepsilon_i | \le \delta | x_i |, \; i = 1, 2, \ldots, n,$$

то аналогичным образом получаем, что нотна с точностью до бесконечно малых более высокого порядка, т.е. асимптотическая нотна, имеет вид

$$N_f(x) = \left(\sum_{1 \le i \le n} | x_i \frac{\partial f(x)}{\partial x_i} | \right) \delta.$$

При практическом использовании рассматриваемой концепции необходимо провести тотальную замену символов x на символы y. В каждом конкретном случае удается показать, что в силу малости погрешностей разность $N_f(y) - N_f(x)$ является бесконечно малой более высокого порядка сравнительно с $N_f(x)$ или $N_f(y)$.

4.3. Основные результаты в вероятностной модели

В классической вероятностной модели элементы исходной выборки $x_1, x_2, ..., x_n$ рассматриваются как независимые одинаково распределенные случайные величины. Как правило, существует некоторая константа $C > 0$ такая, что в смысле сходимости по вероятности

$$\lim_{n \to \infty} N_f(x) = C\Delta. \qquad (2)$$

Соотношение (2) доказывается отдельно для каждой конкретной задачи.

При использовании классических статистических методов в большинстве случаев используемая статистика $f(x)$ является асимптотически нормальной. Это означает, что существуют константы a и σ^2 такие, что

$$\lim_{n \to \infty} P\left(\sqrt{n} \frac{f(x) - a}{\sigma} < x \right) = \Phi(x),$$

где $\Phi(x)$ — функция стандартного нормального распределения с математическим ожиданием 0 и дисперсией 1. При этом обычно оказывается, что

$$\lim_{n \to \infty} \sqrt{n}(Mf(x) - a) = 0 \text{ и } \lim_{n \to \infty} nDf(x) = \sigma^2,$$

а потому в классической математической статистике средний квадрат ошибки статистической оценки равен

$$M(f(x) - a)^2 = (Mf(x) - a)^2 + Df(x) = \frac{\sigma^2}{n}$$

с точностью до членов более высокого порядка.

В статистике интервальных данных ситуация совсем иная — обычно можно доказать, что средний квадрат ошибки равен

$$\max_{\{\varepsilon\}} M(f(x) - a)^2 = \frac{\sigma^2}{n} + N_f^2(y) + o\left(\Delta^2 + \frac{1}{n} \right). \qquad (3)$$

Из соотношения (3) вытекает ряд важных следствий. Правая часть этого равенства, в отличие от правой части соответствующего классического равенства, не стремится к 0 при безграничном возрастании объема выборки. Она остается больше некоторого положительного числа, а именно, квадрата нотны. Следовательно, статистика $f(x)$ не является состоятельной оценкой параметра a. Более того, состоятельных оценок вообще не существует.

Пусть доверительным интервалом для параметра a, соответствующим заданной доверительной вероятности γ, в классической математической статистике является интервал $(c_n(\gamma); d_n(\gamma))$. В статистике интервальных данных аналогичный доверительный интервал является более широким. Он имеет вид $(c_n(\gamma) - N_f(y); d_n(\gamma) + N_f(y))$. Таким образом, его длина увеличивается на две нотны. Следовательно, при увеличении объема выборки длина доверительного интервала не может стать меньше, чем $2C\Delta$ (см. формулу (2)).

В статистике интервальных данных методы оценивания параметров имеют другие свойства по сравнению с классической математической статистикой. Так, при больших объемах выборок метод моментов может быть заметно лучше, чем метод максимального правдоподобия (т.е. иметь меньший средний квадрат ошибки — см. формулу (3)), в то время как в классической математической статистике второй из названных методов всегда не хуже первого.

4.4. Рациональный объем выборки

Анализ формулы (3) показывает, что в отличие от классической математической статистики нецелесообразно безгранично увеличивать объем выборки, поскольку средний квадрат ошибки остается всегда большим квадрата нотны. Поэтому представляется полезным ввести понятие «рационального объема выборки» n_{rat}, при достижении которого продолжать наблюдения нецелесообразно.

Как установить «рациональный объем выборки»? Можно воспользоваться идеей «принципа уравнивания погрешностей», выдвинутой в монографии [3]. Речь идет о том, что вклад погрешностей различной природы в общую погрешность должен быть примерно одинаков. Этот принцип дает возможность выбирать необходимую точность оценивания тех или иных характеристик в тех случаях, когда это зависит от исследователя. В статистике интервальных данных в соответствии с «принципом уравнивания погрешностей» предлагается определять рациональный объем выборки n_{rat} из условия равенства двух величин — метрологической составляющей, связанной с нотной, и статистической составляющей — в среднем квадрате ошибки (3), т.е. из условия

$$\frac{\sigma^2}{n_{rat}} = N_f^2(y),\ n_{rat} = \frac{\sigma^2}{N_f^2(y)}.$$

Для практического использования выражения для рационального объема выборки неизвестные теоретические характеристики необходимо заменить их оценками. Это делается в каждой конкретной задаче по-своему.

Исследовательскую программу в области статистики интервальных данных можно «в двух словах» сформулировать так: для любого алгоритма анализа данных (алгоритма прикладной статистики) необходимо вычислить

нотну и рациональный объем выборки. Или иные величины из того же понятийного ряда, возникающие в многомерном случае, при наличии нескольких выборок и при иных обобщениях описываемой здесь простейшей схемы. Затем проследить влияние погрешностей исходных данных на точность оценивания, доверительные интервалы, значения статистик критериев при проверке гипотез, уровни значимости и другие характеристики статистических выводов. Очевидно, классическая математическая статистика является частью статистики интервальных данных, выделяемой условием $\Delta = 0$.

Поясним теоретические концепции статистики интервальных данных на простых примерах.

4.5. Оценивание математического ожидания

Пусть необходимо оценить математическое ожидание случайной величины с помощью обычной оценки — среднего арифметического результатов наблюдений, т.е.

$$f(x) = \frac{x_1 + x_2 + ... + x_n}{n}.$$

Тогда при справедливости ограничений (1) на абсолютные погрешности имеем $N_f(x) = \Delta$. Таким образом, нотна полностью известна и не зависит от многомерной точки, в которой берется. Вполне естественно: если каждый результат наблюдения известен с точностью до Δ, то и среднее арифметическое известно с той же точностью. Ведь возможна систематическая ошибка — если к каждому результату наблюдения добавить Δ, то и среднее арифметическое увеличится на Δ.

Поскольку

$$D(\bar{x}) = \frac{D(x_1)}{n},$$

то в ранее введенных обозначениях

$$\sigma^2 = D(x_1).$$

Следовательно, рациональный объем выборки равен

$$n_{rat} = \frac{D(x_1)}{\Delta^2}.$$

Для практического использования полученной формулы надо оценить дисперсию результатов наблюдений. Можно доказать, что, поскольку Δ мало, это можно сделать обычным способом, например, с помощью несмещенной выборочной оценки дисперсии

$$s^2(y) = \frac{1}{n-1} \sum_{1 \le i \le n} (y_i - \bar{y})^2.$$

Здесь и далее рассуждения часто идут на двух уровнях. Первый — это уровень «истинных» случайных величин, обозначаемых «x», описывающих реальность, но неизвестных специалисту по анализу данных. Второй — уровень известных этому специалисту величин «y», отличающихся

погрешностями от истинных. Погрешности малы, поэтому функции от x отличаются от функций от y на некоторые бесконечно малые величины. Эти соображения и позволяют использовать $s^2(y)$ как оценку $D(x_1)$.

Итак, выборочной оценкой рационального объема выборки является

$$n_{sample-rat} = \frac{s^2(y)}{\Delta^2}.$$

Уже на этом первом рассматриваемом примере видим, что рациональный объем выборки находится не где-то вдали, а непосредственно рядом с теми объемами, с которыми имеет дело любой практически работающий статистик. Например, если статистик знает, что

$$\Delta = \frac{\sigma}{6},$$

то $n_{rat} = 36$. А именно такова погрешность контрольных шаблонов во многих технологических процессах! Поэтому, занимаясь управлением качеством, необходимо обращать внимание на действующую на предприятии систему измерений.

По сравнению с классической математической статистикой доверительный интервал для математического ожидания (для заданной доверительной вероятности γ) имеет другой вид:

$$\left(\bar{y} - \Delta - u(\gamma)\frac{s}{\sqrt{n}}; \bar{y} + \Delta + u(\gamma)\frac{s}{\sqrt{n}} \right), \quad (4)$$

где $u(\gamma)$ — квантиль порядка $(1 + \gamma)/2$ стандартного нормального распределения с математическим ожиданием 0 и дисперсией 1.

По поводу формулы (4) была довольно жаркая дискуссия среди специалистов. Отмечалось, что она получена на основе Центральной предельной теоремы теории вероятностей и может быть использована при любом распределении результатов наблюдений (с конечной дисперсией). Если же имеется дополнительная информация, то, по мнению отдельных специалистов, формула (4) может быть уточнена. Например, если известно, что распределение x_i является нормальным, в качестве $u(\gamma)$ целесообразно использовать квантиль распределения Стьюдента. К этому надо добавить, что по небольшому числу наблюдений нельзя надежно установить нормальность, а при росте объема выборки квантили распределения Стьюдента приближаются к квантилям нормального распределения.

Вопрос о том, часто ли результаты наблюдений имеют нормальное распределение, подробно обсуждался среди специалистов. Выяснилось, что распределения встречающихся в практических задачах результатов измерений почти всегда отличны от нормальных [25]. А также и от распределений из иных параметрических семейств, описываемых в учебниках.

Применительно к оцениванию математического ожидания (но не к оцениванию других характеристик или параметров распределения) факт существования границы возможной точности, определяемой точностью

исходных данных, неоднократно отмечался в литературе ([26, с. 230–234], [27, с. 121] и др.).

4.6. Оценивание дисперсии

Для статистики $f(y) = s^2(y)$, где $s^2(y)$ — выборочная дисперсия (несмещенная оценка теоретической дисперсии), при справедливости ограничений (1) на абсолютные погрешности имеем

$$N_f(y) = \frac{2\Delta}{n-1}\sum_{i=1}^{n} | y_i - \bar{y} | + O(\Delta^2).$$

Можно показать, что нотна $N_f(y)$ сходится к

$$2\Delta M | x_1 - M(x_1) |$$

по вероятности с точностью до $o(\Delta)$, когда n стремится к бесконечности. Это же предельное соотношение верно и для нотны $N_f(x)$, вычисленной для исходных данных. Таким образом, в данном случае справедлива формула (2) с

$$C = 2M | x_1 - M(x_1) |.$$

Известно [28], что случайная величина

$$\frac{s^2 - \sigma^2}{\sqrt{n}}$$

является асимптотически нормальной с математическим ожиданием 0 и дисперсией $D(x_1^2)$.

Из сказанного вытекает: в статистике интервальных данных асимптотический доверительный интервал для дисперсии σ^2 (соответствующий доверительной вероятности γ) имеет вид

$$(s^2(y) - A;\ s^2 + A),$$

где

$$A = \frac{u(\gamma)}{\sqrt{n(n-1)}}\sqrt{\sum_{i=1}^{n}\left(y_i^2 - \frac{1}{n}\sum_{j=1}^{n} y_j^2\right)^2} + \frac{2\Delta}{n-1}\sum_{i=1}^{n} | y_i - \bar{y} |,$$

здесь $u(\gamma)$ обозначает тот же самый квантиль стандартного нормального распределения, что и выше в случае оценивания математического ожидания.

Рациональный объем выборки при оценивании дисперсии равен

$$n_{rat} = \frac{D(x_1^2)}{4\Delta^2 (M | x_1 - M(x_1) |)^2},$$

а выборочную оценку рационального объема выборки $n_{sample-rat}$ можно вычислить, заменяя теоретические моменты на соответствующие выборочные и используя доступные статистику результаты наблюдений, содержащие погрешности.

Что можно сказать о численной величине рационального объема выборки? Как и в случае оценивания математического ожидания, она отнюдь не выходит за пределы обычно используемых объемов выборок. Так, если распределение результатов наблюдений x_i является нормальным с

математическим ожиданием 0 и дисперсией σ^2, то в результате вычисления моментов случайных величин в предыдущей формуле получаем, что

$$n_{rat} = \frac{\sigma^2}{\pi\Delta^2},$$

где π — отношение длины окружности к диаметру, π = = 3,141592… Например, если $\Delta = \sigma/6$, то n_{rat} = 11. Это меньше, чем при оценивании математического ожидания в предыдущем примере.

4.7. Статистика интервальных данных в прикладной статистике

Кратко рассмотрим положение *статистики интервальных данных* (СИД) среди других методов описания неопределенностей и анализа данных.

Нечеткость и СИД. С формальной точки зрения описание нечеткости интервалом — это частный случай описания ее нечетким множеством. В СИД функция принадлежности нечеткого множества имеет специфический вид — она равна 1 в некотором интервале и 0 вне его. Такая функция принадлежности описывается всего двумя параметрами (границами интервала). Эта простота описания делает математический аппарат СИД гораздо более прозрачным, чем аппарат теории нечеткости в общем случае. Это, в свою очередь, позволяет исследователю продвинуться дальше, чем при использовании функций принадлежности произвольного вида.

Интервальная математика и СИД. Можно было бы сказать, что СИД — часть интервальной математики, что СИД так соотносится с прикладной математической статистикой, как интервальная математика — с математикой в целом. Однако исторически сложилось так, что интервальная математика занимается прежде всего вычислительным погрешностями. С точки зрения интервальной математики две известные формулы для выборочной дисперсии, а именно

$$s^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 = \frac{1}{n}\sum_{i=1}^{n}x_i^2 - \left(\bar{x}\right)^2,$$

имеют разные погрешности. А с точки зрения СИД эти две формулы задают одну и ту же функцию, и поэтому им соответствуют совпадающие нотны и рациональные объемы выборок. Интервальная математика прослеживает процесс вычислений, СИД этим не занимается. Необходимо отметить, что типовые постановки СИД могут быть перенесены в другие области математики, и, наоборот, вычислительные алгоритмы прикладной математической статистики и СИД заслуживают изучения. Однако и то, и другое — скорее дело будущего. Из уже сделанного отметим применение методов СИД при анализе такой характеристики финансовых потоков, как *NPV* — чистая текущая стоимость [29, гл.9].

Математическая статистика и СИД. Математическая статистика и СИД отличаются тем, в каком порядке делаются предельные переходы $n \to \infty$ и $\Delta \to 0$. При этом СИД переходит в математическую статистику при $\Delta = 0$. Правда, тогда исчезают основные особенности СИД: нотна становится равной 0, а рациональный объем выборки — бесконечности. Рассмотренные выше методы СИД разработаны в предположении, что погрешности малы (но не исчезают), а объем выборки велик. СИД расширяет классическую математическую статистику тем, что в исходных статистических данных каждое число заменяет интервалом. С другой стороны, можно считать СИД новым этапом развития математической статистики.

Статистика объектов нечисловой природы и СИД. *Статистика объектов нечисловой природы* (СОНП) [30] расширяет область применения классической математической статистики путем включения в нее новых видов статистических данных. Естественно, при этом появляются новые виды алгоритмов анализа статистических данных и новый математический аппарат (в частности, происходит переход от методов суммирования к методам оптимизации). С точки зрения СОНП частному виду новых статистических данных — интервальным данным — соответствует СИД. Напомним, что одно из двух основных понятий СИД — нотна — определяется как решение оптимизационной задачи. Однако СИД, изучая классические методы прикладной статистики применительно к интервальным данным, по математическому аппарату ближе к классической математической статистике, чем другие части СОНП, например, статистика бинарных отношений.

Робастные методы статистики и СИД. Если понимать робастность согласно [3] как теорию устойчивости статистических методов по отношению к допустимым отклонениям исходных данных и предпосылок модели, то в СИД рассматривается одна из естественных постановок робастности. Однако в массовом сознании специалистов термин «робастность» закрепился за моделью засорения выборки большими выбросами (модель Тьюки-Хубера), хотя эта модель не имеет большого практического значения [31]. К этой модели СИД не имеет отношения.

Теория устойчивости и СИД. Общей схеме устойчивости [3, 32, 33] математических моделей социально-экономических явлений и процессов по отношению к допустимым отклонениям исходных данных и предпосылок моделей СИД полностью соответствует. Она посвящена математико-статистическим моделям, используемым при анализе статистических данных, а допустимые отклонения — это интервалы, заданные ограничениями на погрешности. СИД можно рассматривать как пример теории, в которой учет устойчивости позволил сделать нетривиальные выводы. Отметим, что с точки зрения общей схемы

устойчивости [3] устойчивость по Ляпунову в теории дифференциальных уравнений — весьма частный случай, в котором из-за его конкретности удалось весьма далеко продвинуться.

Минимаксные методы, типовые отклонения и СИД. Постановки СИД относятся к минимаксным. За основу берется максимально возможное отклонение. Это — «подход пессимиста», применяемый, например, в теории антагонистических игр. Использование минимаксного подхода позволяет подозревать СИД в завышении роли погрешностей измерения. Однако примеры изучения вероятностно-статистических моделей погрешностей, проведенные, в частности, при разработке методов оценивания параметров гамма-распределения [4, 10], показали, что это подозрение не подтверждается. Влияние погрешностей измерений по порядку такое же, только вместо максимально возможного отклонения (нотны) приходится рассматривать математическое ожидание соответствующего отклонения. Подчеркнем, что применение в СИД вероятностно-статистических моделей погрешностей не менее перспективно, чем минимаксных.

Подход научной школы А.П. Вощинина и СИД. Если в математической статистике неопределенность только статистическая, то в научной школе А.П. Вощинина — только интервальная. Можно сказать, что СИД лежит между классической прикладной математической статистикой и областью исследований научной школы А.П. Вощинина. Другое отличие состоит в том, что в этой школе разрабатывают новые методы анализа интервальных данных, а в СИД в настоящее время изучается устойчивость классических статистических методов по отношению к малым погрешностям. Подход СИД оправдывается распространенностью этих методов, однако в дальнейшем следует переходить к разработке новых методов, специально предназначенных для анализа интервальных данных.

Анализ чувствительности и СИД. При анализе чувствительности, как и в СИД, рассчитывают производные по используемым переменным, или непосредственно находят изменения при отклонении переменной на ±10% от базового значения. Однако этот анализ делают по каждой переменной отдельно. В СИД все переменные рассматриваются совместно, и находится максимально возможное отклонение (нотна). При малых погрешностях удается на основе главного члена разложения функции в многомерный ряд Тейлора получить удобную формулу для нотны. Можно сказать, что СИД — это многомерный анализ чувствительности.

4.8. Заключительные замечания

Асимптотической математической статистике интервальных данных посвящены главы в учебниках [31, 34, 35]. Развиваются научные

исследования как в научной школе А.П. Вощинина [36, 37], так и в СИД [38, 39].

По нашему мнению, во все виды статистического программного обеспечения должны быть включены алгоритмы интервальной статистики, «параллельные» обычно используемым в настоящее время алгоритмам прикладной математической статистики. Это позволит в явном виде учесть наличие погрешностей у результатов наблюдений (измерений, испытаний, анализов, опытов).

Статистика интервальных данных является составной частью системной нечеткой интервальной математики [40, 41] – перспективного направления теоретической и вычислительной математики.

ГЛАВА 5. ВЕРОЯТНОСТНО-СТАТИСТИЧЕСКИЕ МОДЕЛИ КОРРЕЛЯЦИИ И РЕГРЕССИИ

Коэффициенты корреляции и детерминации широко используются при статистическом анализе данных. При этом достаточно часто допускаются те или иные ошибки. Некоторые из них рассмотрены ниже.

В соответствии с подходом системной нечеткой интервальной математики рассмотрим системы моделей корреляции и регрессии.

Ограничимся сначала случаем двух переменных. Пусть (X, Y) - двумерный случайный вектор. Наиболее часто используют линейный парный коэффициент корреляции Пирсона и непараметрические ранговые коэффициенты Спирмена и Кендалла.

Согласно теории измерений [1] коэффициент корреляции Пирсона можно применять к переменным, измеренным в шкале интервалов (и в шкалах с более узкой группой допустимых преобразований, например, в шкале отношений). Его нельзя применять при анализе порядковых данных (например, для анализа связи успеваемости по двум учебным предметам). Непараметрические ранговые коэффициенты Спирмена и Кендалла предназначены для оценки связи порядковых переменных. Их можно использовать и в шкалах с более узкой группой допустимых преобразований, например, в шкалах интервалов или отношений. Исходя из теории устойчивости [2], одни и те же данные целесообразно обработать разными способами и сравнить результаты. В частности, целесообразно рассчитать все упомянутые выше коэффициенты корреляции.

Если X и Y - независимые случайные величины, то коэффициенты корреляции равны 0. Обратное неверно - из равенства 0 коэффициента корреляции не следует, что случайные величины X и Y - независимы.

5.1. Значимость отличия от 0 и "шкала Чеддока"

Выборочные коэффициенты корреляции - случайные величины. Их распределения являются асимптотически нормальными.

Часто проверяют нулевую гипотезу о том, что тот или иной теоретический коэффициент корреляции равен 0. Если эта гипотеза отклоняется, то можно утверждать, что случайные величины X и Y зависимы. Гипотеза отклоняется на уровне значимости α, если выборочный коэффициент корреляции по абсолютной величине больше граничного значения $C(\alpha)f(n)$, где n - объем выборки, C и f - некоторые функции, причем

$$\lim_{n\to\infty} f(n) = 0.$$

Для коэффициента корреляции Пирсона функция f зависит от распределения случайного вектора (X, Y). Распространенные таблицы рассчитаны для случая двумерного нормального распределения (X, Y). Хорошо известно, что распределения подавляющего большинства реальных данных не являются нормальными. Следовательно, применение правил, сформированных для двумерного нормального распределения, как правило, не является обоснованным.

Для непараметрических коэффициентов ранговой корреляции Спирмена и Кендалла свойства правил проверки гипотезы о том, что теоретический коэффициент корреляции равен 0, не зависят от распределения данных.

Иногда показателям тесноты связи (модулям коэффициентов корреляции) пытаются дать качественную оценку (т.н. шкала Чеддока, см. табл.1):

Таблица 1. Шкала Чеддока

Количественная мера тесноты связи	**Качественная характеристика силы связи**
0,1 - 0,3	Слабая
0,3 - 0,5	Умеренная
0,5 - 0,7	Заметная
0,7 - 0,9	Высокая
0,9 - 0,99	Весьма высокая

Такая рекомендация не вполне адекватна. При малых объемах выборки значение коэффициента корреляции 0,5 или 0,7 вполне совместимо со справедливостью гипотезы о том, что теоретический коэффициент корреляции равен 0. А при достаточно большом объеме

выборки коэффициент 0,1 может свидетельствовать о необходимости отклонения такой гипотезы.

5.2. Активный и пассивный эксперименты

Вопреки часто встречающимся мнениям и предложениям, коэффициенты корреляции можно обоснованно использовать лишь для прогнозирования, но не для управления.

Рассмотрим упрощенный пример. Пусть X - число телевизоров в городе, Y - число преступлений в этом городе, Z - число психических заболеваний в нем. Были собраны данные по нескольким сотням городов (англосаксонских стран). Выборочный коэффициент корреляции между X и Y оказался равным практически 1. Весьма мало отличался от 1 и выборочный коэффициент корреляции между X и Z. С высокой степенью точности справедливы зависимости $Y = aX$ и $Z = bX$. С помощью этих зависимостей можно надежно прогнозировать число преступлений и число психических заболеваний по число телевизоров в городе.

В подобных ситуациях часто возникает желание использовать зависимости $Y = aX$ и $Z = bX$ для управления. Однако очевидно, что прекращение телевещания (переход к $X = 0$) не приведет к резкому снижению число преступлений и число психических заболеваний. В чем причина неудачи, казалось бы, естественного подхода к управлению? Дело в том, что значения всех трех рассматриваемых переменных определяются значениями четвертой переменной (латентной, скрытой) - числа жителей города W. А именно, с высокой точностью $X = cW$, $Y = dW$, $Z = eW$, откуда $Y = (d/c)X$, $Z = (e/c)X$.

Проблема в том, что при анализе реальных данных не всегда ясно наличие или отсутствие латентных переменных, определяющих успех управления по регрессионным зависимостям. Полезны понятия "пассивный эксперимент" и "активный эксперимент". При пассивном эксперименте данные накапливаются путем пассивного наблюдения, другими словами, информацию получают в условиях обычного функционирования изучаемых объектов. Активный эксперимент проводится с применением искусственного воздействия на изучаемые объекты по специальной программе.

При пассивном эксперименте существуют только факторы в виде входных контролируемых, но неуправляемых переменных, и экспериментатор находится в положении пассивного наблюдателя. Задача планирования в этом случае сводится к оптимальной организации сбора информации и решению таких вопросов, как выбор количества и частоты измерений, выбор метода обработки результатов измерений.

Наиболее часто целью пассивного эксперимента является построение математической модели объекта. Хорошим примером пассивного

эксперимента являются измерения метеорологических параметров (температуры, скорости ветра и т.д.).

Активный эксперимент основан на задании экспериментатором значений факторов. Такой эксперимент позволяет быстрее и эффективнее решать задачи исследования, но более сложен, требует больших материальных затрат и может помешать нормальному ходу технологического процесса. Иногда отсутствует возможность проведения активного эксперимента (например, при исследовании явлений природы). Тем не менее, учитывая преимущества активного эксперимента, тогда, когда это возможно, предпочтение отдают ему. Теория планирования экспериментов [3, 4] посвящена прежде всего активным экспериментам.

5.3. Влияние выбросов на коэффициент корреляции

Акад. АН СССР С.Н. Бернштейн еще в 1932 г. рассмотрел [5] следующую проблему: "Определить наименьшее возможное значение коэффициента корреляции Пирсона R между величинами X и Y, если известно, что математические ожидания их равны 0 и что существуют две константы L и λ такие, что всегда

$$0 \le \lambda \le \frac{Y}{X} \le L\,.\text{"}$$

Пусть $\sigma^2 = M(X^2),\ \sigma_1^2 = M(Y^2),\ \sigma_1/\sigma = u$. В [5] показано, что минимум коэффициента корреляции R достигается при $u = \sqrt{L\lambda}$ и равен

$$\frac{2\sqrt{L\lambda}}{L+\lambda}\,.$$

Для достижения минимума необходимо и достаточно, чтобы постоянно выполнялось одно из равенств

$$Y - Lx = 0,\ Y - \lambda X = 0\,.$$

Таким образом, минимум R достигается, когда Y есть функция X, которую можно даже предполагать монотонной, если имеем, например,

$$Y = \begin{cases} \lambda X, & |X| < 1, \\ LX, & |X| \ge 1. \end{cases}$$

Рассмотрев численный пример, С.Н. Бернштейн заканчивает статью [5] так: "... достаточно, чтобы только один из 701 индивида не подчинился господствующему закону пропорциональности $Y = 0{,}1X$, чтобы коэффициент корреляции понизился до значения 0,198".

Таким образом, влияние выбросов на коэффициент корреляции может быть весьма велико. Следовательно, перед расчетом коэффициента корреляции необходимо исключить выбросы из выборки. Хорошо известно [1], что обоснованное исключение выбросов может быть

проведено только на основе соображений предметной области, поскольку математико-статистические алгоритмы являются крайне неустойчивыми по отношению к отклонениям от функции распределения, принятой в вероятностно-статистической модели.

5.4. Вздувание коэффициентов корреляции

Это явление обнаружил А.Н. Колмогоров в работе 1933 г. «К вопросу о пригодности найденных статистическим путем формул прогноза» [6]. Предположим, что имеется много наборов предикторов (факторов, признаков). Для каждого из них строится наилучшее приближение отклика с помощью линейной функции от предикторов. Показателем качества приближения служит коэффициент корреляции между откликом и наилучшей линейной функцией от предикторов (в настоящее время чаще используют его квадрат, называемый коэффициентом детерминации). Эффект «вздувания» коэффициента корреляции состоит в том, что при увеличении числа проанализированных наборов предикторов заметно растет максимальный из соответствующих коэффициентов корреляции - показателей качества приближения. Создается впечатление, что тот набор предикторов, на котором достигается рассматриваемый максимум, дает хорошее приближение для отклика. Однако это впечатление развеивается при попытке использовать соответствующую зависимость для прогноза – по новым данным коэффициент корреляции между откликом и ранее найденной линейной функцией от предикторов оказывается значительно меньшим.

В настоящее время весьма популярны методы поиска «наиболее информативного множества признаков» в регрессионном и дискриминантном анализе. Соответствующие алгоритмы, как правило, основаны на переборе большого числа наборов признаков. Поэтому, как показано в [7], актуальность работы А.Н. Колмогорова [6] в настоящее время существенно повысилась. Эффект «вздувания» коэффициента корреляции является одним из проявлений неклассического поведения статистических характеристик в ситуации, когда одна и та же статистическая процедура осуществляется многократно, например, при множественных проверках статистических гипотез [8].

В течение полувека А.Н.Колмогоров интересовался статистическими постановками, в которых число неизвестных параметров растет вместе с объемом данных. К ним относится и кратко рассмотренная выше работа [6]. А в 1970-х годах он стимулировал исследования по т.н. «асимптотике Колмогорова» $p \to \infty,\ n \to \infty,\ p/n \to \lambda$, где p - число параметров, n – объем выборки. Эта асимптотика весьма актуальна как для многомерного статистического анализа [9], так и для статистики объектов нечисловой природы [10], а также для задач статистического приемочного контроля [11].

5.5. Коэффициент детерминации

Как уже отмечалось, для модели линейной регрессии с одним признаком (фактором) *X* коэффициент детерминации равен квадрату линейного парного коэффициента корреляции Пирсона между *X* и откликом *Y*. Необходимо подчеркнуть, что такая интерпретация корректна только тогда, когда анализируемые данные являются выборкой из двумерного распределения. Чуть подробнее: исходные данные рассматриваются как независимые одинаково распределенные случайные вектора. Отсюда следует, что если фактор *X* детерминирован (например, время), то коэффициент детерминации не является квадратом коэффициента корреляции, поскольку понятие коэффициента корреляции для подобной постановки не определено. Следовательно, коэффициент детерминации не является показателем качества зависимости, построенной с помощью метода наименьших квадратов.

Распространенная ошибка состоит в использовании коэффициента детерминации для оценки качества восстановления зависимости методом наименьших квадратов. Часто заявляют, что близость к 1 коэффициента детерминации свидетельствует об успешном восстановлении зависимости. При этом взгляд на данные (на корреляционное поле) может дать совершенно иной вывод. Например, все точки, кроме одной, лежат в небольшой по диаметру области и вытянуты вдоль гиперболы. Оставшаяся точка расположена далеко вправо вверху. Формальное применение метода наименьших квадратов приводит к тому, что единственный "выброс" меняет гиперболу на возрастающую линейную зависимость (сопоставьте с примером С.Н. Бернштейна, рассмотренным выше в п.4).

Формально рассчитанный коэффициент детерминации в рассматриваемой постановке может быть сколь угодно близким к 1. Однако использование этого факта для обоснования утверждения о высоком качестве восстановления зависимости скорее всего является примером неверной интерпретации. Во-первых, из-за неисключенных выбросов. Во-вторых, из-за нарушения предпосылок вероятностно-статистической модели выборки (если фактор *X* детерминирован).

Практическая рекомендация состоит в предварительном проведении отбраковки "выбросов" и проверке выполнения предпосылок вероятностно-статистической модели.

5.6. Многообразие моделей и методов регрессионного анализа

За столетия разработки математических методов исследования накоплен огромный массив научных результатов. Так, еще 30 лет назад мы оценивали [14] число статей и книг в этой области как 10^6, в том числе актуальных для современных исследователей - как 10^5. Сколькими

статьями и книгами может овладеть один человек? Для большинства - 10^3, для отдельных наиболее продвинутых лиц - 10^4, что на порядки меньше, чем объем накопленных научных результатов. Следовательно, необходимы работы по упорядочению накопленных научных результатов. Для успешной работы важно единообразное понимание терминов. Необходимо знание фактов и тенденций развития. Обсудим эти вопросы на примере научной области "модели регрессионного анализа (восстановления зависимостей)" с целью сформировать единую методологическую базу для обсуждения различных частных вопросов этой области. Рассмотрим четыре метода восстановления зависимости.

В простейшем случае есть одна независимая количественная переменная t и одна зависимая количественная переменная x. Требуется указать (как говорят, восстановить) функцию, описывающую зависимость x от t.

В простейшем случае принимают, что эта зависимость - линейная: $x(t) = at + b$. Исходные данные - набор n двумерных векторов ж Предполагается, что имеются отклонения от линейности, т.е. $x_i = at_i + b + e_i$, где e_i, $i = 1, 2, ..., n$, - погрешности (отклонения, невязки). Необходимо оценить неизвестные параметры a и b.

Как известно, оценивание можно провести разными способами. Есть графический метод. Он состоит в том, что точки (t_i, x_i), $i = 1, 2, ..., n$, надо нанести на плоскость и провести с помощью линейки прямую линию, наилучшим образом приближающую эти точки (можно использовать миллиметровую бумагу или опцию "Корреляционное поле" в программном продукте для работы с электронными таблицами EXCEL). Недостатки - субъективизм и невозможность указать точность оценивания зависимости и ее параметров.

Чаще используют расчетные методы. Основная идея состоит в том, чтобы минимизировать одновременно все отклонения $x_i - at_i - b$. Реализовать эту идею можно различными способами. В методе наименьших модулей минимизируют по a и b функцию

$$g(a,b) = \sum_{i=1}^{n} | x_i - at_i - b | .$$

В методе минимакса в качестве показателя суммарного отклонения вместо суммы модулей минимизируют максимальное отклонение

$$h(a,b) = \max_{1 \le i \le n} | x_i - at_i - b | .$$

В 1794 г. К. Гаусс разработал метод наименьших квадратов, основанный на минимизации функции

$$f(a,b) = \sum_{i=1}^{n} (x_i - at_i - b)^2 .$$

двух переменных a и b.

Метод наименьших квадратов выглядит менее естественным, чем метод наименьших квадратов и метод минимакса. Действительно, почему квадрат, а не другая степень? Однако используют и применяют именно метод наименьших квадратов, а остальные два метода - маргинальные, ими занимаются отдельные энтузиасты. Почему в конкурентной борьбе победил именно метод наименьших квадратов? По нашему мнению, дело в том, что оценки параметров a и b метода наименьших квадратов, полученные в результате минимизации $f(a, b)$, задаются элементарными формулами (см., например, [1]), в то время как оценки параметров для двух других методов могут быть найдены лишь с помощью численных алгоритмов [15]. Причина сказанного в том, что для минимизации $f(a, b)$ можно использовать частные производные этой функции по параметрам a и b, в то время как $g(a, b)$ и $h(a, b)$ не дифференцируемы из-за наличия в них модуля. Наличие точных формул не только облегчает вычисление оценок метода наименьших квадратов, но и позволяет глубоко изучить свойства этих оценок.

В проведенных рассуждениях не было никаких вероятностно-статистических моделей. Действительно, метод наименьших квадратов и другие ранее упомянутые методы можно рассматривать в рамках теории приближений. Однако, если целесообразно перенести выводы с набора точек (t_i, x_i), $i = 1, 2, ..., n$, на более широкую совокупность, то необходимо ввести вероятностно-статистические модели, нацеленные на переход от выборки к генеральной совокупности.

Рассмотрим два основных типа вероятностно-статистических моделей.

5.7. Модели с детерминированной независимой переменной

Широко применяются модели с детерминированной независимой количественной переменной t. Для зависимой количественной переменной x случайность вводится с помощью равенств $x_i = at_i + b + e_i$, в правой части которых стоят случайные погрешности (отклонения, невязки) $e_1, e_2, ..., e_n$. Отличительная черта этого типа моделей состоит в том, что независимая переменная является детерминированной, а зависимая - случайной.

В базовой модели случайные величины $e_1, e_2, ..., e_n$. предполагаются независимыми и одинаково распределенными. Каково их общее распределение? В устаревших литературных источниках часто принимают, что их распределение является нормальным (гауссовским). Однако хорошо известно, что практически все распределения реальных данных не являются нормальными [1, 16]. Поэтому согласно новой парадигме математической статистики [17] следует считать распределение случайные величины $e_1, e_2, ..., e_n$ произвольным, с одним ограничением - для получения предельных распределений оценок параметров и значений

задающей зависимость функции целесообразно предположить выполнение условий центральной предельной теоремы.

Согласно [18] модель восстановления зависимости с независимыми одинаково распределенными случайными погрешностями, имеющими распределения произвольного вида, называется непараметрической. Именно ее следует использовать на практике, поскольку параметрическая модель регрессионного анализа, особенно с нормальными ошибками, не соответствует реальности. Здесь под параметрической моделью понимают модель, в которой распределения погрешностей принадлежат тому или иному параметрическому семейству - некоторому подсемейству четырехпараметрического семейства К. Пирсона [19]. Если в описании алгоритма регрессионного анализа используются распределения Стьюдента или Фишера, то необходимо констатировать, что распределения погрешностей предполагаются нормальными, следовательно, алгоритм не соответствует новой парадигме математической статистики. Отметим, что при непараметрической модели погрешностей сама зависимость может являться параметрической, например, линейной. Как показано в дальнейшем, есть много вариантов постановки задач непараметрической регрессии.

Простейшая модель обобщается в двух направлениях - переход от линейной модели к более общей параметрической зависимости и отказ от независимости и одинаковой распределенности погрешностей. Параметрическая зависимость должна быть линейной по параметрам. Например, типовой является зависимость

$$x_i = a_1 f_1(t_i) + a_2 f_2(t_i) + ... + a_m f_m(t_i) + e_i,\ i = 1, 2, ... , n, \qquad (1)$$

где функции $f_1(t)$, $f_2(t)$, ... , $f_m(t)$ заданы, а параметры a_1, a_2, ... , a_m подлежат оценке методом наименьших квадратов. В частном случае, когда $f_k(t) = t^{k-1}$, $k = 1, 2, ..., m$, зависимость (1) является многочленом. Если же зависимость не является линейной по параметрам, то минимизацию в методе наименьших квадратов можно провести лишь численно, а теоретическое изучение свойств оценок встречает сложности.

Переход от одной независимой переменной к нескольким не представляет методологических сложностей.

Много постановок порождает отказ от независимости и одинаковой распределенности погрешностей. Например, дисперсии независимых погрешностей могут зависеть от независимой переменной t, например, линейно. Тогда абсолютные отклонения в методе наименьших квадратов заменяют относительными. Отказ от независимости погрешностей приводит к более сложным моделям, поскольку зависимость можно моделировать многими способами. Наиболее простой является модель, в которой все пары погрешностей имеют одинаковые коэффициенты корреляции. В рассматриваемой области необходимы новые исследования.

5.8. Модели анализа случайных векторов

Второй основной тип вероятностно-статистических моделей основан на выборке случайных векторов. В таких моделях исходные данные в простейшем случае - двумерные случайные вектора $(x_i(\omega), y_i(\omega)), i = 1,2,...,n,$ определенные на одном и том же вероятностном пространстве. В базовой модели все эти случайные вектора независимы и одинаково распределены с вектором $(x(\omega), y(\omega))$. В качестве оцениваемой зависимости рассматривают условное математическое ожидание $y(\omega)$ при условии заданного значения $x(\omega)$.

Пусть случайный вектор $(x(\omega), y(\omega))$ имеет плотность $p(x, y)$. Как известно из теории вероятностей, плотность условного распределения $y(\omega)$ при условии $x(\omega) = x_0$ имеет вид

$$p(y \mid x) = p(y \mid x(\omega) = x_0) = \frac{p(x, y)}{\int\limits_{-\infty}^{+\infty} p(x, y)dy}.$$

Условное математическое ожидание, т.е. регрессионная зависимость y от x, имеет вид

$$f(x) = \int\limits_{-\infty}^{+\infty} yp(y \mid x)dy = \frac{\int\limits_{-\infty}^{+\infty} yp(x, y)dy}{\int\limits_{-\infty}^{+\infty} p(x, y)dy}.$$

Таким образом, для нахождения оценок регрессионной зависимости достаточно найти оценки совместной плотности распределения вероятности $p_n(x, y)$ такие, что

$$p_n(x, y) \to p(x, y)$$

при $n \to \infty$. Тогда непараметрическая оценка регрессионной зависимости

$$f_n(x) = \frac{\int\limits_{-\infty}^{+\infty} yp_n(x, y)dy}{\int\limits_{-\infty}^{+\infty} p_n(x, y)dy}$$

при $n \to \infty$ является состоятельной оценкой регрессии как условного математического ожидания, т.е.

$$f_n(x) \to f(x).$$

Общий подход к построению непараметрических оценок плотности распределения вероятностей в пространствах различной природы развит в ряде публикаций (см., например, [1]), крайняя по времени статья [20].

Таким образом, если выборка $(x_i(\omega), y_i(\omega)), i = 1,2,...,n,$ состоит из случайных векторов, то базовая модель восстановления зависимости является двойной непараметрической, т.е. зависимость является непараметрической и распределение двумерного вектора является

произвольным. Как уже отмечалось, принимать гипотезу многомерной нормальности нет оснований. В некоторых случаях полезны параметрические модели зависимости, например,

$$y = b_1\varphi_1(x) + b_2\varphi_2(x) + \ldots + b_m\varphi_m(x) + e_y. \qquad (2)$$

где функции $\varphi_1(x)$, $\varphi_2(x)$, ..., $\varphi_m(x)$ заданы, а параметры b_1, b_2, ... , b_m подлежат оценке методом наименьших квадратов. В отличие от (1), в правой части (2) все слагаемые - случайные величины.

Итак, две основные модели основаны на детерминированной независимой переменной и выборке случайных векторов соответственно. Хотя расчетные алгоритмы метода наименьших квадратов во многом совпадают, но интерпретации результатов расчетов могут различаться. Так, об оценке дисперсии независимой переменной можно говорить только в модели на основе выборки случайных векторов, равно как и о коэффициенте детерминации как критерии качества модели. В случае модели на основе детерминированной независимой переменной попытки применять коэффициенте детерминации в качестве критерия качества модели могут привести к грубым ошибкам.

5.9. Сглаживание временных рядов

С формальной точки зрения временные ряды являются частным случаем моделей с детерминированной независимой переменной, в качестве которой рассматривается время t. При этом для зависимой переменной $X(t)$ часто рассматривают аддитивную модель

$$X(t) = T(t) + P(t) + R(t), \qquad (3)$$

где $T(t)$ - тренд, задающий центральную тенденцию, $P(t)$ - периодическая составляющая, $R(t)$ - случайная составляющая. Иногда рассматривают мультипликативную модель $X(t) = T(t)\ P(t)\ R(t)$, однако она не имеет самостоятельного значения, поскольку после логарифмирования переходит в модель (3) для логарифмов включенных в модель составляющих.

Для модели (3) рассматривают различные варианты непараметрики. Например, тренд $T(t)$ может задаваться линейной функцией, а периодическая составляющая $P(t)$ быть произвольной. Методы непараметрического оценивания периодической составляющей для такой модели разработаны в [21].

От независимости отклонений приходится отказаться при движении от дискретного времени к непрерывному. В пределе отклонения моделируются случайным процессом с непрерывными траекториями. Так поступают при моделировании динамики курсов акций и валют. Математическая теория оценивания в случае непрерывных случайных процессов существенно отличается от таковой в случае выборок погрешностей.

5.10. Методы восстановления зависимостей в пространствах общей природы

Обсудим модели регрессионного анализа в общем виде. Сначала рассмотрим параметрические постановки задач регрессионного анализа (восстановления зависимостей) в пространствах произвольной природы, затем — непараметрические, после чего перейдем к оцениванию нечисловых параметров в классической ситуации, когда отклик и факторы принимают числовые значения.

Задача аппроксимации зависимости (параметрической регрессии). Пусть X и Y — некоторые пространства. Пусть имеются статистические данные — n пар (x_k, y_k), где $x_k \in X$, $y_k \in Y$, $k = 1, 2, \ldots, n$. Задано параметрическое пространство Θ произвольной природы и семейство функций $g(x,\theta)$: $X \times \Theta \to Y$. Требуется подобрать параметр $\theta \in \Theta$ так. чтобы $g(x_k, \theta)$ наилучшим образом приближали y_k, $k = 1, 2, \ldots, n$. Пусть f_k — последовательность показателей различия в Y. При сделанных предположениях параметр θ естественно оценивать путем решения экстремальной задачи:

$$\theta_n = Arg \min_{\theta \in \Theta} \sum_{k=1}^{n} f_k(g(x_k, \theta), y_k). \qquad (4)$$

Часто, но не всегда, все f_k совпадают. В классической постановке, когда $X = R^k$, $Y = R^1$, функции f_k различны при неравноточных наблюдениях, например, когда число опытов меняется от одной точки x проведения опытов к другой.

Если $f_k(y_1, y_2) = f(y_1, y_2) = (y_1 - y_2)^2$, то получаем общую постановку метода наименьших квадратов:

$$\theta_n = Arg \min_{\theta \in \Theta} \sum_{k=1}^{n} (g(x_k, \theta) - y_k)^2.$$

В рамках детерминированного анализа данных остается единственный теоретический вопрос — о существовании θ_n. Если все участвующие в формулировке задачи (4) функции непрерывны, а минимум берется по бикомпакту, то θ_n существует. Есть и иные условия существования θ_n [10].

При появлении нового наблюдения x в соответствии с методологией восстановления зависимости рекомендуется выбирать оценку соответствующего y по правилу

$$y^* = g(x, \theta_n).$$

Обосновать такую рекомендацию в рамках детерминированного анализа данных невозможно. Это можно сделать только в вероятностной теории, равно как и изучить асимптотическое поведение θ_n, доказать состоятельность этой оценки.

Как и в классическом случае, вероятностную теорию целесообразно строить для трех различных постановок.

1. Переменная x — детерминированная (например, время), переменная y — случайная, ее распределение зависит от x.

2. Совокупность (x_k, y_k), $k = 1, 2, \ldots, n$, — выборка из распределения случайного элемента со значениями в $X \times Y$.

3. Имеется детерминированный набор пар (x_{k0}, y_{k0}), $k = 1, 2, \ldots, n$, результат наблюдения (x_k, y_k) является случайным элементом, распределение которого зависит от (x_{k0}, y_{k0}). Это — постановка т.н. конфлюэнтного анализа.

Во всех трех случаях

$$f_n(\omega, \theta) = \sum_{k=1}^{n} f_k(g(x_k, \theta), y_k),$$

однако случайность входит в правую часть по-разному в зависимости от постановки, от которой зависит и определение предельной функции $f(\theta)$.

Проще всего выглядит $f(\theta)$ в случае второй постановки при $f_k \equiv f$:

$$f(\theta) = Mf(g(x_1, \theta), y).$$

В случае первой постановки

$$f(\theta) = \lim_{n\to\infty} \sum_{k=1}^{n} Mf_k(g(x_k, \theta), y_k(\omega))$$

в предположении существования указанного предела. Ситуация усложняется для третьей постановки:

$$f(\theta) = \lim_{n\to\infty} \sum_{k=1}^{n} Mf_k(g(x_k(\omega), \theta), y_k(\omega)).$$

Во всех трех случаях на основе общих результатов о поведении решений экстремальных статистических задач можно изучить асимптотику оценок θ_n методами нечисловой статистики [10]. При выполнении соответствующих внутриматематических условий регулярности оценки оказываются состоятельными, т.е. удается восстановить зависимость.

Аппроксимация и регрессия. Соотношение (4) дает решение задачи аппроксимации. Поясним, как эта задача соотносится с нахождением регрессии. Согласно [10] для случайной величины (ξ, η) со значениями в $X \times Y$ регрессией η на ξ относительно меры близости f естественно назвать решение задачи

$$Mf(g(\xi), \eta) \to \min_{g}, \qquad (5)$$

где $f: Y \times Y \to R^1$, $g: X \to Y$, минимум берется по множеству всех измеримых функций.

Можно исходить и из формально другого определения. Для каждого $x \in X$ рассмотрим случайную величину $\eta(x)$, распределение которой является условным распределением η при условии $\xi = x$. В соответствии с определением математического ожидания в пространстве общей природы назовем условным математическим ожиданием решение экстремальной задачи

$$M(\eta \mid \xi = x) = Arg \min\{Mf(y, \eta(x)), y \in Y\}.$$

Оказывается, при обычных предположениях измеримости решение задачи (5) совпадает с $M(\eta \mid \xi = x)$. (Внутриматематические уточнения типа «равенство имеет место почти всюду» здесь опущены.)

Если заранее известно, что условное математическое ожидание $M(\eta \mid \xi = x)$ принадлежит некоторому параметрическому семейству $g(x, \theta)$, то задача нахождения регрессии сводится к оцениванию параметра θ в соответствии с рассмотренной выше второй постановкой вероятностной теории параметрической регрессии.

Если же нет оснований считать, что регрессия принадлежит некоторому параметрическому семейству, можно использовать непараметрические оценки регрессии. Они строятся с помощью непараметрических оценок плотности [1, 20].

Непараметрические методы восстановления зависимости. Пусть ν_1 — мера в X, ν_2 — мера в Y, а их прямое произведение $\nu = \nu_1 \times \nu_2$ — мера в $X \times Y$. Пусть $g(x, y)$ — плотность случайного элемента (ξ, η) по мере ν. Тогда условная плотность $g(y|x)$ распределения η при условии $\xi = x$ имеет вид

$$g(y \mid x) = \frac{g(x, y)}{\int\limits_Y g(x, y)\nu_2(dy)} \tag{6}$$

(в предположении, что интеграл в знаменателе отличен от 0). Следовательно,

$$Mf(y, \eta(x)) = \int\limits_Y f(y, a)g(a \mid x)\nu_2(da),$$

а потому

$$M(\eta \mid \xi = x) = Arg \min_{y \in Y} Mf(y, \eta(x)) =$$

$$= Arg \min_{y \in Y} \int\limits_Y f(y, a)g(a \mid x)\nu_2(da).$$

Заменяя $g(x,y)$ в (6) непараметрической оценкой плотности $g_n(x,y)$, получаем оценку условной плотности

$$g_n(y \mid x) = \frac{g_n(x, y)}{\int\limits_Y g_n(x, y)\nu_2(dy)}. \tag{7}$$

Если $g_n(x,y)$ — состоятельная оценка $g(x,y)$, то числитель (7) сходится к числителю (6). Сходимость знаменателя (7) к знаменателю (6) обосновывается с помощью предельной теории статистик интегрального типа [12]. В итоге получаем утверждение о состоятельности непараметрической оценки (7) условной плотности (6).

Непараметрическая оценка регрессии ищется как

$$M_n(\eta \mid \xi = x) = Arg \min_{y \in Y} \int\limits_Y f(y, a)g_n(a \mid x)\nu_2(da).$$

Состоятельность этой оценки следует из приведенных выше общих результатов об асимптотическом поведении решений экстремальных статистических задач.

5.11. Оценивание объектов нечисловой природы в классических постановках регрессионного анализа

Нечисловая статистика тесно связана с классическими областями прикладной статистики. Ряд трудностей в классических постановках удается понять и разрешить лишь с помощью общих результатов прикладной статистики. В частности, это касается оценивания параметров, когда параметр имеет нечисловую природу.

Рассмотрим типовую прикладную постановку задачи восстановления регрессионной зависимости, линейной по параметрам. Исходные данные имеют вид $(x_i, y_i) \in R^2,\ i = 1, 2, ..., n$. Цель состоит в том, чтобы с достаточной точностью описать y как многочлен (полином) от x, т.е. модель имеет вид

$$y_i = \sum_{k=0}^{m} a_k x_i^k + \varepsilon_i, \quad i = 1, 2, ..., n, \tag{8}$$

где m — неизвестная степень полинома; a_0, a_1, a_2, …, a_m — неизвестные коэффициенты многочлена; $\varepsilon_i,\ i = 1,\ 2, ..., n$, — погрешности, которые для простоты примем независимыми и имеющими одно и то же нормальное распределение с нулевым математическим ожиданием и дисперсией σ^2.

В прикладной статистике часто используют следующую технологию анализа данных. Сначала пытаются применить модель (8) для линейной функции (m = 1), при неудаче (неадекватности модели) переходят к многочлену второго порядка (m = 2), если снова неудача, то берут модель (8) с m = 3 и т.д. Адекватность модели обычно проверяют по F-критерию Фишера, *основанному на предположении нормальности погрешностей*.

Обсудим свойства этой процедуры. Если степень полинома задана ($m = m_0$), то его коэффициенты оценивают методом наименьших квадратов, свойства этих оценок хорошо известны. Однако в рассматриваемой постановке m тоже является неизвестным параметром и подлежит оценке. Таким образом, требуется оценить объект (m, a_0, a_1, a_2, …, a_m), множество значений которого можно описать как $R^1 \cup R^2 \cup R^3 \cup \ldots$. Это — объект нечисловой природы, обычные методы оценивания для него неприменимы. Разработанные к настоящему времени методы оценивания степени полинома носят в основном эвристический характер (см., например, гл. 12 монографии [22]). Рассмотрим некоторые из них.

Замечание. Здесь наглядно проявляется одна из причин живучести вероятностно-статистических моделей на основе нормального распределения. Такие модели, как правило, не адекватны реальной ситуации, о чем сказано выше. Однако с математической точки зрения они позволяют глубже проникнуть в суть изучаемого явления. Поэтому такие модели полезны для первоначального анализа ситуации. В ходе дальнейших исследований необходимо снять нереалистическое предположение нормальности и перейти к непараметрическим моделям.

Оценивание степени полинома. Полезно рассмотреть основной показатель качества регрессионной модели (8). Одни и те же данные можно обрабатывать различными способами. На первый взгляд, показателем отклонений данных от модели может служить остаточная сумма квадратов *SS*. Чем этот показатель меньше, тем приближение лучше, значит, и модель лучше описывает реальные данные. Однако это рассуждение годится только для моделей с одинаковым числом параметров. Ведь если добавляется новый параметр, по которому можно минимизировать, то и минимум, как правило, оказывается меньше.

В качестве основного показателя качества регрессионной модели используют следующую оценку остаточной дисперсии

$$\tilde{\sigma}^2(m) = \frac{SS}{n-m-1}.$$

Таким образом, вводят корректировку на число параметров, оцениваемых по наблюдаемым данным. Корректировка состоит в уменьшении знаменателя на указанное число. В модели (8) это число равно (m + 1). В случае задачи восстановления линейной функции одной переменной оценка остаточной дисперсии имеет вид

$$\hat{\sigma}^2 = \frac{SS}{n-2},$$

поскольку число оцениваемых параметров $m + 1 = 2$.

Еще раз — почему *при подборе вида модели* знаменатель дроби, оценивающей остаточную дисперсию, приходится корректировать на число параметров? Если этого не делать, то придется заключить, что всегда многочлен второй степени лучше соответствует данным, чем линейная функция, многочлен третьей степени лучше приближает исходные данные, чем многочлен второй степени, и т.д. В конце концов доходим до многочлена степени (n – 1) с n коэффициентами, который проходит через все заданные точки. Но его прогностические возможности, скорее всего, существенно меньше, чем даже у линейной функции. *Излишнее усложнение статистических моделей вредно.*

Типовое поведение скорректированной оценки $v(m) = \tilde{\sigma}^2(m)$ остаточной дисперсии в случае расширяющейся системы моделей (т.е. при возрастании натурального параметра m) выглядит так. Сначала наблюдаем заметное убывание. Затем оценка остаточной дисперсии колеблется около некоторой константы (дисперсии погрешности). Поясним ситуацию на примере модели восстановления зависимости, выраженной многочленом:

$$x(t) = a_0 + a_1 t + a_2 t^2 + a_3 t^3 + \ldots + a_m t^m.$$

Пусть эта модель справедлива при $m = m_0$. При $m < m_0$ в скорректированной оценке остаточной дисперсии учитываются не только погрешности измерений, но и соответствующие (старшие) члены многочлена (предполагаем, что коэффициенты при них отличны от 0). При $m \geq m_0$ имеем

$$\lim_{n\to\infty} v(m) = \sigma^2.$$

Следовательно, скорректированная оценка остаточной дисперсии будет колебаться около указанного предела. Поэтому представляется естественным, что в качестве оценки неизвестной статистику степени многочлена (полинома) можно использовать первый локальный минимум скорректированной оценки остаточной дисперсии, т.е.

$$m^* = \min\{m : v(m-1) > v(m), \quad v(m) \le v(m+1)\}.$$

В работе [23] найдено предельное распределение этой оценки параметра, принимающего целые значения - степени многочлена.

Теорема. При справедливости некоторых условий регулярности

$$\lim_{n\to\infty} P(m^* < m_0) = 0, \ \lim_{n\to\infty} P(m^* = m_0 + u) = \lambda(1-\lambda)^u,$$
$$u = 0,1,2,....,$$

где

$$\lambda = \Phi(1) - \Phi(-1) = \frac{1}{\sqrt{2\pi}} \int_{-1}^{1} \exp\left\{-\frac{x^2}{2}\right\} dx \approx 0,68268.$$

Таким образом, предельное распределение оценки m^* степени многочлена (полинома) является геометрическим. Это означает, в частности, что оценка не является состоятельной. При этом вероятность получить меньшее значение, чем истинное, исчезающе мала. Далее имеем:

$$P(m^* = m_0) \to 0,68268, \ P(m^* = m_0 + 1) \to 0,68268(1 - 0,68268) = 0,21663,$$

$$P(m^* = m_0 + 2) \to 0,68268(1 - 0,68268)^2 = 0,068744,$$

$$P(m^* = m_0 + 3) \to 0,68268(1 - 0,68268)^3 = 0,021814...$$

Разработаны и иные методы оценивания неизвестной степени многочлена, например, путем многократного применения процедуры проверки адекватности регрессионной зависимости с помощью критерия Фишера. Предельное поведение таких оценок — таково же, как в приведенной выше теореме, только значение параметра λ иное. Для степени многочлена давно предложены состоятельные оценки [24]. Для этого достаточно уровень значимости (при проверке адекватности регрессионной зависимости с помощью критерия Фишера) сделать убывающим при росте объема выборки.

Построение информативного подмножества признаков. В более общем случае многомерной линейной регрессии данные имеют вид (y_i, X_i), $i = 1, 2, \ldots, n$, где $X_i = (x_{i1}, x_{i2}, \ldots, x_{iN}) \in R^N$ — вектор предикторов (факторов, объясняющих переменных), а модель такова:

$$y_i = \sum_{j\in K} a_j x_{ij} + \varepsilon_i, \quad i = 1,2,...,n \qquad (9)$$

(здесь K — некоторое подмножество множества $\{1, 2, \ldots, n\}$; ε_i — те же, что и в модели (8); a_j — неизвестные коэффициенты при предикторах с номерами из K). Множество K называют *информативным подмножеством признаков*, поскольку согласно формуле (9) остальные признаки можно отбросить без потери информации. Проблема состоит в

том, что при анализе реальных данных неизвестно, какие признаки входят в K, а какие нет. Ясна важность оценивания информативного подмножества признаков.

Модель (8) сводится к модели (9), если

$$x_{i1} = 1,\ x_{i2} = x_i,\ x_{i3} = x_i^2,\ x_{i4} = x_i^3,\ ..., x_{ij} = x_i^{j-1},\ ...$$

В модели (8) есть естественный порядок ввода предикторов в рассмотрение — в соответствии с возрастанием степени многочлена, а в модели (9) естественного порядка нет, поэтому здесь приходится рассматривать произвольное подмножество множества предикторов. Есть только частичный порядок — чем мощность подмножества меньше, тем лучше. Модель (9) особенно актуальна в технических исследованиях (см. многочисленные примеры в журнале «Заводская лаборатория. Диагностика материалов»). Она применяется в задачах управления качеством продукции и других технико-экономических исследованиях, в медицине, экономике, маркетинге и социологии, когда из большого числа факторов, предположительно влияющих на изучаемую переменную, надо отобрать по возможности наименьшее число значимых факторов и с их помощью сконструировать прогнозирующую формулу (9).

Задача оценивания модели (9) разбивается на две последовательные задачи: оценивание множества K — подмножества множества всех предикторов, а затем — неизвестных параметров a_j. Методы решения второй задачи хорошо известны и подробно изучены (обычно используют метод наименьших квадратов). Гораздо хуже обстоит дело с оцениванием объекта нечисловой природы K. Существующие методы — в основном эвристические, они зачастую не являются даже состоятельными. Даже само понятие состоятельности в данном случае требует специального определения.

Определение. Пусть K_0 — истинное подмножество предикторов, т.е. подмножество, для которого справедлива модель (9), а подмножество предикторов K_n — его оценка. Оценка K_n называется состоятельной, если

$$\lim_{n\to\infty} Card(K_n \Delta K_0) = 0,$$

где Δ — символ симметрической разности множеств; $Card(K)$ означает число элементов множества K, а предел понимается в смысле сходимости по вероятности.

Задача оценивания в моделях регрессии, таким образом, разбивается на две — оценивание структуры модели и оценивание параметров при заданной структуре. В модели (8) структура описывается неотрицательным целым числом m, в модели (9) — множеством K. Структура — объект нечисловой природы. Задача ее оценивания сложна, в то время как задача оценивания численных параметров при заданной структуре хорошо изучена, разработаны эффективные (в смысле прикладной математической статистики) методы. Такова же ситуация и в других методах многомерного

статистического анализа — в факторном анализе (включая метод главных компонент) и в многомерном шкалировании, в иных оптимизационных постановках проблем прикладного многомерного статистического анализа.

Множество K и параметры a_j линейной зависимости можно оценивать путем решения задачи оптимизации

$$\sum_{i=1}^{n}\left(y_i - \sum_{j\in K} a_j x_{ij}\right)^2 \to \min, \qquad (10)$$

в которой минимум берется по K, a_j, $j \in K$. Математическая природа множества, по которому проводится минимизация, весьма сложна. Это и объясняет тот факт, что к настоящему времени разработано много эвристических методов оценивания информативного множества параметров K, свойства которых плохо изучены. На основе общих результатов нечисловой статистики об асимптотическом поведении решений экстремальных статистических задач удалось показать, что оценки, полученные путем решения задачи (7), являются состоятельными (см., например, [7]).

К рассматриваемой тематике относится также эффект "вздувания коэффициентов корреляции", рассмотренный выше в п.5.

В соответствии с рекомендациями теории устойчивости статистических выводов при допустимых отклонениях исходных данных и предпосылок рассматриваемой вероятностно-статистической или иной модели задача восстановления зависимости должна рассматриваться в рамках системы моделей, а не лишь одной модели.

5.12. Регрессионный анализ интервальных данных

Иногда рассматривают модели, в которых как входная, так и выходная переменные имеют погрешности, определяемые значениями этих переменных. В простейшем случае вместо "истинных" данных (t_i, x_i), $i = 1, 2, ..., n$, наблюдают данные с погрешностями (q_i, y_i), $i = 1, 2, ..., n$. где q_i, $= t_i + \varepsilon_i$, $y_i = x_i + \delta_i$. Здесь ε_i и δ_i - погрешности измерений (наблюдений, регистрации, опытов, анализов). Требуется восстановить зависимость между "истинными" переменными t и x.

Есть несколько подходов к решению. этой задачи. Если заданы ограничения на значения погрешностей, наложенных на случайные величины, то плодотворен подход разработанной нами статистики интервальных данных [25]. Восстановлению линейной зависимости в соответствии с подходом статистики интервальных данных посвящена статья [26]. Подробному изложению статистики интервальных данных посвящены развернутые главы в монографиях [1, 10, 27, 28].

Уходит в прошлое подход т.н. конфлюэнтного анализа, согласно которому погрешности измерений ε_i и δ_i имеют нормальные распределения. Поскольку, как уже отмечалось, распределения

практически всех реальных величин не являются нормальными, конфлюэнтный анализ не является адекватным реальным ситуациям и потому не имеет практических перспектив. Точно также распределения Стьюдента и Фишера не адекватны реальности и могут иметь лишь теоретическое значение. Вместе с тем отметим, что, например, неизвестен непараметрический аналог критерия Фишера, предназначенного для проверки адекватности регрессионной модели (скажем, для проверки адекватности линейной модели, когда альтернативой является квадратическая).

5.13. Заключительные замечания

Как уже отмечалось [12], основная проблема современной науки - всеобщее невежество научных работников. Мы постарались показать, что нельзя бездумно применять распространенные программные продукты (ср. [13]). Необходимо владеть основами прикладной статистики. Иначе вместо обоснованных результатов статистического анализа данных можно получить ошибочные заключения.

Отметим, что многие важные результаты (в частности, принадлежащие А.Н. Колмогорову и С.Н. Бернштейну) были получены много десятилетий назад. Следовательно, грубо ошибочна встречающаяся иногда ориентация исследователей и редакций научных журналов только на публикации последних 5 лет.

Анализ многообразия моделей регрессионного анализа приводит к выводу, что не существует единой "стандартной модели". В каждом конкретном случае необходимо описывать используемую модель (а лучше - систему моделей) и обосновывать ее.

Исследования в рассматриваемой области прикладной статистики ведутся активно, но много задач всё еще требует решения. Некоторые такие задачи отмечены выше. Например, разработанные в XX в. модели и методы, основанные на предположении нормальности, требуют осмысления и доработки (как теоретической, так и алгоритмической) с позиций непараметрической статистики. Критический разбор устоявшихся взглядов необходим для квалифицированного развития и применения математических методов исследования, в частности, для перехода на современную парадигму математической статистики [17].

ГЛАВА 6. ОЦЕНИВАНИЕ РАЗМЕРНОСТИ ВЕРОЯТНОСТНО-СТАТИСТИЧЕСКОЙ МОДЕЛИ

По статистическим данным необходимо оценивать две составляющие вероятностно-статистических моделей - структуру моделей и параметры. Методы расчета состоятельных оценок параметров хорошо известны (например, применяют метод одношаговых оценок, который пришел на смену методу максимального правдоподобия). Структура модели обычно выбирается исследователем (можно сказать, что используются экспертные методы). Некоторые параметры структуры можно оценивать с помощью математико-статистических методов. Например, степень многочлена в регрессионной зависимости или число слагаемых в модели смеси распределений, используемой для классификации. Для подобных параметров модели используется общий термин - размерность вероятностно-статистической модели. Более общая составляющая модели - информативное подмножество признаков.

6.1. О содержании этой главы

В настоящей главе рассмотрено асимптотическом поведении оценок размерностей ряда моделей. Изучено асимптотическое поведение ряда оценок степени полинома при восстановлении зависимости. Получены состоятельные оценки размерности и структуры модели в регрессии. Рассмотрены подходы к оцениванию числа элементов смеси в задачах классификации. Обсуждаются оценки размерности модели в факторном анализе и многомерном шкалировании. С целью обоснования последовательного выполнения этапов статистического анализа данных анализируются проблемы "стыковки" алгоритмов классификации и регрессии. Полезными оказываются оптимизационные формулировки ряда задач прикладной статистики. Основные результаты касаются состоятельности оценок. Краткие формулировки ряда теорем содержатся в ранее вышедших публикациях. Проблема оценивания размерности вероятностно-статистической модели как самостоятельное направление прикладной статистики впервые в монографическом издании подробно рассмотрена здесь. Публикуются доказательства включенных в настоящую главу теорем. Эти доказательства и являются основными научными результатами этой главы.

6.2. Асимптотическое поведение ряда оценок степени полинома в регрессии

Во многих прикладных задачах требуется установить зависимость переменной y от переменных x_1, x_2, ..., x_m. Простейшая вероятностно-статистическая модель имеет вид

$$y = a_1 x_1 + a_1 x_1 + ... + a_m x_m + \varepsilon , \qquad (1)$$

где a_j - коэффициенты линейной регрессии, j = 1, 2, ..., m, а ε - остаточный член, рассматриваемый обычно как погрешность измерения или результат влияния неучтенных факторов.

Исходные данные для определения (т.е. оценивания) коэффициентов регрессии имеют вид

$$(y_i, x_{1i}, x_{2i}, ..., x_{mi}),\ i = 1, 2, ..., n, \qquad (2)$$

Рассмотрим модель с детерминированными x_{ji}, $j = 1, 2, ..., m$, $i = 1, 2, ..., n$. В классической вероятностно-статистической модели предполагается, что

$$y_i = \sum_{1 \le j \le m} a_j x_{ji} + \varepsilon_i, \qquad (3)$$

где $\varepsilon_i, i = 1,2,...,n$, - независимые нормальные случайные величины с нулевым математическим ожиданием и дисперсией σ^2. Модель (3) обычно записывают в матричной форме:

$$Y = aX + \mathrm{E}, \qquad (4)$$

где $Y = (y_1, y_2, ..., y_n)^T$ - вектор значений зависимой переменной, $a = (a_1, a_2, ..., a_m)$ - вектор неизвестных коэффициентов, $X = ||x_{ji}||$ - матрица значений независимых переменных, называемая также матрицей плана (в терминах теории планирования экспериментов), $\mathrm{E} = (\varepsilon_1, \varepsilon_2, ..., \varepsilon_n)^T$ - вектор погрешностей, T - символ транспонирования.

Литература по регрессионному анализу практически необозрима (миллионы названий статей и книг). В частности, многообразие моделей регрессионного анализа обсуждается в [1, 2]. Классическая теория изложена в [3, 4]. Неклассический подход развит в [5]. Вычислительные вопросы рассмотрены в [6]. Оптимальный выбор матрицы плана - предмет теории планирования эксперимента [7]. Ряд ссылок будет дан ниже.

Для модели (3) - (4) теория хорошо развита. Параметры оценивают методом наименьших квадратов, проверяют различные гипотезы. Однако в практических исследованиях часто возникает необходимость выделения "информативного подмножества признаков (независимых переменных)". При этом вместо (3) предполагается справедливой модель

$$y_i = \sum_{j \in J} a_j x_{ji} + \varepsilon_i,\ i = 1,2,...,n, \qquad (5)$$

где J - подмножество множества $\{1, 2, ..., m\}$. Например, если модель (3) используется для управления технологическим процессом или для иного массового применения, то сокращение числа независимых переменных приносит ощутимый экономический эффект от сокращения числа измерений. В научных исследованиях выделение "информативного подмножества признаков" позволяет установить основные факторы, влияющие на изучаемое явление или процесс, и т.д. В дискуссии по прикладной статистике, проведенной во время IV международной вильнюсской конференции по теории вероятностей и математической статистике (Вильнюс, 1985 г.), именно проблема выделения ""информативного подмножества признаков" J была признана наиболее актуальной.

Возникает задача построения состоятельной оценки J_n множества J, т.е. оценки, удовлетворяющей соотношению

$$\lim_{n\to\infty} Card(J_n \Delta J) = 0, \qquad (6)$$

где *Card* (*A*) - число элементов конечного множества *A*.

Разработано большое число методов выделения информативного подмножества признаков (см., например, [8, гл.12], [9, гл.6]). Однако они обычно излагаются как эвристические, свойства их не изучены, неизвестно даже, справедливо ли свойство состоятельности (6). А если оно не выполнено, то, вообще говоря, нельзя гарантировать, что линия регрессии оценивается состоятельно. В рамках статистики нечисловых данных может быть получено (6) на основе общих результатов для решений экстремальных статистических задач [10, 11].

Хорошо известно, распределения реальных данных, как правило, не является нормальным [12, 13]. Однако математический аппарат в случае нормальности зачастую является более простым. Это связано с тем, что глубоко развита теория квадратичных форм в евклидовом пространстве (квадратичные формы стоят в степени экспоненты, описывающей плотность многомерного распределения). Поэтому для первоначального теоретического изучения считаем возможным использовать основанные на нормальности частные случаи регрессионных моделей.

В ряде случаев представляется естественным рассмотреть последовательность моделей вида (1) - (4).Например, изучается зависимость y от t. Естественно попытаться приблизить зависимость сначала константой, при недостаточной точности такого приближения попробовать использовать линейную функцию, при неудаче - квадратическую, затем, если необходимо, - параболу третьего порядка, и т.д. [14]. Приближение y с помощью полинома порядка m - 1 описывается с помощью модели (1) - (4), если положить

$$x_1 = 1, x_2 = t, x_3 = t^2, ..., x_m = t^{m-1}. \qquad (7)$$

В связи с (7) подчеркнем, что x_j в модели (1) - (4) не обязательно являются результатами прямых измерений. Более важным является случай, когда $x_j = f_j(x)$, j = 1, 2, ..., где x - исходные переменные, $f_j(x)$ - некоторые функции. (Модель (1) - (4) - частный случай такой формулировки, когда $x = (x_1, x_2, ..., x_m)$ и $f_j(x) = x_j$, $j = 1, 2, ..., m$. При этом x может иметь произвольную природу, в частности, быть объектом нечисловой природы.

В постановке (7) естественно считать, что модель (1) - (4) имеет место при некотором $m = m_0$, и искать это m_0, увеличивая m на 1, пока модель не будет адекватно описывать данные (подробнее см. ниже).Если априори задано достаточное (наверняка) число переменных M, то информативное подмножество признаков J естественно искать не среди всех подмножеств множества $\{1, 2, ..., M\}$, а среди подмножеств $J(m) = \{1, 2, ..., m\}$, образующих расширяющуюся систему подмножеств

$J(m) \subset J(m+1)$, m = 1, 2, ... Этим постановка (7) отличается от общей постановки (5). Другими словами, в случае (7) структуру модели задает не подмножество J, а натуральное число m_0, которое в соответствии с [15] называем *размерностью модели*.

Рассмотрим два метода, используемых прикладниками [9, 14, 16, 17] для оценки размерности модели m_0. Они основаны на применении "кажущейся ошибки", т.е. величины

$$\Delta_m = \frac{1}{n-m-1} \sum_{1 \le i \le n} (y_i - y_{im})^2 = \frac{1}{n-m-1} \Delta_m^0, \tag{8}$$

где y_{im} - сглаженные по методу наименьших квадратов значения зависимой переменной, полученные при принятии модели (7) с данным m.

Первый метод состоит в том, что в качестве оценки размерности модели, т.е. необходимого числа базисных функций, берут первый локальный минимум "кажущейся ошибки", т.е.

$$m_{1n} = \min\{m : \Delta_{m-1} > \Delta_m,\ \Delta_m \le \Delta_{m+1}\}. \tag{9}$$

Второй метод основан на проверке адекватности модели (3). При этом начинают с m =1 и увеличивают на 1 число параметров только в случае неадекватности, т.е. отклонения гипотезы о том, что данные (2) описываются моделью (3) при используемом m (в постановке (7) при этом увеличивается число используемых базисных функций f_j, т.е. степень полинома, но не число исходных независимых переменных). При известной дисперсии σ^2 для проверки указанной гипотезы можно воспользоваться тем, что Δ_m^0 имеет распределение $\sigma^2 \chi^2_{n-m-1}$. Если σ^2 неизвестно, то применяют известный критерий Фишера: при $m_2 > m_1$ и справедливости (3) с $m = m_1$ статистика

$$f(m_1, m_2) = \frac{(n-m_2-1)(\Delta_{m_1}^0 - \Delta_{m_2}^0)}{(m_2-m_1)\Delta_{m_2}^0} \tag{10}$$

имеет распределение Фишера с числом степеней свободы числителя m_2 - m_1 и знаменателя n - m_2 - 1, и гипотеза H_0: $m = m_1$ отвергается, если

$$f(m_1, m_2) \ge F_\alpha(m_2 - m_1; n - m_2 - 1), \tag{11}$$

где F_α есть $(1-\alpha)$-квантиль распределения Фишера с указанными степенями свободы, α - уровень значимости. Метод оценки размерности модели основан на том, что, рассматривая последовательно m_1 = 1, 2, ..., мы проверяем гипотезу H_0: $m = m_1$ с помощью (10) - (11) (выбор m_2 может быть проведен различными способами, например, $m_2 = m_1 + 1$ или $m_2 = 2m_1$) и останавливаемся на таком наименьшем $\hat{m}$, что рассматриваемая гипотеза не отвергается. В постановке (7) наиболее естественно применять $m_2 = m_1 + 1$. При этом мы используем статистики

$$\xi_k = \frac{(g-k-2)(\Delta_k^0 - \Delta_{k+1}^0)}{\Delta_{k+1}^0},\ k = 1,2,... \tag{12}$$

При $k \geq m_0$ статистика ξ_k имеет распределение Фишера с числом степеней свободы числителя 1 и знаменателя n - k - 2. В качестве оценки размерности модели используют согласно (11)

$$m_{2n} = \min\{k : \xi_k < F_\alpha(1,\ n-k-2).\} \tag{13}$$

Изучим поведение статистик m_{1n} и m_{2n} как оценок истинной размерности модели m_0. Заметим, что если модель (3) адекватна при $m = m_0$, то она адекватна и при любом $m' > m_0$ - достаточно положить $a_{m+1} = a_{m+2} = ... = a_{m'} = 0$. Поэтому истинная размерность m_0 - это минимальное m, при котором модель (3) адекватна.

Воспользуемся геометрической интерпретацией метода наименьших квадратов, рассмотренной А.Н. Колмогоровым [18] и изложенной, например, в [19, §§11,12]. Введем вектора

$$T_j = (x_{j1}, x_{j2}, ..., x_{jn})^T,\ j = 1, 2, ..., m. \tag{14}$$

В постановке (7) они имеют вид

$$T_1 = (1,1,...,1)^T,\ T_j = (t_1^{j-1}, t_2^{j-1}, ..., t_n^{j-1}),\ j = 2,..., m\,. \tag{15}$$

Тогда

$$Y = \sum_{1 \leq j \leq m} a_j T_j + \mathrm{E}\,, \tag{16}$$

где Y и E - те же, что и равенстве (4).

Введем в рассмотрение линейную оболочку

$$L_m = L_m\,(T_1, T_2, ..., T_m) \tag{17}$$

векторов T_1, T_2, ..., T_m. Ясно, что задача оценки параметров методом наименьших квадратов является частным случаем так называемой "общей линейной модели" [19, с.129]. Следовательно, наилучшей оценкой (в модели (3)) для вектора

$$Z = Y - \mathrm{E} = \sum_{1 \leq j \leq m} a_j T_j \tag{18}$$

является проекция Y как элемента евклидова пространства R^n на подпространство L_m. В случае линейной независимости векторов T_1, T_2, ..., T_m проекция однозначно определяет оценки $\hat{a}_j$ коэффициентов a_j, j = 1, 2, ..., m, а именно, оценкой a_j является коэффициент $\hat{a}_j$ в разложении проекции Y_{1m} по базису T_1, T_2, ..., T_m:

$$Y_{1m} = \mathrm{Pr}\,oj_{L_m} = \sum_{1 \leq j \leq m} \hat{a}_j T_j\,. \tag{19}$$

Имеем

$$Y = Y_{1m} + Y_{2m}, \tag{20}$$

где Y_{1m} - проекция Y на L_m, а Y_{2m} - проекция Y на ортогональное дополнение к L_m.. При этом

$$\Delta_m^0 = \| Y_{2m} \|^2\,. \tag{21}$$

Пусть Q_{1n}, Q_{2n}, ..., Q_{mn} - ортонормированный базис в L_m (в предположении dim (L_m) = m). а $Q_{(m+1)n}$, $Q_{(m+2)n}$, ..., Q_{nn} - ортонормированный базис в $L_m^{\perp}$ - ортогональном дополнении к L_m . Тогда

$$Z = \sum_{1 \le j \le m} a_j T_j = \sum_{1 \le j \le n} b_{jn} Q_{jn}\,, \tag{22}$$

и

$$Y = \sum_{1 \le j \le n} \beta_{jn} Q_{jn}\,, \tag{23}$$

где

$$\beta_{jn} = b_{jn} + \sigma\delta_j \tag{24}$$

при $j = 1, 2, ..., m_0$ и

$$\beta_{jn} = \sigma\delta_j \tag{25}$$

при $j = m_0 + 1, ..., n$, где $\delta_1, \delta_2, ..., \delta_n$ - независимые нормальные случайные величины с нулевым математическим ожиданием и единичной дисперсией.

Что можно сказать о случайных величинах $|\beta_{jn}|$? Если исходный базис являлся ортогональным, то в пространстве L_m естественно использовать ортонормированный базис

$$Q_j = Q_{jn} = \frac{T_j}{\| T_j \|}\,. \tag{26}$$

Следовательно, в этом случае

$$\beta_{jn} = a_j \| T_j \| + \sigma\delta_j \tag{27}$$

при $j = 1, 2, ..., m_0$, а при $m > m_0$ величины β_{jn} задаются формулой (25). Поскольку

$$\| T_j \| = \sum_{1 \le i \le n} x_{ji}^2\,, \tag{28}$$

в частности, в постановке (7) $\| T_1 \| = \sqrt{n}$, то для типичных прикладных задач

$$n^{-1} \| T_j \|^2 = O(1),\ \ n \| T_j \|^{-2} = O(1) \tag{29}$$

при $n \to \infty$.

Изучим асимптотическое поведение Δ_m при $n \to \infty$. При этом с изменением n вектора T_j размерности n, разумеется, меняются, и базис Q_{1n}, Q_{2n}, ..., Q_{mn} в L_m тоже, вообще говоря, меняется вместе с коэффициентами b_{jn}, $j = 1, 2, ..., m$, даже в случае, когда ортонормальный базис получаем ортогонализацией $T_1, T_2, ..., T_m$.

Для дальнейших рассуждений есть два пути. Один из них применили М.В. Гальченко и В.А. Гуревич [20]. Они ввели предположение, что матрица плана такова, что при каждом n вектора $T_1, T_2, ..., T_m$ ортогональны. Примером является план [20, с.55] с

$$f_j(x) = \sqrt{2}\cos(j \arccos x),\ \ x \in [-1.1],\ \ x_{in} = \cos\left(\frac{2i-1}{2n}\pi\right). \tag{30}$$

Кроме того, они предполагают, что $a_j \neq 0$ при $j = 1, 2, ..., m_0$.

Специальный вид плана, на наш взгляд, излишнее ограничение. Дальнейшие рассуждения верны для плана "общего вида", нужны лишь некоторые условия регулярности, гарантирующие от "вырождения". Это и есть второй путь.

Имеем

$$\Delta_m = \frac{1}{n-m-1}\left(\beta^2_{(m+1)n} + ... + \beta^2_{nn}\right). \tag{31}$$

В силу (25) и справедливости (по теореме Чебышева) закона больших чисел для δ_j^2 имеем

$$\Delta_m \to \sigma^2 \tag{32}$$

по вероятности при $n \to \infty$, если $m \geq m_0$.

Пусть теперь $m < m_0$. Представим Δ_m в виде суммы двух слагаемых

$$\Delta_m = \frac{1}{n-m-1}\left(\beta^2_{(m+1)n} + ... + \beta^2_{m_0 n}\right) + \frac{1}{n-m-1}\left(\beta^2_{(m_0+1)n} + ... + \beta^2_{nn}\right). \tag{33}$$

Из (32) следует, что второе из них сходится по вероятности к σ^2 при $n \to \infty$. Если

$$\varliminf_{n\to\infty} \frac{b^2_{jn}}{n} = \gamma_j > 0, \quad j = 1,2,...,m_0, \tag{34}$$

то

$$\Delta_{m-1} - \Delta_m = \frac{1}{n}\beta^2_{mn}(1 + o(1)) \tag{35}$$

и для любого $\varepsilon > 0$

$$\lim_{n\to\infty} P(\Delta_{m-1} - \Delta_m \geq \gamma_m - \varepsilon) = 1. \tag{36}$$

Из (36) следует, что

$$\lim_{n\to\infty} P\{m_{1n} < m_0\} = 0. \tag{37}$$

Пусть теперь $m \geq m_0$. Имеем

$$\Delta_m - \Delta_{m+1} = \frac{1}{n-m-1}\left(\beta^2_{(m+1)n} - \frac{1}{n-m-2}\left(\beta^2_{(m+2)n} + ... + \beta^2_{nn}\right)\right). \tag{38}$$

В силу (32)

$$\lim_{n\to\infty} P\{\Delta_m - \Delta_{m+1} \leq 0\} = P\{\beta^2_{(m+1)n} \leq \sigma^2\} = \lambda, \tag{39}$$

где

$$\lambda = P\{\delta^2_{m+1} \leq 1\} = \frac{1}{\sqrt{2\pi}}\int_{-1}^{1} \exp\left\{-\frac{x^2}{2}\right\} dx = 0{,}68268... \tag{40}$$

Из (37) и (39) вытекает, что

$$\lim_{n\to\infty} P\{m_{1n} = m_0\} = \lim_{n\to\infty} P\{\Delta_{m_0} \leq \Delta_{m_0+1}\} = \lambda. \tag{41}$$

В силу (38) и (32) величина

$$P\{m_{1n} = m_0 + k \mid m_{1n} \geq m_0\} = \lim_{n\to\infty} P\{\Delta_{m_0} > \Delta_{m_0+1}, \Delta_{m_0+1} > \Delta_{m_0+2}, \\ ..., \Delta_{m_0+k-1} > \Delta_{m_0+k}, \Delta_{m_0+k} \leq \Delta_{m_0+k+1}\} \tag{42}$$

сходится при $n \to \infty$ к

$$P\{\beta^2_{(m_0+1)n} > \sigma^2, ..., \beta^2_{(m_0+k)n} > \sigma^2, \beta^2_{(m_0+k+1)n} \leq \sigma^2\}. \tag{43}$$

Из независимости $\beta_{(m_0+1)n}, ..., \beta_{(m_0+k+1)n}$ соотношений (25) и (39) вытекает, что

$$\lim_{n\to\infty} P\{m_{1n} = m_0 + k\} = \lambda(1-\lambda)^k, \quad k = 0,1,2,..., \tag{44}$$

где λ определено в (40). Итак, доказана следующая теорема, впервые полученная в [21].

Теорема 1. Пусть модель (1) - (4) верна при $m = m_0$. Пусть справедливы условия регулярности (34). Тогда имеют место предельные соотношения (37) и (44), т.е. распределение оценки m_{1n} в пределе является геометрическим.

Следствие. Оценка m_{1n} не является состоятельной (в смысле, принятом в математической статистике).

Замечание. Просматривается аналогия с последовательным анализом. В частности, соотношения типа (43) - (44) справедливы для декартовых последовательных критериев [22, с.485]. Специфика рассматриваемой задачи состоит в том, чтобы избавиться от зависимости последовательных проверок, что удается сделать в асимптотике с помощью соотношений типа (32). Представляется перспективным использование оптимальных правил остановки, разработанных в статистическом последовательном анализе [23]. Однако необходимо отметить, что типичные задачи последовательного анализа, в частности, задачи разладки и задачи последовательного различения простых гипотез с помощью критерия отношения вероятностей, существенно отличаются от рассматриваемых нами задач регрессионного анализа.

Условие (34) - это условие типа того, что мы находимся в ситуации "общего положения" (ср. [24]), т.е. отсутствует "вырождение". Если при всех n базис T_1, T_2, ..., T_m является ортогональным, как для плана (30), то согласно (23) и (26) $b_{jn} = a_j \|T_j\|$, а потому соотношение (34) эквивалентно тому, что $a_j \neq 0$ при $j = 1, 2, ..., m_0$ и

$$\underline{\lim}_{n\to\infty}\left(\frac{1}{n}\|T_j\|^2\right) = \gamma_j', \ \gamma_j' = \frac{\gamma_j}{a_j}, j = 1,2,...,m_0, \tag{45}$$

Соотношение (45) справедливо, например, для плана (30). Грубо говоря, условия (34) и (45) означают, что "вклады" вновь добавляемых переменных "не вырождаются", т.е. по порядку такие же, как вклад $T_1 = (1, 1, ..., 1)$ в постановке (7).

Рассмотрим теперь оценку m_{2n}. Согласно (10) и (21) имеем

$$f(m_1, m_2) = \frac{\frac{1}{m_2 - m_1}\left(\beta^2_{(m_1+1)n} + ... + \beta^2_{m_2 n}\right)}{\frac{1}{n - m_2 - 1}\left(\beta^2_{(m_2+1)n} + ... + \beta^2_{nn}\right)}. \tag{46}$$

Пусть выполнено условие (34). Тогда в силу (24) для любого $\varepsilon > 0$

$$\lim_{n\to\infty} P(\beta^2_{mn} > n\gamma_m(1-\varepsilon)) = 1, \ m = 1,2,...,m_0. \tag{47}$$

Если $m_1 < m_0$, то для числителя в (46) имеем:

$$\lim_{n\to\infty} P\left(\frac{1}{m_2 - m_1}\left(\beta^2_{(m_1+1)n} + ... + \beta^2_{m_2 n}\right) > \frac{n(1-\varepsilon)}{m_2 - m_1}\left(\gamma_{m_1+1} + ... + \gamma_{m_3}\right)\right) = 1 \tag{48}$$

для любого $\varepsilon > 0$, где $m_3 = \min(m_2, m_0)$.

Если $m_1 < m_0, m_2 \geq m_0$, то из (32) и (48) следует, что существует $C > 0$ такое, что

$$\lim_{n\to\infty} P\{f(m_1, m_2) > Cn\} = 1. \tag{49}$$

Пусть оценка m_{2n} размерности модели m_0 определяется с помощью последовательности троек

$$(m_1(k), m_2(k), F(k)),\ k = 1, 2, ...,\ m_1(k) < m_2(k), \tag{50}$$

где последовательности натуральных чисел $m_1(k)$, $m_2(k)$ возрастают. Гипотеза H_0: $m = m_1(k)$ против альтернативы H_1: $m = m_2(k)$ проверяется с помощью статистики $f(m_1(k), m_2(k))$, критическое значение выбирается согласно (11) с уровнем значимости $\alpha = F^{-1}_{m_2-m_1, n-m_2-1}(F(k))$. Это описание получения оценки m_{2n} - несколько более общее, чем данное ранее (формулы (10) - (13)), когда предполагалось, что $m_1(k) \equiv k$ и $F(k) \equiv F_\alpha$. Если гипотеза H_0 отвергается при $k = 1, 2, ..., k_0$ и впервые принимается при $k = k_0 + 1$, то полагаем $m_{2n} = m_1(k_0 + 1)$.

Теорема 2 [15]. Пусть выполнены условия (34), (52). Тогда

$$\lim_{n\to\infty} P(m_{2n} < m_0) = 0. \tag{51}$$

Доказательство вытекает из соотношения (49), согласно которому при достаточно больших n гипотеза H_0 может быть принята только при $m_1(k) \geq m_0$. если известно, что $m_2(k) > m_0$. Остается рассмотреть случай, когда $m_2(k) \leq m_0$. Для того, чтобы гипотеза H_0 отвергалась при любом $F(k)$ с вероятностью, стремящейся к 1 при $n \to \infty$, необходимо и достаточно, чтобы для любого $m < m_0$ было выполнено соотношение

$$\lim_{n\to\infty}\left(nb^2_{mn}\left(\sum_{m+1\leq j\leq m} b^2_{mn} \right)^{-1} \right) = \infty. \tag{52}$$

Замечание. Как видно из проведенных рассуждений, для справедливости (51) нет необходимости требовать выполнения (34), достаточно справедливости (52) и условия

$$\lim_{n\to\infty} b^2_{mn} = \infty,\ \ j = 1,2,...,m_0 \tag{53}$$

Теорема 3 [15]. Пусть оценка размерности модели m_{2n} определяется с помощью последовательности проверок (50). Пусть выполнено (51). Тогда для любого целого $q \geq 0$ существует

$$p(q) = \lim_{n\to\infty} P\{m_{2n} = m_0 + q\}. \tag{54}$$

Доказательство. С помощью (25) и (32) получаем из (46) и (11), что

$$\begin{aligned} &\lim_{n\to\infty} P\{m_{2n} = m_0 + q\} = P\{\delta^2_{m_1(k)+1} + ... + \delta^2_{m_2(k)} \geq F(k)(m_2(k) - m_1(k)), \\ &k_1 \leq k < k_2,\ \delta^2_{m_1(k)+1} + ... + \delta^2_{m_2(k)} < F(k)(m_2(k) - m_1(k)),\ k = k_2\}, \end{aligned} \tag{55}$$

где $\{\delta_1, \delta_2, ..., \delta_m, ...\}$ - последовательность независимых нормальных случайных величин с нулевым математическим ожиданием и единичной

дисперсией, $k_1 = \min\{k: m_1(k) \geq m_0\}$, число k_2 таково, что $m(k_2) = m_0 + q$. Если же $m_0 + q$ не принадлежит множеству $\{m_1(k), k = 1, 2, ...\}$, то очевидно, что $p(q) = 0$.

Теорема 4 [15, 25]. Пусть $m_1(k) = k$, $m_2(k) = k + 1$, $F(k) = F$, $k = 1, 2, ...$ Пусть выполнены условия (52), (53). Тогда

$$p(k) = \lambda(1-\lambda)^q, \ q = 0,1,2,... \tag{56}$$

где

$$\lambda = P\{\delta_1^2 < F) = \Phi\left(\sqrt{F}\right) - \Phi\left(-\sqrt{F}\right). \tag{57}$$

Доказательство. При данном в теореме 4 виде последовательности (50) статистика $f(m_1, m_2)$ переходит в ξ_k из (12). Согласно теореме 2 справедливо (51). Согласно теореме 3

$$p(q) = P\{\delta_k^2 \geq F, \ m_0 \leq k < m_0 + q, \ \delta_{m_0+q}^2 < F\} = [P\{\delta_1^2 \geq F\}]^q P\{\delta_1^2 < F\}, \tag{58}$$

откуда и следует требуемое. Сравним предельное распределение оценки m_{1n} (формулы (44), (40)) и предельное распределение оценки m_{2n} (формулы (56), (57)). Видим, что при $F = 1$, т.е. при $\lambda = 0{,}68268...$, предельные распределения этих оценок совпадают. Поэтому можно сказать, что оценка m_{2n} обобщает оценку m_{1n}.

Обсудим значение основных предпосылок, при которых получены теоремы 1 - 4, а именно, нормальности погрешностей ε_i в (3), "условия невырожденности" (34) и аналогичных ему условий (52) - (53).

Нормальность распределений случайных величин ε_i используется для получения следующих двух утверждений: после ортогонализации базиса, т.е. перехода от $\{T_j\}$ к $\{Q_{jn}\}$ (см. (22) - (23)) ошибки по-прежнему независимы и одинаково распределены; параметр λ в (44) и (56) выражается через нормальное распределение по формулам (40) и (57) соответственно.

Сохранение независимости ошибок при переходе к другому базису - характеризационное свойство нормального распределения. Это - следствие известного цикла характеризационных теорем [26], начатого работой С.Н. Бернштейна 1941 г. [27] и продолженного в исследованиях Б.В. Гнеденко [28], В.П. Скитовичем, Г. Дармуа, Ю.В. Линником, А.А. Зингером и др.

Отказаться от нормальности можно в предположении, указанном в [25] и принятом за основу в [20], что план эксперимента имеет специальный вид, обеспечивающий ортогональность базиса $\{T_j\}$ (тогда переход к $\{Q_{jn}\}$ не нужен). Примером является план (30). Пусть в этом случае ошибки ε_i - независимые одинаково распределенные случайные величины с конечным начальным вторым моментом $M(\varepsilon_i^2) = \mu_2 < \infty$. Пусть выполнено (34). Тогда, как нетрудно убедиться, проследив проведенные выше выкладки, выполнены соотношения (37) и (44) с $\lambda = P\{\varepsilon_1^2 \leq \mu_2\}$. Если выполнены условия (52) и (53), то справедливы соотношения (51) и (57) с $\lambda = P\{\varepsilon_1^2 \leq F\}$. Очевидно, можно отказаться и от предположения одинаковой

распределенности помех ε_i, как это сделано в [20], но это делать здесь не будем, поскольку принципиально новых результатов при этом не получено, а демонстрировать владение техникой предельных теорем нет необходимости.

Невыполнение одного из условий (34), (53) в силу (29) практически эквивалентно (в предположении, что $\{T_j\}$ - ортогональный базис при всех n) тому, что модель (3) верна при $m = m_0$, но при некотором $j < m_0$ имеем $a_j = 0$. Каковы свойства оценок m_{1n} и m_{2n} в этом случае?

Для упрощения описания поведения оценок предположим, что существуют

$$\lim_{n\to\infty}\frac{b_{tn}^2}{n} = \rho_t > 0,\ \ t = 1,2,...,m_0. \tag{59}$$

Тогда согласно [25]

$$\lim_{n\to\infty} P\{m_{1n} = j\} = P\left\{\delta_j^2 \le \frac{1}{\sigma^2}\left(\rho_{j+1} + ... + \rho_{m_0}\right) + 1\right\} \tag{60}$$

и

$$\lim_{n\to\infty} P\{m_{2n} = j\} = P\left\{\frac{\sigma^2\delta_j^2}{\rho_{j+1} + ... + \rho_{m_0} + \sigma^2} \le \left[\Phi^{-1}\left(1 - \frac{\alpha}{2}\right)\right]^2\right\},\ \ m_1(k) = k,\ m_2(k) = k+1, \tag{61}$$

т.е. с достаточно высокой вероятностью произойдет преждевременный останов. От предположений (59) можно избавиться, заменив предельные переходы на сближение левых и правых частей (60) и (61) и ρ_t на $\frac{b_{tn}^2}{n}$.

6.3. Состоятельные оценки размерности и структуры модели в регрессии

Рассмотренные в предыдущем разделе методы оценки истинной размерности модели (3) не являются состоятельными:

$$\lim_{n\to\infty} P(m_{in} = m_0) \ne 1,\ \ i = 1,2. \tag{62}$$

В настоящем разделе рассмотрим построение состоятельных оценок m_n^* параметра m_0, т.е. оценок, для которых

$$\lim_{n\to\infty} P(m_n^* = m_0) = 1. \tag{63}$$

В [20] предложена состоятельная модификация m'_{1n} оценки m_{1n}. В отличие от (9) в качестве оценки взят не первый локальный минимум "кажущейся ошибки" Δ_m, а первый локальный минимум линейной функции от неё $A_{nm}\Delta_m + B_{nm}$, где A_{nm} и B_{nm} - некоторые константы. Предположение [20] о специальном виде плана излишне, от него можно избавиться методами предыдущего раздела. Другие подходы рассмотрены в [8, 9, 29, 30, 31].

Состоятельную модификацию оценки m_{2n} можно получить, заменив правую часть в (11) на величину, растущую с увеличением n так, что

правая часть в (55) стремится к 0 при $n \to \infty$, но при этом выполнено (51). В частности, рассмотрим оценку

$$m_{3n} = \min\{k : \xi_k < \omega(n)\}, \tag{64}$$

где ξ_k определено в (12), $\omega(n)$ - некоторая последовательность. Для справедливости (55) необходимо и достаточно, чтобы

$$\lim_{n\to\infty} \omega(n) = +\infty. \tag{65}$$

Для справедливости (51) согласно доказательству теоремы 2 достаточно выполнения соотношений (34), (52) и

$$\lim_{n\to\infty} \frac{\omega(n)}{n} = 0. \tag{66}$$

Из проведенных рассуждений вытекает следующая теорема [25].

Теорема 5. Пусть выполнены соотношения (34), (52), (65) и (66). Тогда оценка $m_n^* = m_{3n}$, заданная формулой (64), является состоятельной оценкой размерности модели, т.е. удовлетворяет соотношению (63).

Рассмотрим некоторые другие методы оценки размерности модели, а также выбора информативного подмножества признаков. При этом весьма полезной оказывается независимость в совокупности получаемых по (23) - (25) оценок β_{jn} параметров регрессии в ортонормальном базисе $\{Q_{jn}\}$.

Упорядочим оценки β_{jn} в порядке убывания их абсолютной величины:

$$|\beta_{j(1)n}| \geq |\beta_{j(2)n}| \geq \ldots \geq |\beta_{j(n)n}|. \tag{67}$$

Предположим сначала, что σ известно. Выберем ν_n из условия

$$2(1-\Phi(\nu_n)) = \frac{1}{n}. \tag{68}$$

Тогда, как известно [32, с.410],

$$\nu_n \cong 2\sqrt{\ln n} \tag{69}$$

и, кроме того,

$$P\left\{\max_{0\leq i\leq n-1} |\delta_i| > \nu_n\right\} \leq \frac{1}{n}. \tag{70}$$

Оценку m^* размерности модели m_0 найдем из условия

$$|\beta_{j(m^*)n}| \geq \sigma\nu_n, \ |\beta_{j(m^*+1)n}| < \sigma\nu_n. \tag{71}$$

Если (см. (28))

$$\lim_{n\to\infty} (\ln n)^{-1/2} \| T_j \| = +\infty, \ \ j = 1,2,..., \tag{72}$$

то условие (71) дает состоятельную оценку размерности модели m_0 = Card J, а множество

$$J_n = \{j(1), j(2), \ldots, j(m^*)\} \tag{73}$$

является состоятельной оценкой информативного подмножества признаков J (см.(5)) в смысле (6).

Пусть теперь σ неизвестно. Укажем семейство оценок σ. Пусть $0 < \theta < 1$. Рассмотрим $\beta_{j([\theta n])n}, \ldots, \beta_{j(n)n}$. При $n \to \infty$ выборочная дисперсия $s^2(\theta)$

этих случайных величин сходится к дисперсии $\sigma\xi$, где ξ - урезанная на отрезок $\left[-\Phi\left(1-\theta/2\right),\ \Phi\left(1-\theta/2\right)\right]$ стандартная нормальная случайная величина, т.е. $s^2(\theta)$ сходится к $\sigma^2(1-\theta)$. Следовательно, оценкой параметра σ^2 является $(1-\theta)^2 s^2(\theta)$. Эту оценку можно использовать в (71). Состоятельность описанных выше оценок при этом сохраняется.

Оценки (71) и (73) рассмотрены согласно [25]. В ситуации, когда исходный базис не является ортонормальным, требуются некоторые пояснения типа тех, что были даны выше в связи с работой М.В. Гальченко и В.А. Гуревича [20] (см. (30)). От (5) следует перейти к аналогичной записи в ортонормальном базисе $\{Q_{jn}\}$, вообще говоря, зависящем от n. Примем, что базис $\{Q_{jn}\}$ получен ортогонализацией и нормированием исходного базиса. Тогда вместо (5) имеем

$$Y = \sum_{j\in J(n)} b_{jn} Q_{jn} + \mathrm{E}, \tag{74}$$

где

$$\max\{j : j \in J(n)\} = ma[\{j : j \in J\} = j_0. \tag{75}$$

Если

$$\lim_{n\to\infty} \min_{1\le j\le j_0} (\ln n)^{-1/2} \mid b_{jn} \mid = +\infty \tag{76}$$

то справедлив аналог состоятельности оценок

$$\lim_{n\to\infty} Card(J_n \Delta J(n)) = 0. \tag{77}$$

Другой поход к нахождению информативного подмножества признаков - метод "всех регрессий" [8] - основан на статистике

$$\hat{J}_{nk} = Arg \min_{a_j, J\in A_k} \sum_{i=1}^{n} \left(\sum_{j\in J} a_j x_{ji} - y_i \right)^2 = Arg \min_{J\in A_k} g, \tag{78}$$

где

$$g = g(J, a_j, x_{ji}, y_i,\ 1\le j\le m,\ 1\le i\le n) = \sum_{i=1}^{n} \left(\sum_{j\in J} a_j x_{ji} - y_i \right)^2, \tag{79}$$

а *Arg* min берется по всем J таким, что $J \in A_k$, т.е.

$$Card(J) \le k. \tag{80}$$

Рассмотрим функцию

$$h_k(\omega) = \min_{J\in A_k} \min_{a_j, j\in J} g. \tag{81}$$

Из результатов об асимптотике решений экстремальных статистических задач [10] следует, что по вероятности

$$\lim_{n\to\infty} h_k(\omega) = h_k, \tag{82}$$

где (в общей ситуации) функция h_k сначала убывает при росте k от $k = 1$ до $k = \mathrm{Card}(J_{ист})$, затем остается постоянной (равной h), а

$$\lim_{n\to\infty} P\left(\hat{J}_{nk} \in \left\{ J : \min_{a_j} M\left(\sum_{j\in J} a_j x_{ji} - y_i \right)^2 = \min_J = h \right\} \right) = 1. \tag{83}$$

Отсюда следует, что метод "всех регрессий", вообще говоря, не дает состоятельных оценок истинного множества информативных признаков $J_{ист}$, а даёт оценки "с завышением", что выражается формулой (83). Это означает, что разнообразные программно-алгоритмические методы нахождения "наилучшей" регрессии [8, гл.12; 9, гл.6], в которых не обращается внимание на отличие (83) от желаемой состоятельности (6), нуждаются в более тщательном изучении.

6.4. Оценивание числа элементов смеси в задачах классификации

Среди задач классификации [33, 34] важное место занимают задачи расщепления смесей. В них принимают, что наблюдается выборка из распределения с плотностью

$$f(x) = \sum_{1 \le i \le m} \pi_i f_i(x), \qquad (84)$$

где плотности $f_i(x)$ описывают отдельные классы, а π_i - веса этих классов, $\pi_i > 0$, $\pi_1 + \pi_2 + ... + \pi_m = 1$. Часто считают, что $f_i(x) = f(x, \theta_i)$, т.е. плотности элементов смеси взяты из некоторого параметрического семейства, $\theta_i \in \Theta$. Запись (84) можно рассматривать также как приближение плотности $f(x)$ с помощью линейной комбинации плотностей $f_1(x)$, $f_1(x)$, ... в этом случае веса π_i не обязаны быть положительными, а вместо равенства (84) имеет быть предельный переход.

Смеси встречаются в различных прикладных задачах. Так, Э.С. Эренбург моделировал продолжительность безотказной работы изделий бытовой техники как смесь двух классов - изделий со скрытыми дефектами и изделий без скрытых дефектов [35].

Если число слагаемых в сумме (84) известно и все $\pi_i > 0$, то с теоретической точки зрения оценивание параметров π_i и θ_i не представляет трудностей - можно применять оценки максимального правдоподобия или одношаговые оценки [36]. Рассмотрим оценивание числа слагаемых. Вначале приведем один известный результат.

Пусть $\xi_1, \xi_2, ..., \xi_n$, - выборка из совокупности с плотностью $f(x, \theta)$, где параметр $\theta \in \Omega$ имеет размерность r. Пусть подпространство $\Omega_0 \subset \Omega$ имеет размерность $r' < r$. Для проверки гипотезы $H_0 : \theta \in \Omega_0$ при альтернативе $H_1 : \theta \in \Omega \setminus \Omega_0$ применяют критерий отношения правдоподобия

$$\lambda(\Omega_0, \Omega) = \sup_{\theta \in \Omega_0} \prod_{1 \le i \le n} f(\xi_i, \theta) \Big(\sup_{\theta \in \Omega} \prod_{1 \le i \le n} f(\xi_i, \theta) \Big). \qquad (85)$$

В [22, §13.8] при некоторых условиях регулярности показано, что при $\theta \in \Omega_0$ распределение случайной величины $(-2 \log \lambda(\Omega_0, \Omega)$ сходится при $n \to \infty$ к распределению хи-квадрат с r - r' степенями свободы. Это

доказывается путем построения r - r' независимых стандартных нормальных случайных величин $\eta_1, \eta_2, ..., \eta_{r-r'}$, таких, что

$$\left(-2\log\lambda(\Omega_0, \Omega) = \sum_{1 \le j \le r-r'} \eta_j^2 + o(1)\right) \quad (86)$$

по вероятности при $n \to \infty$.

Рассмотрим последовательность описанных выше задач. Пусть $\Omega_0 \subset \Omega_1 \subset \Omega_2 \subset ...$ - последовательность пространств параметров,

$$\dim \Omega_i = r' + iq, \ i = 0,1,2,... \quad (87)$$

при некоторых r' и q. Пусть проводится проверка гипотез $H_i : \theta \in \Omega_i$ при альтернативах H_{i+1} последовательно при $i = 0, 1, 2, ...$ Проверки проводятся с помощью статистики $\lambda(\Omega_i, \Omega_{i+1})$ (см. (85)), гипотеза H_i отвергается, если $\left(-2\log\lambda(\Omega_i, \Omega_{i+1}\right) > \lambda_\gamma$, где λ_γ есть $100(1-\gamma)$-процентная точка распределения χ^2 с q степенями свободы. Пусть впервые при $i = m^*$ гипотеза H_i не отвергнута. Какова предельное распределение m^* при $n \to \infty$?

Пусть $\theta \in \Omega_{m(0)}$ и $\theta \notin \Omega_{m(0)-1}$ Так же, как в разделе 2 настоящей главы, можно показать, что при некоторых условиях регулярности [22]

$$\lim_{n\to\infty} P(m^* < m(0)) = 0, \ \lim_{n\to\infty} P(m^* = m(0) + a) = \gamma^a(1-\gamma), \ a = 0,1,2,... \quad (88)$$

При доказательстве используется независимость главных членов в разложениях типа (85) для $\left(-2\log\lambda(\Omega_i, \Omega_{i+1})\right)$. Как и в разделе 3 настоящей главы, состоятельную оценку $m(0)$ получаем, сделав γ зависящим от n.

С формальной точки зрения частным случаем рассматриваемой последовательности проверок является определение числа элементов смеси (параметра m в модели (84)). При этом $\Omega_s = \{(\pi_1, ..., \pi_s, \theta_1, ..., \theta_s)\}$. Тогда в (87) $r' = \dim\theta, \ q = \dim\theta + 1$.

Однако в силу специфики модели (84) соотношения (88) верны не всегда, в частности, они неверны, если рассматривается смесь нормальных распределений [37]. Поскольку необходимо $\Omega_i \subset \Omega_{i+1}$, а точка $\theta \in \Omega_0$ в (85) должна быть внутренней, то ограничения $\pi_i > 0$ или $\pi_i \ge 0$ противоречат условиям регулярности Уилкса. Поэтому не будем принимать эти ограничения. Далее, информационная матрица вырождается, если $f_i(x, \theta_i)$ и $f_{i+1}(x, \theta_{i+1})$ могут совпадать, как это имеет место для смеси нормальных распределений. Действительно, если $f_i(x, \theta_i) = f_{i+1}(x, \theta_{i+1})$, то

$$\pi_i f_i(x, \theta_i) = \pi'_i \ f_i(x, \theta_i) + (\pi_i - \pi'_i) f_{i+1}(x, \theta_{i+1}), \quad (89)$$

т.е. разложение в (84) неоднозначно. Поэтому предложение использовать критерий Уилкса для нормальных смесей нельзя признать обоснованным.

Предельное распределение (88), полученное для смеси (84) в [38], имеет место при справедливости условий регулярности Уилкса, например, когда задана последовательность линейно независимых плотностей $f_1(x)$,

$f_2(x)$, ... и $\pi_i \in R^1$. Интересные результаты получены А.М. Никифоровым [39].

6.5. Оценка размерности модели в факторном анализе и многомерном шкалировании

Идея многомерного шкалирования состоит в представлении каждого объекта точкой геометрического пространства небольшой размерности (обычно размерности 1, 2 или 3), координатами которой служат "скрытые значения факторов", в совокупности достаточно адекватно описывающих объект. Размерности 1 - 3 позволяют провести визуальный анализ (о нем на примере клинической медицины см. [40]). В прикладном многомерном статистическом анализе имеется большое число методов снижения размерности - факторный анализ, метод главных компонент, многомерное шкалирование [41, 42]), целенаправленное проецирование [43, 44]) (этой группе методов посвятил свой доклад П. Хубер на Первом Всемирном Конгрессе Общества математической статистики и теории вероятностей им. Бернулли [45]). Цель всех этих методов - от большого числа признаков перейти к существенно меньшему, вообще говоря, вновь сконструированных признаков, которые тем не менее достаточно адекватно описывают рассматриваемые объекты. Многомерное шкалирование использует не сами объекты (как вектора в многомерном пространстве), а расстояния между ними ρ_{ij}, вычисленные по координатам векторов или заданные иными способами, например, с использованием экспертов. Требуется подобрать точки-представители в евклидовом пространстве небольшой размерности так, чтобы расстояния между ними r_{ij} мало отличались от расстояний между объектами ρ_{ij}. Согласно одной из формализаций (в т.н. метрическом шкалировании) должна достигать минимума величина

$$S = \sum_{i<j} | \rho_{ij} - r_{ij} | . \qquad (90)$$

В настоящем разделе мы не будем пытаться подробно рассматривать многообразие методов рассматриваемого типа (см. указанную выше литературу и наши публикации [46, 47]), а разберем модельную постановку оценки размерности итогового пространства.

Пусть объекты описываются точками d_1, d_2, ..., d_n, в k-мерном евклидовом пространстве. Пусть L_m - пространство размерности m. Пусть $\rho(d_i, L_m)$ - расстояние между точкой d_i и линейным пространством L_m, и

$$f_n(m) = \min_{\{L_m\}} \sum_{1 \le i \le n} \rho(d_i, L_m) \ - \qquad (91)$$

- сумма расстояний точек d_1, d_2, ..., d_n до их наилучшего приближения гиперплоскостью размерности m. Пусть в рассматриваемой вероятностной модели

$$d_i = d_i^0 + \varepsilon_i, \qquad (92)$$

где ε_i - независимые нормальные случайные вектора с математическим ожиданием 0 и ковариационной матрицей $\sigma^2 I$, где I - единичная матрица, точки d_i^0 лежат в гиперплоскости размерности m_0 и не лежат (одновременно все вместе) ни в какой гиперплоскости меньшей размерности. Тогда методами раздела 2 настоящей главы установлено [46, с.68-70], что при $n \to \infty$ и соответствующих условиях регулярности (типа данных выше в разделе 2)

$$f_n(m) \to f(m) = f_1(m) + \sigma^2(k-m), \ m = 1,2,...,k, \qquad (93)$$

по вероятности, где $f_1(m)$ - функция, зависящая от расположения точек $d_1^0, d_2^0, ..., d_n^0$. Примем для первичного анализа ситуации, что эти точки имеют круговое нормальное распределение в том подпространстве размерности m_0, в котором они лежат, т.е.

$$d_i^0 = \xi_i(1)e(1) + \xi_i(2)e(2) + ... + \xi_i(m_0)e(m_0), \qquad (94)$$

где $e(1)$, $e(2)$, ..., $e(m_0)$ - ортонормальный базис в этом пространстве, а $\xi_i(1), \xi_i(2), ..., i = 1,2,...,n,$ - независимые нормальные случайные величины с математическими ожиданиями 0 и одинаковыми дисперсиями σ_0^2. Тогда в силу (93) имеем

$$f_1(m) = \begin{cases} \sigma_0^2(m_0 - m), & m < m_0, \\ 0, & m \ge m_0. \end{cases} \qquad (95)$$

Таким образом, функция $f(m)$ из (93) линейна на отрезках [1, m_0] и [m_0, k], причем на первом отрезке она убывает быстрее, чем на втором. Отсюда следует, что статистика

$$m^* = Arg \max_m (f_n(m+1) - 2f_n(m) + f_n(m-1) \qquad (96)$$

является состоятельной оценкой истинной размерности m_0 модели многомерного шкалирования.

Примечание. Если справедлива модель (94), упомянутые выше условия регулярности (типа рассматриваемых в разделе 2 настоящей главы) выполнены.

Итак, из вероятностно-статистической теории вытекает рекомендация - определять размерность факторного пространства по правилу (96). Отметим: подобная рекомендация была сформулирована как эвристическая одним из основателей многомерного шкалирования Краскалом на основе опыта практического использования этого метода и вычислительных экспериментов (см., например, [42]). Вероятностная теория позволила обосновать эту эвристическую рекомендацию. Точнее, выше показано, что в достаточно естественной модели она приводит к состоятельной оценке.

К тематике настоящего раздела относятся также работы [48, 49], подробный анализ содержания которых опустим.

6.6. Регрессия после классификации

Известно, что регрессионный анализ дает доступные интерпретации результаты лишь применительно к достаточно однородным совокупностям (см. обсуждение понятия "однородность" в [50]. Поэтому исходные данные рекомендуют разбить на однородные группы и лишь затем применять регрессионный анализ к каждой из них по отдельности.

Программный продукт по прикладной статистике обычно включает в себя ряд методов классификации и регрессии. Поскольку статистическое исследование включает в себя, как правило, последовательное применение не одного, а многих алгоритмов, работа предыдущего алгоритма может, вообще говоря, нарушать условия применимости последующих. Поэтому раздел 6 Рекомендаций ВНИИС [51] посвящен вопросам "стыковки" последовательно выполняемых алгоритмов: "При последовательном применении нескольких методов обработки данных необходимо обеспечить проверку условий применения каждого последующего метода" [51, с.9].

Рассмотрим "стыковку" алгоритмов классификации и регрессии [52]. Пусть в результате работы некоторого алгоритма классификации выделена группа "однородных" наблюдений. Можно ли применять тот или иной метод регрессионного анализа [1] к элементам этой группы? Во-первых, эти элементы, вообще говоря, не являются независимыми, т.к. границы группы определяются по исходной выборке, а не задаются априорно. Во-вторых, наблюдения не могут иметь нормальное распределение, поскольку элементы группы ограничены по крайней мере с некоторых сторон (например, несколькими гиперплоскостями). Следовательно, обычные предпосылки регрессионного анализа не выполнены, а потому влияние отклонений от этих предпосылок на свойства алгоритмов требуют специального изучения (прежде всего, в рамках общей схемы устойчивости [53]).

В качестве примера рассмотрим "стыковку" алгоритмов классификации и регрессии, когда классификация сводится к расщеплению смеси (см. раздел 4 выше). Пусть для простоты $m = 2$ в смеси (84). Находят состоятельные оценки параметров смеси и строят с их помощью дискриминантную поверхность

$$g(x, \alpha_n) = \beta_n \qquad (97)$$

где x - элемент того пространства, в котором лежат наблюдения, функция g задает вид дискриминантной поверхности (в простейшем случае g - линейная функция), α_n и β_n - оценки параметров дискриминантной (разделяющей) поверхности. Если $g(\xi_j, \alpha_n) > \beta_n$, то наблюдение ξ_j относят к первому классу (совокупности)ЯЯ, в противном случае - ко второму. Зависимость наблюдений, попавших в один класс, имеет своей причиной

то, что параметры α_n и β_n определяются по всей исходной выборке, в том числе и по тем наблюдениям, что попали в рассматриваемый класс. Однако обычно существуют предельные значения α и β такие, что $\alpha_n \to \alpha$ и $\beta_n \to \infty$ по вероятности при $n \to \infty$. Тогда, как легко видеть, совместное распределение фиксированного конечного числа элементов одного класса стремится к совместному распределению независимых случайных элементов, распределение которых получено из рассмотрения соответствующего слагаемого в исходной смеси (84) усечением на область $\{x : g(x,\alpha) > \beta\}$ (для первого класса) или на область $\{x : g(x,\alpha) \le \beta\}$ (для второго класса).

Хотя в каждом из двух классов (кластеров) наблюдения и являются асимптотически независимыми, их распределения отличаются от $f_1(x)$ и $f_2(x)$ соответственно, т.е. от распределений, описывающих исходные классы. В частности, математические ожидания и ковариационные матрицы отличаются от исходных, поэтому с помощью выборочны характеристик, рассчитанных по кластерам, нельзя непосредственно оценивать характеристики исходных классов. Аналогичные выводы справедливы и для иных способов кластеризации [38].

Укажем два практически важных способа корректной "стыковки" алгоритмов классификации и регрессии. Один из них основан на объединении двух задач в одну. Так, принимая модель смеси (84), параметры регрессии определяют при помощи оценок параметров π_j и θ_j в (84). Действительно, при расщеплении смеси нормальных распределений оценивают математические ожидания и ковариационные матрицы каждого из исходных классов (описываемых плотностями $f(x,\theta_j)$), а этого достаточно для нахождения регрессии. Недостатками этого способа "стыковки" являются: "привязка" к определенной параметрической модели (84), ограничение свободы выбора алгоритма классификации, большой объем вычислений.

Второй способ основан на использовании методов устойчивой регрессии, не опирающихся на предположение нормальности. При этом метод предварительной классификации может быть любым, но результаты расчетов относятся не к исходным классам в модели типа (84), а именно к тем таксонам (кластерам), что выделены алгоритмом классификации.

Мы видим, что двухэтапность обработки данных, при которой на первом этапе выделяются объекты нечисловой природы - кластеры, влечет необходимость выполнения определенных требований на втором этапе, а также предъявляется определенные требования к интерпретации результатов расчетов. Здесь методология статистики нечисловой природы вторгается в классическую область многомерного статистического анализа.

6.7. Использование оптимизационной формулировки ряда задач прикладной статистики

Основные задачи прикладной статистики допускают оптимизационную формулировку [11, 12], а потому предельная теория решений экстремальных статистических задач [10] позволяет получать полезные следствия для них. Так, результаты, относящиеся к экстремумам аддитивных статистик, можно непосредственно применить к статистикам минимального контраста. Частными случаями оценок минимального контраста являются оценки максимального правдоподобия, устойчивые оценки Тьюки-Хубера, оценки параметров в задаче аппроксимации (параметрической регрессии). Состоятельность оценок минимального контраста означает состоятельность всех перечисленных оценок, а также справедливость законов больших чисел в пространствах произвольной природы. (Отметим, что результаты [10 - 12] обобщают результаты [54].) Поэтому каждая общая теорема типа полученных в [10 - 12] влечет за собой соответствующие следствия, касающиеся перечисленных и других конкретных областей. Так, например, в задаче конструирования факторов [55] результаты [10 - 12] описывают поведение отношения, аппроксимирующего систему матриц.

В качестве примера рассмотрим подробнее метод главных компонент. Пусть $\xi_1, \xi_2, ..., \xi_n$ - независимые одинаково распределенные случайные вектора размерности p. Кратко опишем экстремальную задачу, решаемую в методе главных компонент. Введем в рассмотрение координаты векторов: $\xi_j = (\xi_j(1), \xi_j(2), ..., \xi_j(p)), j = 1,2,...,n$. Рассмотрим p' линейных комбинаций

$$z_\infty(i) = \sum_{1 \le k \ge p} c_{ik} (\xi_1(k) - M\xi_1(k)), \ i = 1,2,..., p'. \qquad (98)$$

В методе главных компонент используется функционал

$$I(C) = \frac{D(z_\infty(1)) + D(z_\infty(2)) + ... + D(z_\infty(p'))}{D(\xi_1(1)) + D(\xi_1(2)) + ... + D(\xi_1(p))}, \qquad (99)$$

где $C=||c_{ik}||$. Формула (99) относится к вероятностной модели. При анализе статистических данных аналогом $I(C)$ является функционал $I_n(C)$, в котором теоретические дисперсии заменены выборочными. Легко видеть, что при $n \to \infty$ для любой матрицы C

$$I_n(C) \to I(C) \qquad (100)$$

(сходимость по вероятности). Рассмотрим решения экстремальных задач

$$C_n = \underset{C}{Arg\,\min}(-I_n(C)), \ C_\infty = \underset{C}{Arg\,\min}(-I(C)) \qquad (101)$$

Легко видеть, что условия асимптотической равномерной разбиваемости [10 - 12] выполнено, а потому

$$\lim_{n \to \infty} C_n = C_\infty \qquad (102)$$

по вероятности, с учетом единственности решений задач (101).

В литературе по методу главных компонент (см., например, обзор [56]), теорему о справедливости соотношения (102) обнаружить не удалось. Основное внимание уделяется нереалистическому случаю многомерной нормальности.

В ряде других задач прикладной статистики решения находятся путем минимизации функционала, также не являющегося аддитивным. Таковы различные варианты задач классификации, решаемые путем минимизации функционала качества, факторный анализ, метод экстремальной группировки признаков, отбор наиболее информативных признаков в моделях дискриминантного анализа, построение множества наиболее информативных переменных в моделях восстановления зависимостей (некоторые постановки разобраны выше в разделе 3), скалярная редукция многокритериальной оптимизационной схемы, т.е. экспертно-статистический метод построения обобщенного показателя "качества" в случае, когда экспертная информация - ранжировки, разбиения или результаты парных сравнений [57]. Во всех перечисленных задачах результаты [10-12] позволяют изучить асимптотическое поведение получаемых решений. Мы не будем подробно расписывать соответствующие результаты, поскольку это означало бы дать обзор основных задач прикладной статистики (см., в частности, [12, 58, 59]), обширный по объему, но не содержащий принципиально новых идей по сравнению со сказанным выше в настоящей главе и предыдущих публикациях.

ГЛАВА 7. ОСНОВНЫЕ ТРЕБОВАНИЯ К МЕТОДАМ АНАЛИЗА ДАННЫХ (НА ПРИМЕРЕ ЗАДАЧ КЛАССИФИКАЦИИ)

Во всех отраслях промышленности, в медицине, социально-экономических исследованиях и других областях деятельности постоянно решаются разнообразные задачи классификации. Разработано много различных математических методов классификации. Строго говоря, их не меньше, чем точек на отрезке. Действительно, ряд методов использует только расстояния между классифицируемыми объектами. Однако, если d - расстояние (метрика), то d^{α} - также метрика при любом α таком, что $0 < \alpha < 1$.

Несмотря на многообразие постановок задач, моделей и методов классификации, алгоритмов расчетов, положение дел в этой области анализа данных далеко от удовлетворительного. Задачи классификации зачастую решаются не наилучшим образом (более того, зачастую не ясно, как сравнивать методы решения). Области применимости различных методов классификации не установлены, свойства методов недостаточно

изучены. Отдельные группы специалистов (кланы) разрабатывают собственные подходы, не слишком интересуясь результатами других. Популярность тех или иных методов зачастую определяется субъективными причинами. Распространен ряд сомнительных концепций и попросту заблуждений. Во многом трудности определяются тем, что накоплено столько теоретических и практических разработок, что отдельный специалист или небольшая группа не в состоянии их осмыслить.

Назрела необходимость навести порядок в методах классификации. Это повысит их роль в решении прикладных задач, в частности, при диагностике материалов. Решить поставленную задачу можно только с помощью добровольной стандартизации. Необходимо проанализировать накопленное и разработать стандарты (предприятий и организаций) по применению признанных наилучшими метолов классификации. Для этого, прежде всего, следует выработать требования, которым должны удовлетворять методы классификации. Первоначальная формулировка таких требований - основное содержание настоящей работы.

Методы классификации рассматриваем как часть прикладной статистики. Ниже приводим ряд примеров нарушения обсуждаемых требований к методам анализа данных, при этом критика конкретной публикации не означает, что в ней нет ничего ценного.

7.1. Требования к методам анализа данных и представлению результатов расчетов

1. Методы должны быть объективными, результат их применения - определяться исходными данными, но не субъективными мнениями и решениями исследователя. В частности, *методы анализа данных должны быть инвариантны относительно допустимых преобразований шкал, в которых измерены данные, т.е. методы должны быть адекватны в смысле теории измерений* [1]. Это требование иногда бывает довольно жестким. Так, в качестве средних величин для данных, измеренных в порядковой шкале, можно использовать только члены вариационного ряда, в частности, медиану, но не среднее арифметическое, среднее гармоническое и т.д. Из всех средних по Колмогорову условие адекватности выделяет для данных, измеренных в интервальной шкале, только среднее арифметическое, а для шкалы отношений - только степенные средние [2].

Иногда градациям порядковых данных пытаются приписать числа, с тем, чтобы потом применять методы, разработанные для количественных шкал. Это - так называемая "оцифровка" [3, 4]. Она частично оправдана лишь в том случае, когда есть уверенность, что наблюдаемые данные получены в результате группировки количественных переменных. Пропаганда методов "оцифровки" вне указанных пределов может привести

к неадекватным рекомендациям и повлечь те или иные потери. Примером неадекватной оцифровки является метод анализа иерархий [5], в котором от порядковых переменных осуществляется переход к измерениям в шкале интервалов.

2. Основой конкретного статистического метода анализа данных всегда является та или иная вероятностная модель. Именно на основе модели осуществляется перенос выводов с выборочной совокупности на более широкую (генеральную) совокупность. Модель должна быть явно описана, ее предпосылки обоснованы - либо из теоретических соображений, либо экспериментально. Так, в математической статистике часто предполагается, что данные представляют собой выборку, т.е. моделируются как реализации набора независимых одинаково распределенных случайных величин. В обосновании нуждаются, в частности, независимость, одинаковая распределенность. Обоснование используемой модели может быть дано либо из содержательных соображений (например, на основе анализа условий наблюдений), либо же путем статистической проверки. Так, критерии независимости результатов наблюдений приведены в [6, 7]. Иногда высказываемое мнение [8], что положениям математической статистики не угрожает опытная проверка, не соответствует действительности. Построением вероятностно-статистических моделей в связи с задачами классификации занимался Л.Г. Малиновский [9].

Модель и метод (алгоритм) - две самостоятельные составляющие процедуры анализа данных. Для одной и той же модели могут быть предложены различные алгоритмы. Например, параметры функции распределения можно оценивать методом моментов, методом максимального правдоподобия и др. Отметим здесь, что итеративные процедуры нахождения оценок максимального правдоподобия применять нецелесообразно: если эти оценки нельзя найти явно, то следует вычислять одношаговые оценки [1].

Более важно, что один и тот же алгоритм в одной модели может быть наилучшим из возможных, а в другой - очень плохим. Так, для проверки однородности двух выборок в классической модели, в которой элементы выборки имеют нормальные распределения, критерий Стьюдента является наилучшим (при условии равенства дисперсий). Если же распределения, из которых взяты выборки, могут быть произвольными, то этот критерий несостоятелен. К сожалению, неправильное понимание критерия Стьюдента укоренилось, например, в медицинской науке. Следует, конечно, переучивать прикладников на непараметрические критерии.

Полезно сказать несколько слов в защиту критерия Стьюдента. Во-первых, распределение статистики Стьюдента устойчиво к малым отклонениям от нормальности. Во-вторых, из Центральной Предельной Теоремы следует, что статистика Стьюдента распределена асимптотически

нормально, если объемы обеих выборок стремятся к бесконечности, а распределения, из которых они взяты, имеют дисперсии. Отсюда следует, что критерий Стьюдента является состоятельным для проверки гипотезы о равенстве математических ожиданий двух распределений. Если последняя гипотеза отвергнута, то однородности нет (подробности см. в [10]).

Аналогичное замечание можно сделать по поводу распространенного неправильного мнения о том, что проверять равенство 0 линейного парного коэффициента корреляции Пирсона можно только в случае, когда результаты наблюдений имеют двумерное нормальное распределение. На самом же деле выборочный коэффициент корреляции асимптотически нормален, а потому при большом объеме выборки можно пользоваться теми же процедурами, что и в предположении нормальности [1].

Проверка однородности - одна из процедур классификации. Именно, проверяется, представляют ли выборки два класса или же их можно объединить в один. Каким же непараметрическим критерием пользоваться? В литературе имеется много предложений. Например, в [7] предлагается применять критерий Вилкоксона. Эта рекомендация не соответствует традициям отечественной вероятностно-статистической научной школы [11], рекомендующей критерии, основанные на эмпирических функциях распределения. Обсудим обоснованность рекомендации по применению критерия Вилкоксона.

В [7] критерий Вилкоксона опирается на модель, в которой одна из функций распределения произвольна, а вторая отличается от нее только сдвигом. Редко можно указать ситуацию, в которой подобная модель обоснована. Разве что при анализе результатов многократных измерений значений физической величины для двух образцов с помощью одного и того же средства измерения, для которого характеристики погрешностей стабильны в рассматриваемом диапазоне.

Если реальная ситуация достаточно изучена, то функции распределения в основном известны. Под таким заявлением обычно понимают то, что они известны с точностью до параметров, а тогда проверка гипотезу однородности проводится с помощью параметрических критериев, в частности, при нормальных распределениях с одинаковыми дисперсиями - с помощью критерия Стьюдента.

Если же реальная ситуация изучена мало, то функции распределения естественно считать произвольными и не связанными друг с другом. Затруднительно представить себе ситуацию, в которой связь между функциями распределения известна почти полностью (с точностью до параметра сдвига), в то время как о самих функциях распределения ничего не известно. Авторы [7] не рассматривают такие ситуации, в соответствующем примере [7, с.87-88] они попросту не обосновывают модель. Таким образом, несведущий в прикладной статистике исследователь, пользуясь [7], может взять произвольную модель,

обработать данные в соответствии с ней, результат расчетов выдать как научно обоснованный.

(Отметим, что название [7] не соответствует содержанию: эту монографию следовало бы назвать "Избранные ранговые статистические методы". В [7] несколько искажена история непараметрической статистики, полностью игнорируются такие ее современные разделы, как непараметрические оценки плотности и регрессии. Современные взгляды на непараметрическую статистику обсуждаются в статье [12]).

Итак, при проверке однородности в непараметрическом случае необходимо принять, что функции распределения выборок произвольны. В такой постановке критерий Вилкоксона не является состоятельным. Значит, его применять нельзя. Чем же пользоваться? Очевидно, состоятельными критериями - Смирнова, типа омега-квадрат (Лемана-Розенблатта) [11] и др. Каким именно? Это - нерешенная проблема, подходов к которой не видно (она стоит первой в "цахкадзорской тетради" [13]). Если известна альтернатива, то можно подобрать наиболее мощный критерий. Но откуда взять альтернативу?

Ясно, что нельзя ждать, пока наука дозреет до решения этой проблемы. В настоящее время мы считаем целесообразным рекомендовать два критерия - двухсторонний критерий Смирнова и типа омега-квадрат (Лемана-Розенблатта). В пользу первого из них говорит то, что разработан быстрый алгоритм вычисления распределения критерия Смирнова при конечных объемах выборок, на основе которого рассчитаны таблицы критических значений, исчерпывающим образом дополняющие таблицы для предельного распределения [14]. (Отметим, что называть этот критерий "критерием Колмогорова - Смирнова", как это сделано в [7], неправильно, поскольку у Колмогорова и Смирнова не было ни одной совместной работы, рассматриваемый критерий был предложен Н.В. Смирновым в 1939 г., причем, вопреки [7], метод нахождения предельного распределения статистики Смирнова никак не связан с методом известной работы А.Н. Колмогорова 1933 г., в которой введен "критерий Колмогорова".) Однако у критерия Смирнова имеется заметный недостаток - его функция распределения растет большими скачками, а потому реальный уровень значимости может сильно отличаться от номинального [15]. Поэтому в настоящее время [16] мы склоняемся к рекомендации о применении типа омега-квадрат (Лемана-Розенблатта).

Приведенная выше критика критерия Вилкоксона относится также и к его обобщениям, применяемым в так называемом "непараметрическом дисперсионном анализе" [7] (кстати, название это неточно, поскольку никаких "дисперсий" в рассматриваемых непараметрических методах нет). В рассматриваемых постановках также необходимо перейти на состоятельные критерии.

Таким образом, на примере проверки гипотезы однородности показана необходимость обоснования вероятностной модели реального явления и ее взаимосвязь с алгоритмом расчетов, а также продемонстрирован ряд типичных ошибок.

3. Методы обработки данных, предназначенные для использования в реальных задачах, должны быть исследованы на устойчивость относительно допустимых отклонений исходных данных и предпосылок модели. В частности, должна указываться точность решений, определяемая по точности исходных данных. При этом каждый отдельный элемент исходных данных (например, элемент выборки) рассматриваем как представитель кластера, сгустка с размытыми границами, определяемыми погрешностями исходных данных. Решения, даваемые моделью, описываются, естественно, как элементы кластера - образа кластера данных. Этот подход подробно рассмотрен в монографиях [17, 18], а применительно к теории классификации - в [19] и других статьях. Здесь отметим только два применения развитой нами общей теории устойчивости.

Анализ погрешностей социологических данных привел нас к выводу, что в социологических (и маркетинговых) анкетах не имеет смысла использовать более 3 - 6 градаций [17, п.2.6]. Различие значений параметров моделей управления запасами, определяемых по методикам тех или иных организаций, приводило отдельных экономистов к выводу о невозможности использования оптимизационных моделей. Анализ с позиций теории устойчивости показал, что все рассматриваемые значения лежат в одном и том же кластере, определяемом погрешностями, а анализ кластера решений дал возможность сделать вывод, что оптимизационная модель позволяет снизить издержки не менее чем в 2 раза [17, п.5.1].

Заслуживает дальнейшего развития связь разработанной нами теории устойчивости с теорией решения некорректных задач [20] и с теорией нечеткости. Отметим, что в [17, 18] указан способ сведения теории нечеткости к теории случайных множеств, что позволяет рассматривать теорию нечеткости как своеобразный частный вероятностно-статистический метод. Ясно также, что нечеткость границ реально существующих кластеров должна учитываться в алгоритмах кластер-анализа, т.е. во многих реальных задачах адекватной является лишь нечеткая классификация.

4. Должна указываться точность решений, даваемых с помощью используемого метода. Понятие "точность" конкретизируется для отдельных классов методов. Так, погрешности решения могут быть связаны с погрешностями исходных данных, с погрешностями округления при компьютерных вычислениях, с погрешностями выбранного численного метода решения строго поставленной математической задачи, с тем, что математическая модель лишь грубо отражает действительность,

и т.д. Особенно важно уметь численно оценивать погрешности при использовании так называемых "эвристических" алгоритмов, таких, как алгоритм [21], о котором авторы честно пишут, что не знают, дает ли он решение поставленной оптимизационной задачи.

Надо констатировать, что каждый метод обработки данных - это косвенное измерение [1, 17, 18]. Перед массовым использованием, как и всякий метод измерения, он должен быть обоснован с позиций метрологии (науки об измерениях). Поскольку аналитические методы при конечных объемах выборок зачастую не разработаны, то напрашивается изучение точности решений с помощью метода Монте-Карло. Однако следует знать, что многие используемые ныне датчики псевдослучайных чисел дают последовательности, свойства которых явно отличаются от номинальных при числе испытаний, скажем, более 2000, как это установлено И.Г. Журбенко и его сотрудниками еще в 1980-х годах [22].

Явный учет погрешностей может привести к неожиданным выводам. Так, для гамма-распределения еще Р. Фишер в 1920-х годах сравнивал по эффективности метод моментов оценки параметров и метод максимального правдоподобия, и последний оказался лучше. Когда же мы в [1] учли погрешности наблюдений, то вывод оказался другим - в обширной области исходных данных метод моментов лучше метода максимального правдоподобия.

Большой материал по рассматриваемым вопросам дан в весьма ценной книге [23]. Однако, по нашему мнению, авторы [23] слишком много внимания уделяют нынешнему состоянию прикладной математики по сравнению с обсуждением путей развития. Кроме того, методы анализа данных предлагаются, по нашей оценке, прежде всего для их массового использования, поэтому, в согласии с [23, гл.2], необходимо их тщательное исследование. Однако в настоящее время бесконтрольно распространяется большое число плохо обоснованных методов (некоторые примеры ошибок даны выше). Это представляет, на наш взгляд, большую опасность, поскольку с развитием цифровизации происходит стандартизация статистического инструментария на основе стандартных пакетов прикладных статистических программ. Опасность состоит в возможности проникновения в стандартные пакеты плохо обоснованных методов. Подобные методы есть даже в лучших современных пакетах [24]. Необходимы широкие и глубокие исследования имеющихся методов анализа данных, нацеленные на создание "золотого фонда", рекомендуемого для массового использования. Пример такой попытки - система государственных стандартов по статистическим методам управления качеством продукции, прежде всего серия ГОСТов по прикладной статистике ГОСТ 11.001-73 - ГОСТ 11.011-83. К сожалению, попытка провалилась - во многих стандартах этой системы были

обнаружены грубые ошибки [25]. Причина - некомпетентность ряда разработчиков.

Очевидно, целесообразно провести анализ методов классификации, нацеленный на создание "золотого фонда". Для этого необходимо провести ряд исследований в духе описанных в статье [19]. Надо также навести порядок в терминологии: вряд ли допустимо, чтобы одна и та же область имела массу названий - кластер-анализ, распознавание образов без учителя, таксономия, автоматическая классификация и т.д.

Нужно обсуждать и показатели качества классификации. Так, например, при классификации на два класса в качестве подобного показателя часто используют долю ошибочно классифицированных объектов. Это, однако, нерационально. Если доля одного из классов сравнительно мала, то вполне обоснованный алгоритм может по этому показателю оказаться хуже тривиального, согласно которому следует отнести все объекты к более многочисленному классу. Так, ряд работ группы И.М. Гельфанда посвящен прогнозированию исхода инфаркта миокарда (использовался алгоритм "Кора-3"). Если для больного прогнозировался неблагоприятный исход (смерть), то за больным следовало установить специальное наблюдение и применять интенсивное лечение - такова практическая польза применения здесь метода классификации. Ясно, что риск смерти целесообразнее несколько переоценить, чем недооценить. На это и ориентировался алгоритм группы И.М. Гельфанда. А вот по доле ошибочной классификации он оказался хуже тривиального, согласно которому предлагалось считать, что никому из больных не угрожает смерть. Одна из возможных рекомендаций [26] - сравнивать методы классификации путем пересчета на модель линейного дискриминантного анализа, в котором классы описываются многомерными нормальными распределениями с одинаковыми ковариационными матрицами. Тогда можно оценить расстояние Махаланобиса между классами и сравнивать методы классификации с его помощью - чем это расстояние больше, тем метод классификации лучше. Пусть доля правильно классифицированных объектов в первом классе есть α, а во втором - β. Тогда оценкой расстояния Махаланобиса между классами является

$$d = \Phi^{-1}(\alpha) + \Phi^{-1}(\beta), \qquad (1)$$

а в качестве оценки прогностической силы алгоритма классификации вместо доли правильных прогнозов рекомендуется использовать прогностическую силу

$$\gamma = \Phi\left(\frac{d}{2}\right) = \Phi\left(\frac{\Phi^{-1}(\alpha) + \Phi^{-1}(\beta)}{2}\right),$$

где $\Phi(.)$ - функция нормального распределения с нулевым математическим ожиданием и единичной дисперсией, а Φ^{-1} - обратная к Φ функция.

5. В большинстве случаев анализируются данные о выборке с целью переноса на более широкую совокупность, в частности, для прогноза поведения вновь появляющегося объекта. Необходимо указывать точностные характеристики метода, т.е. точность оценивания по выборке параметров и характеристик модели. В вероятностных моделях это делается с помощью доверительных множеств, которыми обычно являются доверительные интервалы.

С прикладной точки зрения метод, для которого неизвестны точностные характеристики, является недостаточно разработанным, другими словами, поисковым, экспериментальным, эвристическим. Его нельзя рекомендовать для массового использования. Его применение может оказаться полезным, а может привести к грубым ошибкам, т.е. он является "магическим" в терминологии В.Н. Тутубалина [27].

Суть дела проста: интуиция обманывает, представляет метод гораздо более точным, чем он есть на самом деле. Современному научно-техническому уровню отвечают работы, в которых наряду с точечными оценками даны доверительные границы. Отходят от этого требования как несведущие в статистике лица, так и, к сожалению, отдельные преподаватели высшей школы, в том числе университетов, что объясняется, видимо, сочетанием "академичности" и отрыва от массы специалистов, обрабатывающих реальные данные.

При публикации результатов статистического анализа реальных данных необходимо указывать их точность (доверительные интервалы). Иначе невозможно использование этих результатов в дальнейших исследованиях в качестве исходных данных (поскольку неизвестны "допустимые отклонения исходных данных" - см. монографии по методам анализа устойчивости выводов [17, 18]), а также сравнение результатов различных исследований. К сожалению, данные социологических, медицинских и иных исследований часто публикуются без указания их точностных характеристик. Потом с содержательной точки зрения (т.е. с точки зрения конкретной прикладной ситуации) обсуждают, например, причины различия показателей для двух групп, в то время как статистические данные, которые можно извлечь из работы, не позволяют заключить о значимости рассматриваемого различия. Имеется в виду частный случай задачи, рассмотренной выше - проверка однородности для независимых выборок из двух биномиальных распределений. Так вот, если есть две выборки объема 100, в первой положительных ответов - 47%, а во второй - 61%, то различие незначимо (на уровне значимости 5%). Но социолог этого не знает - точностные характеристики не указаны - и начинает наводить теорию ... В журнале "Химия и жизнь" (1976, №4, с.112-113) всерьез обсуждалась связь между специальностью ученого и знаком Зодиака, под которым он родился, хотя элементарный подсчет по критерию хи-квадрат показывает, что никакой связи нет (см. подробный

разбор в [28, гл.2]). Достойно сожаления, что отдельные специалисты по математическим методам в социологии всерьез воспринимают так называемый "детерминационный анализ" [29], котором используются сравнительно малые по численности группы и игнорируются точностные характеристики, что толкает на получение неадекватных выводов (отметим, что с математической точки зрения "детерминационный анализ" покрывается одним из параграфов книги Г.С. Лбова [30]. Малограмотны и претенциозны высказывания о статистических методах в науковедении в книге [31] ... Впрочем, все ошибки не перечислишь. Напомним хотя бы о хроническом непонимании области применимости критерия Колмогорова, разобранном нами в статье [32] и других работах.

По нашему мнению, неточны слова К. Джини [33, с.29]: "Нельзя предпочесть метод, который не отвечает определенной цели, методу, отвечающему цели, только на том основании, что в одном случае вычислена, а в другом еще не вычислена вероятная ошибка". Как можно знать, что "метод отвечает цели", если его точность неизвестна? В частности, лучше ли он тривиального метода - принять решение априори, а на данные вообще не смотреть. Из сказанного ясно, что мы считаем неверным и мнение Е.С. Вентцель [34] о том, что построению доверительных интервалов не следует уделять большого внимания.

В последние десятилетия получили распространение "невероятностные методы обработки данных", или "анализ данных" (в узком смысле). Типичными публикациями по анализу данных являются статья [21] и книги [29, 30, 35]. Как правило, методы анализа данных - это эвристические методы, вопрос о точностных характеристиках которых даже не ставится. Справедливо сказано в [35, с.15]: "Анализ данных применяется на первых этапах теоретического познания исследуемого явления". Очевидно, за первыми этапами должны следовать дальнейшие, имеющие целью развитие вероятностно-статистической теории, т.е. построение адекватной вероятностной модели явления и на ее основе теоретически обоснованных правил принятия решений (например, решений о необходимости наладки технологического процесса). Таким образом, анализ данных содержит методы, которые можно сравнить с "времянками": они первыми появляются на месте будущих зданий, а после окончания строительства подлежат сносу. Это поисковые, магические, а не научно обоснованные методы, их нельзя рекомендовать для широкого использования, включать в нормативно-техническую документацию - до оценок точности получаемых с их помощью решений, что в большинстве случаев возможно лишь с помощью вероятностной модели. Последняя необходима, если полученные по выборке результаты распространяются на более широкую совокупность. Если же интересующие специалиста включены в исследование, то точность понимается в соответствии с теорией устойчивости [17, 18]. Реальная опасность состоит в том, что в

условиях современного обилия публикаций и программ, оборотной стороной чего является относительное невежество специалистов (нельзя знать и 5% от более чем миллиона актуальных к настоящему времени публикаций по математической статистике), распространение получат недостаточно обоснованные методы анализа данных. Ясно ведь, что времянку легче построить, чем здание ... Отметим, что в строительстве времянки стоят десятки лет. Как говорят: "Нет ничего более постоянного, чем "временное"".

Отметим, что математические методы исследования делятся на "разведочный анализ" и "доказательную статистику". Разведочный анализ нацелен на обнаружении нового, в то время как цель доказательной статистики - строго обосновать выводы. Например, разведочный анализ дает возможность сформулировать статистическую гипотезу, а доказательная статистика позволяет ее обосновать (принять) на выбранном заранее уровне значимости.

Многие методы анализа данных основаны на максимизации какого-либо функционала. Надо подчеркнуть, что наличие оптимизации не делает метод более научным, она - средство, а не цель. В связи с обсуждением оценивания параметров гамма-распределения [1] уже приводились примеры того, что не основанные на оптимизации методы могут быть лучше оптимизационных. Польза от экстремальной формулировки основных задач прикладной статистики состоит в основном в том, что можно едиными методами изучать асимптотическое поведение решений этих задач [36], а также единообразно строить алгоритмы их решения. Наиболее естественная оптимизационная постановка задач кластер-анализа дана А.Н. Колмогоровым (см. [17]).

6. Специфические требования к методам обработки данных возникают в связи с их "стыковкой" при последовательном выполнении [13, 19]: результаты работы предыдущего алгоритма должны удовлетворять условиям, наложенным на исходные данные последующего. Так, "восстановление пропущенных данных" по какому-либо алгоритму приводит к тому, что полученная матрица "объект-признак" не может рассматриваться как составленная из независимых случайных векторов, т.е. классическое предположение математической статистики: "наблюдения есть выборка" (конечная последовательность независимых одинаково распределенных случайных элементов" - не выполнено; следовательно, применение основанных на этом предположении методов не является обоснованным. Аналогичная ситуация имеет быть при "преобразовании данных", если параметры преобразования определяются по исходным данным. Неясной остается на настоящий момент обоснованность регрессионного анализа, если степень полинома, описывающего линию регрессии, подбирается по экспериментальным данным, поскольку распространенные оценки этой степени

несостоятельны [37]. Продолжать можно долго. К сожалению, нельзя априори надеяться, что влияние указанных нарушений исходных предпосылок мало. Так, в критериях согласия Колмогорова, омега-квадрат и др. возникает желание вместо неизвестных параметров подставить их оценки. Этот прием аналогичен рассмотренным выше, но, в отличие от них, последствия его применения хорошо изучены. Влияние велико и не уменьшается с ростом выборки, например, при применении критерия Колмогорова для проверки нормальности процентные точки должны быть уменьшены примерно в 1,5 раза по сравнению с классическими [32].

Распространена рекомендация - разбить совокупность на однородные классы и затем анализировать каждый класс отдельно. Рекомендация рациональна (в смысле [23]). Так, при обработке данных о течении острой пневмонии [38] коэффициент корреляции между возрастом и длительностью заболевания оказался сравнительно малым (r = 0,21). Когда же мы выделили группы курящих и некурящих, то в первой из них связь оказалась гораздо более выраженной (r = 0,53), во второй же - незначимой.

В рассмотренной задаче классы выделены по априорным соображениям. Если же дискриминирующая поверхность (разделяющая классы) строится на основе анализа экспериментальных данных, то попавшие в один класс наблюдения, вообще говоря, не образуют выборку (нарушается независимость), а распределения их не являются нормальными. Для естественной модели показано [19], что при росте объема выборки независимость в определенном смысле восстанавливается, в то время как распределение элементов кластера отнюдь не приближается к распределению соответствующего члена в смеси, описывающей исходную совокупность (в частности, плотность этого распределения равна 0 для обширной области пространства). Следовательно, нельзя применять регрессионный анализ, основанный на предположении нормальности.

7. Требования к представлению результатов статистического анализа частично рассмотрены выше. Результаты должны приводиться вместе с точностными характеристиками, с указанием конкретного метода, с помощью которого они получены, и степени его обоснованности. При использовании информационно-коммуникационных технологий следует указывать тип (марку, название) компьютера, язык программирования, время счета и другие необходимые характеристики.

Кроме указанных выше, можно сформулировать ряд иных требований к методам обработки данных и представлению их результатов [39].

7.2. О границах применимости вероятностно-статистических методов

Этой теме посвящены многочисленные публикации [8, 9, 27, 40 - 43]. Мы ее также кратко касались [17, 44, 45]. Здесь отметим только два обстоятельства, весьма кратко и не претендуя на окончательность.

1. По нашему мнению, применение вероятностных методов не имеет принципиальных отличий от применений других областей математики, как более старых (геометрия, дифференциальные уравнения), так и более новых (теория нечеткости [45]). Схема применения однотипна: строится модель на основе соответствующей области математики, тем или иным способом она обосновывается, на основе модели реального явления изучаются интересующие специалистов вопросы, полученные выводы интерпретируются и используются для принятия решений. Поразительно, что отдельные авторы полностью игнорируют многочисленные способы проверки адекватности вероятностной модели.

2. Не менее поразительно, что возможность применения вероятностных моделей связывают с "темными понятиями" устойчивости частот, статистической однородности, статистического ансамбля [8, 27, 43]. Вот уже более 80 лет теория вероятностей является аксиоматической наукой (мы основываемся на аксиоматике А.Н. Колмогорова [46]; его основополагающая монография впервые издана в 1933 г. на немецком языке и в 1936 г. на русском). В ней нет места перечисленным "темным понятиям", как и бессмысленному, вообще говоря, понятию "генеральная совокупность" (оно имеет смысл лишь в случае выбора из конечного множества). Понятие статистического ансамбля, как и выражение "теория вероятностей изучает закономерности массовых явлений" - это реликты начала XX века, когда не отделяли математическую теорию вероятностей от её приложений. Попытки применить эти понятия сводятся к бездоказательным общим рассуждениям (другими словами, демагогии), поскольку любая научно обоснованная проверка должна опираться на вероятностную модель явления. На наш взгляд, движение одной-единственной частицы или развитие уникальной экономической системы вполне могут описываться случайными процессами (в терминологии теории вероятностей) - если соответствующие вероятностные модели обоснованы. Например, как проверить, что $w(t)$ - траектория винеровского процесса? Возьмем разности $w(\tau)-w(0), w(2\tau)-w(\tau), w(3\tau)-w(2\tau),\ldots,$, где τ достаточно мало по сравнению с интервалом наблюдения. Тогда, как известно, все эти разности независимы и имеют одинаковое нормальное распределение с нулевым математическим ожиданием и дисперсией τ - при справедливости гипотезы о винеровости. Остается проверить, справедливо ли утверждение, сформулированное в предыдущей фразе, с помощью широкого известных статистических критериев [1, 6, 7, 11 и др.].

7.3. О некоторых постановках задач классификации

1. Если классы полностью описаны или заданы обучающими выборками, классификацию можно рассматривать как измерение. При статистическом контроле качества единицы продукции классифицируются на годные и бракованные. Врач ставит диагноз больному, относя тем самым его заболевание к одной из нозологических форм. Измерение в номинальной шкале есть разбиение объектов на классы, а в порядковой - на упорядоченные классы [17]. Ясно, что результат измерения должен быть воспроизводимым, допускать сравнение с результатами других измерений. Вообще, классификация как средство измерения должна удовлетворять требованиям, устанавливаемым метрологией. Необходимым условием этого является стандартизация правил классификации (это условие не является, однако, достаточным: сплошь и рядом контролирующие органы обнаруживают, что пропущенные службами контроля качества изделия не удовлетворяют требованиям соответствующих нормативных документов). Ясно, что без стандартизации правил классификации не могут работать различные автоматизированные системы управления, действующие на предприятиях и в регионах. В статистике говорят о точном определении используемых понятий, рассматриваемых совокупностей [33, 47].

Хотя с необходимостью применения стандартных классификаций обычно никто не спорит, на практике стандартизация не всегда осуществлена. Сотрудники вузов хорошо это знают, сравнивая оценки в школьных аттестатах и на экзаменах. Мне уже приходилось упоминать [45] о двух группах медиков, по определению одной из которых "затяжное течение острой пневмонии" имело место в 6% случаев, а по мнению другой - в 60% (для той же совокупности из 461 больного)! Неточности классификаций приводят к тому, что экономико-статистические данные имеют относительные ошибки 5-10% [47].

2. В ряде случаев "мы хотим разбить объекты на группы независимо от того, естественны границы разбиения или нет" [48, с.437]. Типичные примеры - использования интервалов группировки в статистике, разбиение студентов специальности по учебным группам.

3. "Проблема классификации (в узком смысле слова - А.О.) состоит в выяснении по эмпирическим данным, насколько элементы "группируются" или распадаются на изолированные "скопления", "кластеры"" " [48, с.467]. Рассматриваемую область прикладной статистики естественно называть кластер-анализом. В этой области наиболее обоснованными являются вероятностно-статистические методы, известные как методы расщепления смесей [19]. При использовании тех или иных алгоритмов возникает проблема "реальности кластера" [19]. Дело в том, что алгоритм кластер-анализа можно применить к любым исходным данным, в том числе к выборке из однородной совокупности. В последнем случае, очевидно,

результат работы алгоритма не будет иметь реального смысла. Как отличить эту ситуацию от противоположной, когда совокупность действительно разбивается на кластеры? Приведем пример ошибочного применения кластер-анализа.

Качество одного из продуктов нефтехимии - фенола - характеризуют 13 показателей. На их измерения тратятся большие средства. Идея состоит в том, чтобы разбить признаки на группы и из каждой группы оставить только один, при этом "каждый из признаков внутри одной группы говорит б образцах почти одно и то же" [49, с.23]. Последнее означает, что коэффициенты корреляции между признаками одной группы близки к 1. По экспериментальным данным нашли матрицу выборочных коэффициентов корреляции [49, с.25]. Максимальный по величине коэффициент корреляции равен 0,85, следующий за ним - 0,46. Отсюда ясно, что только 2 признака из 13 связаны между собой настолько, что имеет смысл прогнозировать значение одного из них по-другому, да и для них прогнозирование не слишком хорошее. Однако это не смущает Ю.П. Адлера, он, не колеблясь, применяет метод корреляционных плеяд и получает 6 групп. Одна из них состоит из двух показателей, коэффициент корреляции между которыми равен 0,21 [49, с.25], т.е. с помощью одного из них можно объяснить лишь 4% дисперсии второго. Обоснованный (с позиций прикладной статистики) ответ в рассматриваемой задаче таков: показатели практически нельзя объединить в группы (за исключением двух, коэффициент корреляции между которыми равен 0,85); чтобы не потерять информацию, надо измерять не менее 12 показателей. Однако Ю.П. Адлер считает, что достаточно 6 - по одному из группы [49, с.24]. Это - введение заказчика в заблуждение с использованием авторитета математических методов. Интересно подсчитать убытки, вызванные описанной рекомендацией Ю.П. Адлера.

Если кластеры являются реальными, то любой разумный алгоритм кластер-анализа должен их достаточно точно выделить. Другими словами, результат кластер-анализа должен быть устойчив относительно выбора алгоритма [17, 18]. Следовательно, для выделения реальных кластеров можно рекомендовать наиболее простой в определенном смысле алгоритм, например, требующий наименьших вычислений, скажем, алгоритм ближнего соседа [1]. Затем следует проверить устойчивость полученных кластеров по отношению к допустимым отклонениям исходных данных [19].

Приведем пример. В [50] мы обрабатывали анкеты (типа социологических) способных к математике школьников. Для кластер-анализа признаков, измеренных в номинальных шкалах, был выбран алгоритм [21], который мы сочли под влиянием [21] наиболее перспективным и обоснованным. Реализация алгоритма на компьютере и счет заняли около полугода. Позже я за полтора часа обработал вручную те

же данные по упомянутому выше алгоритму ближнего соседа. Результаты (дендрограммы) практически совпали. Более того, алгоритм ближнего соседа дал дополнительную информацию о структуре данных Итак, в случае работы [50] цена ошибочного выбора алгоритма - полгода лишней работы плюс стоимость машинного времени (вторая составляющая в рассматриваемое время была заметной).

Самый радикальный способ сократить затраты на кластер-анализ - заранее объявить совокупность однородной. Так, Ю.Н. Тюрин [51] пишет: "При проведении экспертного опроса обычно считают, что по интересующему предмету существует истинная точка зрения". Если же выявились кластеры различных мнений, то "надо признать, что экспертный опрос не достиг окончательной цели" [51, с.11]. По моему мнению, это слишком категоричное заявление. Оно может повлечь исключение из процедур обработки экспертных данных этапа кластер-анализа, а это может привести к ошибкам в содержательных областях. На практике мнения экспертов зачастую разделяются (например, мнения научных работников и производственников). Мы полагаем, что при применении экспертных технологий необходим этап классификации мнений экспертов (отметим, что в [52] в модели люсианов порождения экспертных оценок удалось из вероятностно-статистических соображений указать ограничение сверху на диаметр кластера, т.е. обосновать выбор итогового разбиения из дендрограммы).

Один из разделов статьи Н. Бурбаки "Архитектура математики", основополагающей для многотомной серии "Элементы математики", называется так: "Стандартизация математических орудий" [53, с.253]. Наша задача - стандартизовать такое мощное орудие, как методы классификации. В настоящей главе раскрыт ряд положений работ [54 - 56].

ГЛАВА 8. ПРИМЕНЕНИЕ МЕТОДА МОНТЕ-КАРЛО ПРИ ИЗУЧЕНИИ СВОЙСТВ СТАТИСТИЧЕСКИХ КРИТЕРИЕВ ОДНОРОДНОСТИ ДВУХ НЕЗАВИСИМЫХ ВЫБОРОК

Среди математических и инструментальных методов экономики важное место занимают метод статистических испытаний (Монте-Карло). Он широко используется при разработке, изучении и применении математических методов исследования в эконометрике, прикладной статистике, организационно-экономическом моделировании, при разработке и принятии управленческих решений.

В развитии математических методов исследования выделяем два важных периода [1]. Первый - начало XX в., когда были разработаны базовые положения современной математической статистики,

сформулированы основные идеи таких ее разделов, как описание данных, оценивание параметров, проверка статистических гипотез. Эти идеи легли в основу учебников, используемых и в настоящее время. Наряду с рациональными приемами анализа данных продолжают пропагандироваться устаревшие воззрения, например, основанные на использовании параметрических семейств распределений вероятностей, в то время как установлено, что практически все распределения реальных данных ненормальны и не описываются с помощью иных семейств распределений вероятностей.

Второй период - с 1980-х годов по настоящее время. Усилиями сотен исследователей разработана новая парадигма прикладной статистики [2]. Фактически речь идет о новой парадигме математических методов исследования [3]. В соответствии с новой парадигмой заложены основы математики XXI в. - системной нечеткой интервальной математики [4]. На первое место вышла статистика нечисловых данных. Так, за десять лет (2006 - 2015) ей посвящены 27,6% всех публикаций раздела "Математические методы исследования" журнала "Заводская лаборатория. Диагностика материалов", т.е. 63,0% статей по прикладной статистике [5].

Новая парадигма математических методов исследования (см. главу 1 выше) опирается на эффективное применение информационно-коммуникационных технологий как при расчете характеристик методов анализа данных, так и при имитационном моделировании. Датчики псевдослучайных чисел лежат в основе многих современных технологий анализа данных. Эти эффективные инструменты исследователя внутренне противоречивы - в них с помощью детерминированных алгоритмов получаем последовательность чисел, обладающих многими свойствами случайных величин. Поэтому свойства таких инструментов требуют тщательного изучения.

8.1. Метод статистических испытаний - инструмент исследователя

Для решения конкретных прикладных задач исследователи постоянно разрабатывают новые методы обработки статистических данных - результатов измерений (наблюдений, испытаний, анализов, опытов) и экспертных оценок. Свойства каждого вновь предлагаемого метода необходимо изучить. Какие интеллектуальные инструменты можно применить для такого изучения?

Мощным инструментом исследователей в области математической статистики являются предельные теоремы теории вероятностей - закон больших чисел, центральная предельная теорема и т.п. Некоторые ориентированные на математику специалисты призывают ими и ограничиться. Однако для практического использования статистических методов предельных теорем недостаточно. Необходимо найти границу -

выяснить, начиная с какого объема выборки можно пользоваться результатами, полученными с помощью предельных теорем. И выяснить, как принимать решения, если объем имеющихся данных меньше этой границы.

С середины XX в. исследователю доступна универсальная "отмычка" - метод статистических испытаний (метод Монте-Карло), другими словами, имитационное моделирование. Он основан на использовании последовательности псевдослучайных чисел, свойства которых напоминают свойства рассматриваемых в теории вероятностей случайных величин. Основная идея состоит в последовательном выполнении следующих этапов: (1) разработке вероятностно-статистической модели реального явления или процесса; (2) планировании статистического испытания, в котором случайные величины заменяются псевдослучайными, полученными с помощью того или иного датчика псевдослучайных чисел; (3) проведении большого числа испытаний (тысяч или миллионов); (4) анализе полученных результатов расчетов.

С каждым этапом связаны соответствующие проблемы адекватности имитационного моделирования. Так, для предельных теорем обычно справедлив тот или иной принцип инвариантности, т.е. в пределе исчезает зависимость от конкретного вида распределения. Однако при изучении скорости сходимости выбор этого конкретного вида весьма важен, поскольку от него зависит итоговый результат статистического моделирования - один для нормального распределения, другой - для логистического, третий - для распределения Коши...

Датчики псевдослучайных чисел лишь имитируют случайность. Алгоритмы получения псевдослучайных чисел имеют достаточно короткое описание, в то время как по определению А.Н. Колмогорова 60-х годов (в рамках теории информации) описание случайной последовательности должно расти пропорционально длине этой последовательности [6]. Кроме этой глобальной причины методологической несостоятельности датчиков псевдослучайных чисел есть и частные недостатки. Например, у некоторых популярных до настоящего времени датчиков три последовательных значения связаны линейной зависимостью.

Значения, рассчитанные с помощью метода Монте-Карло, имеют погрешности, определяемые конечностью числа испытаний. При оценивании вероятности события погрешность достигает величины $1/(2\sqrt{N})$, где N - число испытаний. Значит, для оценивания вероятности с точностью 10^{-6} необходимо 10^{12} / 4 испытаний. На практике провести такое количество испытаний невозможно.

8.2. Дискуссия о современном состоянии и перспективах развития статистического моделирования

Проблемы теории и практики статистических испытаний (Монте-Карло) заслуживают тщательного обсуждения. В 2016 г. журнал "Заводская лаборатория. Диагностика материалов" начал дискуссию о современном состоянии и перспективах развития статистического моделирования, т.е. теории и практики применения метода статистических испытаний (Монте-Карло), различных вариантов имитационного моделирования. Предыдущая дискуссия о свойствах таких датчиков была проведена в журнале "Заводская лаборатория. Диагностика материалов" в 1985 - 1993 гг.

"Затравкой" дискуссии послужили статьи [7] и [8]. В первой из них рассмотрены задачи повышения эффективности вычислений методом Монте-Карло. Отмечено, что ключевую роль в их решении играют вопросы выбора объема статистических испытаний (количества моделируемых случайных чисел), а также качества соответствующих датчиков случайных чисел. Обсуждены проблемы реализации алгоритмов методов Монте-Карло, обусловленные требованиями повышения скорости сходимости асимптотических решений к истинным решениям.

В статье [8] констатируется, что цель прикладной математической статистики - разработка методов анализа данных, предназначенных для решения конкретных прикладных задач. С течением времени подходы к разработке таких методов менялись. Сто лет назад принимали, что распределения данных имеют определенный вид, например, являются нормальными, и исходя из этого предположения развивали статистическую теорию. На следующем этапе на первое место в теоретических исследованиях выдвинулись предельные теоремы. Под «малой выборкой» понимают такую выборку, для которой нельзя применять выводы, основанные на предельных теоремах. В каждой конкретной статистической задаче возникает необходимость разделить конечные объемы выборки на два класса: для одного можно применять предельные теоремы, а для другого делать этого нельзя из-за риска получения неверных выводов. Для выбора границы часто используют метод Монте-Карло (статистических испытаний). Более сложные проблемы возникают при изучении влияния на свойства статистических процедур анализа данных тех или иных отклонений от исходных предположений. Такое влияние также часто изучают, используя метод Монте-Карло. Основная и пока не решенная в общем виде проблема при изучении устойчивости выводов при наличии отклонений от параметрических семейств распределений состоит в том, какие распределения использовать для моделирования. Сформулированы и другие нерешенные проблемы.

Подборка из трех статей опубликована в мартовском номере 2017 г. О.И. Кутузов и Т.М. Татарникова [9] рассмотрели две задачи, обусловленные особенностями применения имитационного моделирования при исследовании сложных технических систем. Одна из них связана с реализацией подхода к повышению эффективности метода Монте-Карло при моделировании редких событий: сочетание расслоенной выборки с равновзвешенным моделированием позволяет значительно ускорить алгоритмический анализ моделей стохастических систем методом имитации. Решение другой задачи выявило проблему, связанную с неадекватностью использования одного и того же датчика псевдослучайных чисел при сопоставлении выборочных значений очередей, полученных на имитационных моделях фрактальной и классической систем массового обслуживания.

И.З. Аронов и О.В. Максимова [10] представили результаты статистического моделирования, характеризующие зависимость времени достижения консенсуса от числа членов технических комитетов по стандартизации (ТК) и их авторитарности. Использована математическая модель обеспечения консенсуса в работе ТК, основанная на модели, предложенной Де Гроотом. Проведен анализ основных проблем достижения консенсуса при разработке консенсусных стандартов в условиях предложенной модели. Показано, что увеличение числа экспертов ТК и их авторитарности негативно влияет на время достижения консенсуса и способствует разобщенности группы.

В комментарии [11] к этой статье проанализировано соотношение консенсуса и истины. Работа технических комитетов по стандартизации - одна из форм экспертных процедур, поэтому ее целесообразно рассматривать в рамках теории и практики экспертных оценок. Тогда проблема консенсуса - это проблема согласованности мнений членов комиссии экспертов. Однако цель работы экспертной комиссии - не достижение согласованности экспертов (консенсуса), а получение (в качестве коллективного мнения) выводов, отражающих реальность, обычно нацеленных на выработку обоснованных управленческих решений, короче говоря, на получение истины. Наблюдаем объективное противоречие между стремлением к выявлению истины и желанием обеспечить консенсус.

Итоги первого этапа дискуссии подведены в [12]. Продолжают публиковаться статьи, посвященные применению метода статистических испытаний (Монте-Карло) для решения различных задач. Так, М.С. Жуков применяет его для изучения свойств алгоритмов нахождения медианы Кемени как итогового мнения комиссии экспертов [13], а И.В. Гадолина и Н.Г. Лисаченко - при разработке метода построения доверительных интервалов для процентилей случайной выборки прочности композитов

[14]. Столь интересно начатая дискуссия заслуживает продолжения и расширения круга обсуждаемых проблем.

Обсудим применение метода статистических испытаний для изучения свойств статистических критериев проверки однородности двух независимых выборок.

8.3. Статистические критерии проверки однородности двух независимых выборок

Исходные данные - две выборки $x_1, x_2,...,x_m$ и $y_1, y_2,...,y_n$ (т. е. наборы из m и n действительных чисел), требуется проверить их однородность.

В общепринятой модели $x_1, x_2, ..., x_m$ - независимые одинаково распределенные случайные величины с функцией распределения $F(x)$, а $y_1, y_2, ..., y_n$- также независимые одинаково распределенные случайные величины, но с, вообще говоря, другой функцией распределения $G(x)$.

Разделяют однородность характеристик (равенство математических ожиданий, или медиан, или дисперсий и т.п.) и однородность (совпадение) функций распределения (абсолютную однородность). Во втором случае речь идет о проверке нулевой гипотезы:

$$H_0 : F(x)=G(x) \text{ при всех } x.$$

Отсутствие однородности означает, что верна альтернативная гипотеза, согласно которой

$$H_1 : F(x_0)\neq G(x_0) \text{ хотя бы при одном значении аргумента } x_0.$$

Если гипотеза H_0 принята, то выборки можно объединить в одну, если нет - то нельзя.

Рассмотрим следующие статистические критерии, предназначенные для проверки однородности двух независимых выборок.

(1) Критерий Крамера-Уэлча T, совпадающий при равенстве объемов выборок ($m = n$) с критерием Стьюдента t [15].

(2) Критерий Лорда, или модифицированный t-критерий [16, табл.3.10, с.42] со статистикой

$$L = \frac{2\left|\bar{x} - \bar{y}\right|}{\left(\max\limits_{1\le i\le m} x_i - \min\limits_{1\le i\le m} x_i\right)+\left(\max\limits_{1\le j\le n} y_j - \min\limits_{1\le j\le n} y_j\right)}.$$

(3) Критерий Вилкоксона (Манна-Уитни) ([16, табл.6.8, с.94]. [17]), основанный на статистике U - сумме рангов элементов первой выборки в общем вариационном ряду.

(4) Критерий Вольфовица (серий) V, основанный на количестве серий в общем (объединенном) вариационном ряду (серия - часть последовательности, состоящая из элементов одной выборки) и разобранный в [16, табл.6.7].

(5) Критерий Ван-дер-Вардена [16, табл.6.9], основанный на статистике

$$X = \sum_{i=1}^{m} \Psi\left\{ \frac{r_i}{m+n+1} \right\},$$

где r_i - ранг i-го элемента первой выборки в общем вариационном ряду, $\Psi(t)$ - функция, обратная к функции стандартного нормального распределения $\Phi(x)$.

(6) Критерий Смирнова [16, 18], основанный на статистике

$$S = D_{m,n} = \sup_x \left| F_m(x) - G_n(x) \right|,$$

где $F_m(x)$ - эмпирическая функция распределения, построенная по первой выборке, а $G_n(x)$ - эмпирическая функция распределения, построенная по второй выборке.

(7) Критерий типа омега-квадрат [16, 18], предложенный Леманом [19] изученный впервые Розенблаттом [20], а потому называемый критерием Лемана-Розенблатта. Этот критерий основан на статистике

$$\omega^2 = \omega_{mn}^2 = \frac{mn}{m+n} \int_{-\infty}^{+\infty} \left(F_m(x) - G_n(x) \right)^2 dH_{m+n}(x),$$

где $H_{m+n}(x)$ - эмпирическая функция распределения, построенная по объединенной выборке.

За пределами перечня остались многие критерии - хи-квадрат [21], Сэвиджа [22], знаков [23], основанные на последовательных рангах [24] и другие.

8.4. Постановка задачи изучения статистических критериев методом статистических испытаний

С помощью вычислительных экспериментов по изучению свойств критериев однородности двух выборок можно выяснить, при каких объемах выборок можно пользоваться предельными распределениями. Ясно, что ответ определяется заданной исследователем точностью (максимально возможным отклонением допредельного распределения от предельного на заданном отрезке или на всей прямой). Можно сравнивать критерии по мощности при тех или иных конкретных альтернативах (например, альтернативах сдвига или масштаба). Представляет интерес анализ "корреляции" критериев на основе изучения доли совпадающих решений по результатам проверки статистических гипотез с помощью этих критериев (эта задача допускает несколько вариантов постановок - можно сравнивать критерии при фиксированном уровне значимости, например, 0,05, можно использовать несколько уровней значимости, можно установить связь между достигаемыми уровнями значимости, ...).

Поскольку статистики ранговых критериев принимают лишь конечное число значений, то их распределения дискретны. Поэтому они "проскакивают" обычно используемые в таблицах [16, 23] номинальные

уровни значимости - 0,01; 0,05; 0,1 и др. Особенно существенным это обстоятельство оказывается для статистик, принимающих небольшое число значений, таких, как статистика Смирнова: реальный уровень значимости статистического критерия может быть в несколько раз меньше номинального - например, равняться 0,02 вместо 0,05 [21, 25]. Сравнение непараметрических критериев затрудняется тем, что по указанной причине невозможно обеспечить совпадение их уровней значимости. Казалось бы, можно использовать рандомизированные критерии. Однако использование таких критериев не соответствует большинству практических задач, в которых проверяется однородность двух конкретных выборок, в то время как рандомизированные критерии нацелены на обработку большого числа однотипных выборок фиксированных объемов.

Таким образом, многообразие перспективных вычислительных экспериментов обширно. Для обеспечения изучения свойств различных критериев проверки гипотез однородности нами совместно с Ю.Э. Камнем и Я.Э. Камнем разработан программный продукт, состоящий из четырех блоков: генерации равномерно распределенных на [0; 1] псевдослучайных чисел; вычисления на их основе псевдослучайных чисел с заданными законами распределения; блока расчета значений статистик критериев и блока сервисных и управляющих программ.

При моделировании использовался датчик равномерно распределенных на множестве $\{1,2,...,2^{15}-1\}$ псевдослучайных чисел [26], построенный на основе рекуррентной формулы

$$x_{n+1} = (1285x_n + 6925) \bmod (2^{15}), n = 1,2,... \quad (1)$$

Тестирование [27] этого датчика с помощью критерия Колмогорова для выборок объема 5000 на уровне значимости 2,5% показало согласие с равномерным распределением. Поскольку далее гипотеза однородности проверяется при уровне значимости 0,05, то погрешность метода Монте-Карло оценивается как

$$\pm\sqrt{\frac{0,05\times 0,95}{5000}} = \pm 0,003 .$$

Как отмечал акад. АН СССР Ю.В. Прохоров (1929 - 2013) на "неформальном обсуждении" проблем статистического моделирования, проведенном в рамках Первого Всемирного конгресса Общества математической статистики и теории вероятностей им. Бернулли [28], применения метода Монте-Карло можно разделить на два класса. В первом из них, появившемся исторически раньше, качество датчика определяется соответствием распределения даваемых датчиком псевдослучайных чисел заданному распределению, например, равномерному. Выполнения этого условия достаточно, например, для вычисления многомерных интегралов. Именно этот класс применений обычно имеется в виду в литературе по методу Монте-Карло [29, 30]. Для применений из второго класса

существенно обеспечить независимость псевдослучайных чисел, точнее, достаточное для успешного применения датчика приближение к независимости. Как показано в работах И.Г. Журбенко с соавторами [31 - 33], датчики типа (1) принципиально не могут обеспечить независимость. Однако из расчетов Г.В. Рыдановой [34] следует, что последовательности из не более чем 24 псевдослучайных чисел, используемые для одного статистического испытания, есть основания рассматривать как модели последовательностей независимых случайных величин.

Коротко говоря, наша позиция по поводу метода Монте-Карло такова. Мы активно используем этот метод в научных исследованиях. В частности, для изучения скорости сходимости распределений статистик - в предельной теории помех, создаваемых электровозами [35], в теории люсианов [36], при изучении свойств критериев однородности [25]. Но одновременно отдаем себе отчет в недостатках этого инструмента и предостерегаем от его бездумного употребления [37].

Для постановки вычислительного эксперимента необходимо задать две функции распределения $F(x)$ и $G(x)$ - функции распределения элементов двух выборок. Обоснованных теорией или практикой рекомендаций по выбору $F(x)$ и $G(x)$ в настоящее время нет. Поэтому для поискового исследования будем использовать привычные нормальные распределения и распределения Вейбулла - Гнеденко.

Функция распределения Вейбулла - Гнеденко имеет вид

$$F(x;a,b,c)=\begin{cases}1-\exp\left\{-\left(\dfrac{x-a}{b}\right)^{c}\right\}, & x>a,\\ 0, & x\le a,\end{cases}$$

где a - параметр сдвига, b - параметр масштаба, c - параметр формы.

Нормально распределенные псевдослучайные числа находились методом обратной функции [38, с. 440, ф-ла (12.10)]. Распределение Вейбулла - Гнеденко моделировалось по [39, с.93].

8.5. Вычислительные эксперименты

Приведем некоторые результаты изучения свойств критериев однородности двух независимых выборок в двух случаях:

$F(x)$ и $G(x)$ - функции нормального распределения;

$F(x)$ и $G(x)$ - функции распределения Вейбулла-Гнеденко

- как с одинаковыми, так и с различными значениями параметров.

В первом случае первая выборка бралась из стандартного нормального распределения с математическим ожиданием 0 и дисперсией 1, а вторая - из нормального распределения с математическим ожиданием m_2 и дисперсией σ_2^2, где значения m_2 и σ_2 приведены в табл.1.

Таблица 1. Проверка равенства математических ожиданий для выборок из нормальных распределений по критерию Крамера-Уэлча

Номер вычислительного эксперимента	Объем выборок $m = n$	Параметры второй выборки		Частота принятия нулевой гипотезы H_0	Вероятность принятия H_0 (исходя из распределения Стьюдента)	Вероятность принятия H_0 (исходя из нормального распределения)
		m_2	σ_2			
1	6	0	1	0,969	0,974	0,950
2	7	0	1	0,954	0,971	0,950
3	8	0	1	0,956	0,968	0,950
4	10	0	1	0,958	0,961	0,950
5	6	1	1	0,596	0,691	0,592
6	6	1,5	1	0,366	0,356	0,262
7	8	2	1	0,048	0,032	0,021
8	12	3	1	0	0	0
9	6	0	1,5	0,948	0,974	0.950
10	8	0	2	0,938	0,968	0,950
11	6	0	3	0,930	0,974	0,950
12	10	0	3	0,934	0,961	0,950

Во втором случае параметр масштаба b функции распределения Вейбулла-Гнеденко во всех выборках принят равным 1. Первая выборка бралась (при всех экспериментах, кроме четырех) при $a = 0$ и $c = 1$, т.е. из экспоненциального распределения с функцией распределения

$$F(x) = \begin{cases} 1 - \exp\{-x\}, & x > 0, \\ 0, & x \le 0. \end{cases}$$

Вторая выборка бралась из распределений Вейбулла-Гнеденко с параметрами a, $b = 1$, c, приведенными в табл.2 (там же оговорены исключения).

Таблица 2. Проверка равенства математических ожиданий для выборок из распределений Вейбулла-Гнеденко по критерию Крамера-Уэлча

Номер вычислительного эксперимента	Объем выборок $m = n$	Параметры второй выборки		Частота принятия нулевой гипотезы H_0	Вероятность принятия H_0 (исходя из распределения Стьюдента)	Вероятность принятия H_0 (исходя из нормального распределения)
		a	c			
1	6	0	1	0,956	0,974	0,950
2	10	0	1	0,954	0,961	0,950
3	6	0,5	1	0,828	0,912	0,861
4	8	0,5	1	0,772	0,874	0,829
5	10	0,5	1	0,750	0,837	0,800
6	6	1	1	0,750	0,689	0,592
7	8	1	1	0,450	0,558	0,484
8	10	1	1	0,348	0,446	0,313
9	6	0	1,5	0,950	0,971	0,946

10	8	0	1,5	0,950	0,963	0,946
11	10	0	1,5	0,956	0,958	0,942
12	6	0	2	0,940	0,949	0,943
13	8	0	2	0,944	0,954	0,938
14	10	0	2	0,928	0,949	0,935
15	12	0	2	0,944	0,950	0,935
16	8	0	3	0,930	0,961	0,942
17	12	0	3	0,942	0,949	0,935
18	6	0	5	0,904	0,971	0,945
19	8	0	5	0,910	0,963	0,944
20	10	0	5	0,920	0,958	0,943
21	12	0	5	0,940	0,955	0,941
22	8	0	1	0,928	0,968	0,950
23*	6	0	3	0,946	0,974	0,950
24*	10	0	3	0,928	0,961	0,950
25*	6	0,5	3	0,292	0,553	0,447
26*	10	0,5	3	0,094	0,273	0,228
27	6	0,5	3	0,690	0,905	0,850
28	10	0,5	3	0,676	0,826	0,781

* В экспериментах 23 - 26 первая выборка взята из распределения Вейбулла-Гнеденко с параметрами $a = 0$, $b = 1$, $c = 3$.

Выбор распределений для экспериментов определяется наряду с желанием сравнить свойства статистик на выборках из нормального семейства распределений, для которых статистики Стьюдента и Крамера-Уэлча имеют определенные оптимальные свойства, так и желанием рассмотреть класс распределений, существенно отличающихся от нормальных, в частности, несимметричностью. Экспоненциальное распределение часто используют при изучении показателей надежности [40], поэтому оно и было включено в эксперименты.

Из семи перечисленных в п.4 критериев однородности критерий Вольфовица (серий), как установлено, имеет малую мощность. Поэтому его исключение из дальнейших рассмотрений не приводит к отрицательным последствиям.

В табл.3 приведены результаты экспериментов для выборок из нормальных распределений. Табл. 3 соответствует табл. 1 - при совпадающих номерах речь идет об одних и тех же экспериментах. В табл. 4 для облегчения анализа свойств критериев приведены относительные мощности критериев по отношению к критерию Крамера-Уэлча (совпадающего с критерием Стьюдента в рассматриваемых экспериментах). В табл. 4 стоят отношения двух случайных величин - оценки мощности рассматриваемого критерия, полученной по 5000 испытаниям, к оценке мощности критерия Крамера-Уэлча.

В табл. 5 приведены результаты экспериментов для выборок из распределений Вейбулла-Гнеденко. Табл. 5 соответствует табл. 2 - при совпадающих номерах речь идет об одних и тех же экспериментах.

Таблица 3. Частоты принятия гипотезы однородности для выборок из нормальных распределений

№	Объем выбо-рок $m = n$	Параметры второй выборки		Частоты принятия нулевой гипотезы H_0 для критериев					
				1	2	3	5	6	7
		m_2	σ_2	t	L	U	X	S	ω^2
1	6	0	1	0,969	0,970	0,976	0,976	0,982	0,976
2	7	0	1	0,954	0,968	0,986	0,964	1,00	0,956
3	8	0	1	0,956	0,958	0,50	0,954	0,994	0,944
4	10	0	1	0,958	0,960	0,974	0,972	0,998	0,958
5	6	1,0	1	0,596	0,624	0,680	0,698	0,754	0,690
6	6	1,5	1	0,366	0,390	0,464	0,474	0,616	0,496
7	8	2,0	1	0,048	0,054	0,064	0,078	0,480	0,084
8	12	3,0	1	0	0	0	0	0	0
9	6	0	1,5	0,948	0,952	0,974	0,976	0,980	0,972
10	8	0	2,0	0,938	0,940	0,930	0,950	0,998	0,904
11	6	0	3,0	0,930	0,924	0,950	0,956	0,934	0,920
12	10	0	3,0	0,934	0,902	0,930	0,946	0,988	0,846

Таблица 4. Мощность критериев относительно критерия Крамера-Уэлча (для экспериментов №№ 5-12 в табл. 1 и 3)

№ эксперимента	Относительная мощность критериев				
	2	3	5	6	7
	L	U	X	S	ω^2
5	0,931	0,792	0,748	0,609	0,767
6	0,727	0,739	0,783	0,000	0,957
7	0,993	0,983	0,968	0,546	0,962
8	1,000	1,000	1,000	1,000	1,000
9	0,923	0,500	0,460	0,385	0,538
10	0,806	1.129	0,806	0,030	1,548
11	1,086	0,714	0,629	0,943	1.140
12	1,485	1,061	0,818	0,182	2,323

Таблица 5. Частоты принятия гипотезы однородности для выборок из распределений Вейбулла-Гнеденко

№	Частоты принятия нулевой гипотезы H_0 для критериев					
	1	2	3	5	6	7
	t	L	U	X	S	ω^2
1	0,956	0,942	0,964	0,968	0,964	0,960
2	0,966	0,940	0,956	0,962	0,998	0,952
3	0,828	0,818	0,840	0,858	0,878	0,840
4	0,772	0,764	0,720	0,746	0,974	0,698
5	0,750	0,724	0,678	0,668	0,962	0,634

6	0,528	0,514	0,534	0,586	0,552	0,482
7	0,450	0,432	0,354	0,392	0,756	0,336
8	0,348	0,344	0,268	0,272	0,662	0,206
9	0,950	0,934	0,958	0,958	0,974	0,958
10	0,950	0,936	0,946	0,954	1,000	0,950
11	0,956	0,934	0,954	0,956	1,000	0,946
12	0,940	0,934	0,962	0,970	0,960	0,952
13	0,944	0,922	0,930	0,972	0,984	0,958
14	0,928	0,900	0,930	0,932	0,990	0,894
15	0,944	0,918	0,938	0,944	0,932	0,898
16	0,930	0,906	0,906	0,924	0,986	0,884
17	0,942	0,908	0,910	0,930	0,816	0,786
18	0,904	0,886	0,934	0,946	0,876	0,866
19	0,910	0,872	0,866	0,896	0,972	0,790
20	0,920	0,872	0,874	0,920	0,968	0,714
21	0,940	0,886	0,862	0,908	0,662	0,606
22	0,928	0,908	0,944	0,952	0,998	0,944
23	0,946	0,948	0,964	0,966	0,978	0,970
24	0,928	0,934	0,948	0,944	0,998	0,936
25	0,292	0,312	0,392	0,408	0,578	0,430
26	0,094	0,100	0,142	0,132	0,654	0,144

При анализе табл. 3 и 5 необходимо иметь в виду отличие реальных уровней значимости $\alpha_р$ от номинальных $\alpha_н$ (см. [25]). Особенно это касается критерия Смирнова. Различие между собой реальных уровней значимости у свободных от распределения статистик делает трудным сравнение между собой критериев по мощности - такое сравнение желательно проводить при одном и том же уровне значимости, но это невозможно.

Как и должно быть согласно теории математической статистики [41], для выборок из нормального распределения наиболее мощным оказался критерий Крамера-Уэлча (Стьюдента). Близким к нему по мощности оказались критерий Лорда и критерий типа омега-квадрат. Отметим: критерий Лорда использует размахи, а потому неустойчив к засорениям на "хвостах"; следовательно, возможность его использования при анализе реальных данных в каждом конкретном случае требует специального обоснования. Критерии Вилкоксона и Ван-дер-Вардена также имеют высокую мощность, особенно в экспериментах №№ 6,8. Малая мощность критерия Смирнова объясняется, видимо, отличием $\alpha_р$ от $\alpha_н$.

Другая картина наблюдается при изменении дисперсии. Критерии Крамера-Уэлча и Лорда слабо реагируют на нее. Еще меньше реагируют линейные ранговые статистики *U* и *X*. Критерий Вилкоксона не может, даже асимптотически, различить нормальные совокупности с одинаковыми математическими ожиданиями, но разными дисперсиями [17]. Обращает на себя внимание высокая мощность критерия омега-квадрат (см. табл.4, эксперименты №№ 10 - 12).

Для выборок из распределений Вейбулла-Гнеденко картина несколько иная. Если распределения отличаются только сдвигом (эксперименты №№ 3 - 8), то наибольшую мощность имеет критерий омега-квадрат, затем идут критерии Вилкоксона и Ван-дер-Вардена, после них - критерии Стьюдента и Лорда, наименьшая мощность у критерия Смирнова. Если же изменяется также и параметр формы (см., например, эксперимент № 21), то наибольшая мощность также у критерия омега-квадрат, следующим является критерий Смирнова (с учетом отличия α_p от $\alpha_н$). Заметно также существенное возрастание мощности с ростом объемов выборок и увеличением различия параметров.

На основе анализа таблиц 3 - 5 можно сформулировать, с понятными оговорками, следующие практические рекомендации.

А. Для проверки гипотезы абсолютной однородности (гипотезы совпадения функций распределения двух выборок) целесообразно использовать критерий Лемана - Розенблатта типа омега-квадрат [18] - во всех случаях.

Б. Если есть основания предполагать, что распределения отличаются в основном сдвигом, то целесообразно использовать линейные ранговые критерии Вилкоксона и Ван-дер-Вардена. Однако даже в этом случае критерий омега-квадрат может оказаться более мощным.

В. Из рассмотренных критериев для проверки гипотезы однородности в общем случае, кроме критерия ω^2, можно использовать критерий Смирнова - с учетом отличия реального уровня значимости от номинального.

8.6. Частота совпадений статистических выводов по разным критериям

По итогам обработки данных с помощью определенного критерия однородности принимают одно из двух решений: "гипотеза однородности отклоняется" или "гипотеза однородности не отклоняется". Решения по разным критериям могут не совпадать. Насколько часты расхождения?

Были изучены доли (в %) расхождений решений по критериям *L, U,X, S,* ω^2 с решениями по критерию Крамера-Уэлча. Для описания полученных результатов введены "зоны". Пусть t_n - критическое значение для критерия Крамера-Уэлча, соответствующее уровню значимости α = 0,05 и объему выборок $m = n$. Используется абсолютное значение статистики критерия Крамера-Уэлча. Введено 8 зон: 1 - [0; $t_n/4$), 2 - [$t_n/4$; $t_n/2$), 3 - [$t_n/2$; $3t_n/4$), 4 - [$3t_n/4$; t_n), 5 - [t_n; $5t_n/4$), 6 - [$5t_n/4$; $3t_n/2$), 7 - [$3t_n/2$; $7t_n/4$), 8 - [$7t_n/4$; $+\infty$).

В качестве примера проведенных исследований приведем в табл. 6 данные по вычислительному эксперименту № 19 для выборок из распределений Вейбулла - Гнеденко (см. табл. 2). В строке "*T* (частота

попадания в зону)" приведено асимптотическое распределение статистики Крамера - Уэлча (сгруппированное по зонам). В каждой строке, соответствующей определенному критерию, для каждой зоны указана доля совпадений решений по этому критерию с решением по критерию Крамера - Уэлча.

Таблица 6. Доли совпадений решений по критериям *L, U, X, S,* ω^2 с решениями по критерию Крамера-Уэлча *T* (эксперимент № 19)

Критерии	Доля принятия H_0 по *T*	Зоны							
		1	2	3	4	5	6	7	8
T (частота попадания в зону)	0,910	0,362	0,274	0,194	0,080	0,042	0,022	0,008	0,018
L	0,872	1	1	0,979	0,575	1	1	1	1
U	0,866	0,956	0,978	0,897	0,775	0,562	0,937	1	1
X	0,896	0,978	0,978	0,928	0,900	0,500	0,875	1	1
S	0,972	1	0,985	0,990	1	0,125	0,125	0,750	0,555
ω^2	0,790	0,889	0,927	0,835	0,600	0,875	1	1	1

В качестве второго примера в табл. 7, построенной аналогично табл. 6, приведена сводка для экспериментов №№ 1-21 с выборками из распределений Вейбулла-Гнеденко. Табл.8 содержит информацию о расхождениях (в %) решений по критериям *L, U,X, S,* ω^2 с решениями по критерию Крамера-Уэлча.

Таблица 7. Сводка для выборок из распределений Вейбулла-Гнеденко (эксперименты №№ 1-21). Проценты расхождений с решениями по критерию Крамера-Уэлча.

Критерии	Зоны							
	1	2	3	4	5	6	7	8
L	0	0	0,9	24,4	6,6	0	0	0
U	1,1	2,2	5,8	22,5	30,7	3,0	0,4	0
X	0,5	1,3	3,5	17,7	34,6	6,2	1,2	0,3
S	3,3	3,4	4,4	10,4	77,8	60,3	46,0	17,0
ω^2	6,5	6,6	11,8	29,5	30,7	2,5	0	0

Таблица 8. Расхождения (в %) решений по критериям *L, U,X, S,* ω^2 с решениями по критерию Крамера-Уэлча.

По критерию Крамера-Уэлча	По другим критериям				
	L	*U*	*X*	*S*	ω^2
Принято 84,6%, из них отвергнуто	3,2	5,1	3,5	4,5	10,5
Отвергнуто 15,4%, из них принято	2,7	13,1	15,8	56,5	12,9
Проверено 100%, из них расхождений	3,1	6,3	5,4	12,4	10,9
По сравнению с критерием Крамера - Уэлча, %	- 2,3	- 2,3	- 0,5	+ 4,9	- 6,9

Из полученных результатов можно сделать ряд выводов.

Наибольший процент расхождений приходится на зоны № 4 (от 10.4% до 29,5% по табл. 7) и № 5 (от 6,6% до 77,8%), что естественно, т.к. при переходе от зоны 4 к зоне № 5 и происходит изменение решения по критерию Крамера-Уэлча. Обратим внимание, что расхождения имеются и в зоне №1 - для 6,5% экспериментов, попавших в эту зону, критерий Лемана - Розенблатта отвергает нулевую гипотезу (т.е. во всех этих случаях гипотеза однородности неверна). Вместе с тем нет ни одного случая, когда бы этот критерий принял гипотезу для экспериментов из зон 7, 8. Другими словами, если критерий Крамера - Уэлча отклоняет нулевую гипотезу с $T > 3{,}0$, то критерий ω^2 также отклоняет гипотезу однородности.

Наибольшее расхождение с критерием Крамера - Уэлча наблюдается у критерия Смирнова, в основном за счет принятия гипотезы в случае, когда T - критерий ее отверг. Это во многом объясняется существенным различием α_p и $\alpha_н$ для критерия Смирнова. Почти такое же суммарное число расхождений у критерия Лемана-Розенблатта, но причина иная - у этого критерия выше мощность, чем у критерия Крамера - Уэлча.

Наиболее близок к T-критерию критерий Лорда. Это подтверждается тем, что расхождения имеются лишь в зонах 4 и 5, и незначительное(0,9%) - в зоне 3.

По числу расхождений критерии Вилкоксона и Ван-дер-Вардена занимают промежуточное положение, они вдвое ближе к статистике Лорда, чем к критериям Смирнова и омега-квадрат. При этом критерий Ван-дер-Вардена ближе к T-критерию, чем критерий Вилкоксона, чего и следовало ожидать, учитывая нацеленность критерия Ван-дер-Вардена на применение к распределениям, близким к нормальным.

При справедливости гипотезы однородности расхождения не превышают 2,2 - 3,2% и проявляются в зонах 3 - 6. При альтернативе изменения параметра формы расхождения возрастают лишь для критериев Смирнова и ω^2 (до 8,4 - 9,9%), оставаясь в пределах 3,0 - 4,7% для остальных критериев, слабо реагирующих на эту альтернативу. При альтернативе сдвига расхождения резко возрастают (до 11,3% у критерия Вилкоксона, 10,2% - у критерия Ван-дер-Вардена, 24,4% - у критерия Смирнова, 15,8% - у критерия ω^2), оставаясь малыми (3,7%) лишь у критерия Лорда.

Можно сделать и ряд других выводов, например, проследить зависимость от объемов выборок и различия параметров. Проведенный нами более детальный анализ подтверждает сформулированные выше практические рекомендации А. Б, В (завершение раздела 6).

Обращает на себя внимание наличие значительного процента расхождений между решениями, принимаемыми по разным критериям. Этот факт необходимо учитывать при обработке конкретных данных в прикладных исследованиях и при разработке нормативно-технической и методической документации, программных продуктов и экспертных систем. В частности, в соответствии с общей теорией устойчивости [42] целесообразно анализировать данные одновременно с помощью нескольких критериев проверки гипотезы однородности двух независимых выборок и затем исходить из выводов, инвариантных относительно выбора критерия.

Таким образом, в соответствии с новой парадигмой математических методов исследования применение метода статистических испытаний (Монте-Карло) позволяет получить рекомендации для конкретных объемов выборок, в частности, оценить скорость сходимости допредельных распределений критериев проверки статистических гипотез к предельным распределениям.

ГЛАВА 9. СИСТЕМНАЯ НЕЧЕТКАЯ ИНТЕРВАЛЬНАЯ МАТЕМАТИКА И СОВРЕМЕННАЯ ЭКОНОМЕТРИКА

Системная нечеткая интервальная математика - основа преподавания современной эконометрики. Покажем это на примере учебного плана преподавания дисциплины "Эконометрика" в МГТУ им. Н.Э. Баумана.

Основополагающий вид управленческих инноваций в образовании - это инновации в содержании образования, в содержании изучаемых дисциплин и самого их перечня. Важным видом инноваций в высшем образовании являются авторские курсы учебных дисциплин, новизна которых состоит во введении в преподавание современных научных результатов. Главный принцип обучения специалистов в МГТУ им. Н.Э. Баумана «образование через науку» реализуется, в частности, путем разработки содержания подобных курсов [1]. В настоящей работе представлена информация об инновационном курсе эконометрики.

Эконометрика - это статистические методы в экономике и управлении. В наших учебниках мы исходим из этого определения [2 - 4]. Оно принято отечественной научной школой в области организационно-экономического моделирования, эконометрики и статистики [5 - 7], а также соответствует более развернутому определению:

"Эконометрика — наука, изучающая количественные и качественные экономические взаимосвязи с помощью математических и

статистических методов и моделей". Такое определение предмета эконометрики было выработано в уставе Эконометрического общества (основано в 1930 г.)" [8]. Однако в нем остается неясным сопоставление "математических и статистических методов и моделей". По нашему мнению, статистические методы опираются на такие разделы математики, как теория вероятностей и математическая статистика", т.е. входят в прикладную математику. Отметим, что изучение свойств статистических моделей и методов математическими средствами следует относить к фундаментальной математике.

9.1. О содержании учебной литературы по эконометрике

По данным Российского индекса научных исследований, в нашей стране выпущено около 400 учебников и учебных пособий по эконометрике. В тройку наиболее цитируемых, наряду с нашими книгами, входят учебники, подготовленные под руководством член-корр. РАН И.И. Елисеевой (см., например, [8]) и многочисленные издания книги Я.Р. Магнуса, П.К. Катышева и А.А. Пересецкого (см., например, [9]). Отметим, что в [8, с.15] описываются наши научные результаты, относящиеся к системной нечеткой интервальной математике, однако без упоминания фамилии автора, получившего эти результаты, и ссылок на литературные источники по рассматриваемым результатам.

Анализ содержания распространенных учебников по эконометрике показывает, что из многообразия статистических методов в экономике и управлении в них рассматриваются лишь небольшая часть - в основном линейные регрессионные модели (метод наименьших квадратов, проверка гипотез, гетероскедастичность, автокорреляция ошибок, спецификация модели т.п.). Все остальные статистические методы в экономике и управлении игнорируются. По нашему мнению, подобное неоправданное сужение сферы эконометрики связано как со слепым копированием устаревших западных учебников, так и с недостаточным знакомством авторов с практикой применения эконометрики при решении задач экономики и управления. Отметим однако, что системы эконометрических уравнений активно использовались во второй трети XX в. в макроэкономике. Возможно, именно это обстоятельство послужило причиной странного для нас сужения сферы эконометрических методов.

Констатируем, что большинство распространенных учебников и учебных пособий соответствует устаревшей парадигме эконометрики, выработанной в XX в., в то время как отечественная научная школа в области организационно-экономического моделирования, эконометрики и статистики развивается в соответствии с новой парадигмой XXI в. Новая парадигма широко обсуждалась научной общественностью [10 - 28]. Важно, что мы нацелены на использование эконометрики специалистами по экономике предприятия и организации производства, т.е. на

микроэкономическом уровне, а не при изучении макроэкономических соотношений.

Эконометрика - базовая научная, практическая и учебная дисциплина для экономистов и менеджеров. Методы эконометрики составляют значительную часть инструментов контроллинга [29 - 35]. При ее преподавании весьма важно преодолеть оковы устаревших взглядов XX в., излагая современную эконометрику. Полезным окажется опыт двадцатилетней реализации на факультете ИБМ МГТУ им. Н.Э. Баумана нашей авторской программы по эконометрике, которой и посвящена настоящая глава. Для определенности рассмотрим содержание курса "Эконометрика", реализованного в 2021 г. для студентов бакалавриата (два семестра, 34 часа лекций и 34 часа семинарских занятий). Изложение опирается на ранее изученные дисциплины "Теория вероятностей и математическая статистика", "Прикладная статистика" и в свою очередь служит основой дисциплины "Организационно-экономическое моделирование" (для магистрантов). В авторском курсе представлены начальные сведения по основным разделам современной эконометрики. В нем раскрыто выделенное нами содержание ядра современной эконометрики. Подходы, модели и методы, включенные в ядро дисциплины, должны составлять содержание учебного курса.

Настоящая глава посвящена основным составляющим современной эконометрики и их отражению в одноименной дисциплине, преподаваемой сотрудниками кафедры "Экономика и организация производства" МГТУ им. Н.Э. Баумана в течение более чем 20 лет.

9.2. Выборочные исследования

В первом семестре после определения эконометрики обосновываем необходимость выборочных исследований. В качестве примера - построение выборочной функции ожидаемого спроса и расчет оптимальной розничной цены при заданной оптовой цене (издержках) [36]. Вводим гипергеометрическую и биномиальную модели выборки значений альтернативных (дихотомических, бинарных) признаков, демонстрируем близость соответствующих им распределений в случае большого объема генеральной совокупности по сравнению с выборочной.

На основе теоремы Муавра-Лапласа теории вероятностей находим при безграничном росте объема выборки асимптотически нормальное распределение выборочной доли (в случае ответов типа «да» - «нет»). На его основе строим асимптотические доверительные границы для вероятности определенного ответа, т.е. разрабатываем метод интервального оценивания вероятности по выборочной доле и объему выборки.

Проверять однородность двух биномиальных выборок приходится, например, при сегментировании потребительского рынка. Для решения

этой задачи разработан метод проверки гипотезы о равенстве вероятностей, основанный на аналоге теоремы Муавра-Лапласа для двух выборок (любопытно, что этой теоремы нет в стандартных курсах теории вероятностей) [37].

9.3. Метод наименьших квадратов

Восстанавливать зависимость между переменными можно разными методами - графическим, наименьших модулей, минимаксным, наименьших квадратов. Из них наиболее используемым при решении задач экономики и управления является метод наименьших квадратов (сотни тысяч публикаций).

Начинаем с рассмотрения метода наименьших квадратов (МНК) для линейной прогностической функции (одна независимая и одна зависимая переменная). Минимизируя сумму квадратов отклонений, получаем точечные оценки параметров. Вводим восстановленные значения. Критерий правильности расчетов основан на равенстве сумм исходных значений зависимой переменной и восстановленных значений.

Поскольку распределения социально-экономических данных, как правило, не являются нормальными [38], принимаем непараметрическую вероятностно-статистическую модель порождения данных. Выводим формулы для оценок параметров. С помощью остаточной суммы квадратов оцениваем дисперсию погрешностей (остатков) в линейной прогностической модели. На основе точечного прогноза строим интервальный прогноз, указываем доверительные интервалы для зависимости (тренда) и индивидуальных значений. Отметим, что Центральная предельная теорема теории вероятностей – основа построения интервального прогноза.

Кратко рассматриваем обобщения базовой модели: МНК для сгруппированных данных, МНК для модели, линейной по параметрам. Обсуждаем оценивание коэффициентов многочлена и преобразования переменных с целью перехода к линейной модели. В случае нескольких независимых переменных (регрессоров) даем, в частности, подход к оцениванию параметров функции Кобба-Дугласа и аналогичных ей [39, 40].

9.4. Эконометрический анализ инфляции

Под инфляцией понимаем рост цен. Начинаем с краткой истории инфляции в СССР и России. Обсуждаем разброс цен (в зависимости от места совершения акта купли - продажи) и возможную точность определения «рыночной цены». Для измерения инфляции нужны инструменты экономиста и управленца - потребительские корзины. Даем определение индекса инфляции как отношение стоимостей потребительской корзины в два момента времени.

Изучаем свойства индекса инфляции. Начинаем с теоремы умножения, позволяющей рассчитывать индекс инфляции за два периода (один из них продолжает второй) как произведение индексов инфляции за периоды. Выясняем связь индекса инфляции "в разах" и индекса инфляции в процентах. Вводим средний индекс (темп) инфляции как среднее геометрическое индексов инфляции по отдельным периодам. Обсуждаем распространенные ошибки, связанные с индексом инфляции. Доказываем теорему сложения для индекса инфляции, позволяющую по групповым индексам инфляции рассчитывать индекс инфляции по объединенной корзине (вплоть до получения дефлятора ВВП).

Обсуждаем различные применения индексов инфляции, основанные на приведении экономических величин к сопоставимым ценам. Рассчитываем реальные проценты по вкладам в банки и кредитам в условиях инфляции. Рассматриваем метод Оршански для оценки прожиточного минимума на основе опыта проведения нами бюджетных обследований. Изучаем курс доллара в РФ в сопоставимых ценах. Проводим международные сопоставления на основе паритета покупательной способности [41, 42].

9.5. Методы экспертных оценок

Обсуждаем эконометрические методы сбора и анализа субъективной информации, полученной от экспертов. Приводим примеры процедур экспертного оценивания. Выделяем основные стадии проведения экспертного исследования с целью организовать работу управленцев, применяющих экспертные оценки.

Даем предварительную классификацию экспертиз. Описываем многообразные варианты организации экспертного исследования, различающиеся по цели (сбор информации или подготовка проекта решения для ЛПР - лица, принимающего решение), числу туров, порядку вовлечения экспертов, способу учета мнений (с весами или без весов), организации общения экспертов (анонимное, заочное, дистанционное, очное в соответствии с регламентом, дискуссия без ограничений). Рассматриваем положительные и отрицательные стороны рассматриваемых вариантов организации экспертного исследования [23, 43 - 45].

Пример экспертного исследования - анализ экспертных упорядочений. Демонстрируем три метода нахождения итогового мнения комиссии экспертов: методы средних арифметических и медиан рангов, построение согласующей ранжировки [46].

9.6. Теория измерений и средние величины

Вводим основные понятия общенаучной теории измерений. Обсуждаем определения, примеры, группы допустимых преобразований

для шкал наименований, порядковой, интервалов, отношений, разностей, абсолютной [1 - 3].

Базовым является требование устойчивости статистических выводов относительно допустимых преобразований шкал. Демонстрируем недопустимость использования среднего арифметического для усреднения данных, измеренных в порядковой шкале [47 - 51].

Перечисляем различные виды средних величин: среднее арифметическое, среднее геометрическое, среднее квадратическое, среднее гармоническое, их обобщение - средние степенные, а также структурные средние (медиана и другие члены вариационного ряда). Обсуждаем свойства средних величин. На примере расчета средней заработной платы работников условного предприятия демонстрируем целесообразность использования, кроме среднего арифметического, медианы и моды зарплат. Обсуждаем логарифмически нормальное приближение к распределению различных видов доходов, в соответствии с –которым среднее арифметическое всегда больше медианы, а та, в свою очередь, всегда больше моды.

Вводим самый общий вид средних - средние по Коши. Даем описание средних по Коши, результат сравнения которых устойчив в порядковой шкале. Это - только члены вариационного ряда, из которых выделяется медиана (при нечетном объеме выборки), левая и правая медианы (при четном объеме выборки) [52].

Вводим средние по Колмогорову (их частный случай - средние степенные). Даем характеризацию средних по Колмогорову, результат сравнения которых устойчив в шкалах интервалов (это только среднее арифметическое) и отношений (средние степенные и среднее геометрическое) [53].

Обсуждаем требование устойчивости выводов при применении статистических методов. Так, коэффициент линейной парной корреляции Пирсона предназначен для анализа данных, измеренных в шкале интервалов, а непараметрический коэффициент ранговой корреляции Спирмена - для анализа данных, измеренных в порядковых шкалах [37].

9.7. Введение в теорию риска

Под риском понимаем нежелательную возможность. Обсуждаем многообразие рисков (личные риски, производственные риски, коммерческие риски, финансовые риски, глобальные риски) [54].

Вводим характеристики рисков (вероятность рискового события, математическое ожидание, медиана, квантили, показатели разброса ущерба). Обсуждаем подходы к учету неопределенности и описанию рисков - вероятностно-статистический, с помощью нечетких множеств, на основе интервальной математики. Согласно этим подходам оценка рисков может проводиться с помощью вероятностно-статистических, нечетких,

интервальных моделей и методов. Обсуждаем постановки многокритериальных задач управления рисков, связанных с минимизацией математического ожидания и дисперсии случайного ущерба. Сведение двухкритериальных задач оптимизации к однокритериальным позволяет корректно решать задачи управления рисками [55, 56].

Строим иерархические системы рисков (частные риски - групповые риски - итоговый риск), в частности, групповые риски "Человек - Машина - Среда" в авиационной отрасли. Для оценки вероятности рискового события применяем аддитивно-мультипликативную модель (АММ) оценки риска. Рассматриваем общую формулировку и частные случаи, использование АММ для управления риском [57].

9.8. Основы статистики нечисловых данных

В качестве примера нечисловых данных рассматриваем бинарные отношения на конечном множестве – подмножества множества пар элементов этого множества. Используем их описание матрицами из 0 и 1. Обсуждаем базовые свойства бинарных отношений (рефлексивность, симметричность, транзитивность). Выделяем наиболее важные виды бинарных отношений (толерантности, разбиения (отношения эквивалентности), кластеризованные ранжировки. Вводим расстояние Кемени между бинарными отношениями и медиану Кемени, позволяющую найти среднее бинарное отношение для совокупности наблюдаемых бинарных отношений, полученных, например, при опросе экспертов [58].

Развиваем оптимизационный подход к определению средних величин в пространствах произвольной природы. Используя расстояния (показатели различия) в таких пространствах, вводим понятие эмпирического среднего. Для случайной величины со значениями в пространстве произвольной природы определяем теоретическое среднее. Рассматриваем примеры эмпирических и теоретических средних. Обсуждаем использование правила большинства при построении эмпирических средних в пространстве всех бинарных отношений и в пространстве подмножеств конечного множества [59].

Для выборки объектов нечисловой природы, состоявшей из независимых одинаково распределенных случайных величин со значениями в рассматриваемом пространстве, формулируем законы больших чисел в пространствах произвольной природы. Доказываем эти законы в частных случаях. Обсуждаем принципиальное значение законов больших чисел при анализе экспертных оценок, в частности, асимптотическое поведение эмпирических средних в случае монотонного распределения элементов выборки [60 - 63].

9.9. Непосредственный анализ статистических данных

При преподавании эконометрики полезно провести непосредственный анализ данных официальной экономической статистики. Обсуждаем динамику выпуска отдельных видов продукции (в натуральных единицах) и макроэкономических показателей РФ.

Роль государства в экономике оцениваем по доле расходной части бюджета в валовом внутреннем продукте. На основе данных Всемирного банка демонстрируем монотонное возрастание в течение XX в. роли государства в экономике в 11 экономически развитых странах в сравнении с ситуацией в России [64 - 66].

Даем представление об эконометрическом анализе демографических процессов. Обсуждаем демографические прогнозы в экономике и их значение для экономики и управления [67].

9.10. Контрольные работы и домашние задания первого семестра

Для контроля знаний предусмотрено 6 самостоятельно выполняемых контрольных работ:

1. Интервальное оценивание вероятностей (с доверительной вероятностью 0,95) и проверка гипотезы о равенстве вероятностей (на уровне значимости 0,05).
2. Метод наименьших квадратов.
3. Индекс инфляции.
4. Анализ экспертных упорядочений.
5. Аддитивно-мультипликативная модель оценки рисков.
6. Вычисление медианы Кемени.

Домашнее задание 1. Соберите информацию о максимально возможной цене (в руб.), которую потребители готовы заплатить за определенный товар или услугу (выбор товара или услуги осуществляется обучающимся самостоятельно или из списка, предлагаемого преподавателем). Опросите не менее 50 человек (не считая отказавшихся от ответа). Постройте выборочную функцию спроса. Найдите розничные цены, максимизирующие прибыль, для пяти различных значений оптовой цены.

Домашнее задание 2. Методом наименьших квадратов восстановите (теоретическую) функцию спроса, используя линейную аппроксимацию. Рассчитайте доверительные границы для функции спроса. Постройте на одном графике восстановленную и выборочную функции спроса. На основе восстановленных зависимостей найдите розничные цены, максимизирующие прибыль, для пяти различных значений оптовой цены, и сопоставьте с результатами оптимизации на основе таблицы выборочной функции спроса (домашнее задание 1). Проделайте аналогичные расчеты, используя степенную аппроксимацию. Ответ на вопрос: "Какая из двух

аппроксимаций позволяет более точно приблизить функцию спроса?" - дается на основе сравнения остаточных сумм квадратов.

9.11. Статистический контроль

Второй семестр начинаем с эконометрических методов управления качеством. Обсуждаем статистический приемочный контроль - выборочный контроль, основанный на теории вероятностей и математической статистике, его необходимость и эффективность. Вводим планы контроля по альтернативному признаку. Внимание уделяем, прежде всего, планам одноступенчатого контроля. Анализ плана контроля основан на оперативной характеристике - вероятности приемки партии в зависимости от входного уровня дефектности. Риски поставщика и потребителя и соответствующие им приемочный и браковочный уровни дефектности задают две выделенные точки на кривой оперативной характеристики. Расчеты для плана (n,0) упрощаются при применении разложения в ряд [68, 69].

Рассматриваем методы синтеза планов. Контроль с разбраковкой - процедура контроля, согласно которой забракованная партия проходит сплошной контроль. Находим средний выходной уровень дефектности (СВУД) как функцию входного уровня дефектности. Максимум СВУД достигается и называется его пределом (ПСВУД). Проводим расчет ПСВУД для плана (n,0) путем решения задачи оптимизации. Выбор плана контроля на основе ПСВУД осуществляется на основе простой формулы, вытекающей из применения второго замечательного предела математического анализа [70].

Синтез одноступенчатого плана контроля по заданным приемочным и браковочным уровням дефектности проводим на основе асимптотических соотношений, вытекающих из теоремы Муавра-Лапласа. Демонстрируем роль предельных теорем в теории статистического контроля [71, 72]. На примере управления качеством обсуждаем проблемы применения в цифровой экономике организационно-экономического моделирования и искусственного интеллекта [73].

9.12. Эконометрический анализ связанных выборок

На основе письма главного инженера химического комбината формулируем проблему обнаружения эффекта (проверки однородности) в связанных выборках. Рассматриваем три варианта обнаружения эффекта путем проверки соответствующих статистических гипотез.

Проверку гипотезы о том, что медиана разностей результатов измерений для двух приборов равна 0, проводим с помощью критерия знаков. Асимптотический метод проверки гипотезы строим на основе теоремы Муавра-Лапласа.

Проверку равенства 0 математического ожидания разностей результатов измерений для двух приборов проводим с помощью непараметрического критерия на основе отношения выборочного среднего к выборочному среднему квадратическому отклонению. Его асимптотическое распределение находим с помощью Центральной предельной теоремы теории вероятностей.

Гипотеза абсолютной однородности (отсутствия эффекта) эквивалентна гипотезе симметрии распределения относительно 0. Для её проверки применяем критерий типа омега-квадрат для проверки симметрии распределения [74]. Правило принятия решения строится на основе асимптотического распределения статистики типа омега-квадрат. Предлагаем табличный алгоритм для расчета значения рассматриваемой статистики [75]. Применяем модель анализа совпадений при расчете непараметрических ранговых статистик [76].

9.13. Основы теории нечетких множеств

Обсуждаем невозможность устранения погрешностей измерений и вычислений, сходство и различие математических, реальных и компьютерных чисел [77].

Парадокс Зенона "Куча" демонстрирует, что невозможность описания расплывчатых величин с помощью однозначно заданных чисел была выявлена еще в Древней Греции. Французский математики Эмиль Борель предложил использовать для описания размытых величин функцию принадлежности (1956). В 1965 г. Л.А. Заде ввел операции над функциями принадлежности и тем самым заложил основы теории нечетких множеств (fuzzy sets - переводят как нечеткие, размытые, расплывчатые, туманные, пушистые множества) [78].

Рассматриваем описание неопределенностей с помощью теории нечетких множеств. Доказываем формулы алгебры нечетких множеств, выявляем сходство и различие с обычной алгеброй множеств. Доказываем законы де Моргана в алгебре нечетких множеств. Рассматриваем треугольные нечеткие числа. Обсуждаем "удвоение математики" путем замены обычных множеств и чисел на нечеткие [79 - 82]..

Вводим понятие случайного множества как случайной величины в пространстве подмножеств (конечного) множества. Вводим распределения случайных множеств и вероятности накрытия. Подробно рассматриваем случай подмножеств конечного множества из трех элементов.

Описываем сведение теории нечетких множеств к теории случайных множеств. Для нечеткого множества с носителем из трех элементов строим случайное множество, для которого вероятности накрытия совпадают со значениями функции принадлежности исходного нечеткого множества [83].

9.14. Элементы статистики интервальных данных

Погрешности измерения описываем как интервальные данные. Вводим операции над интервальными числами [37].

Изучаем основную модель статистики интервальных данных. Вводим базовое понятие нотны - максимально возможного отклонения значения функции, вызванного интервальностью статистических данных. Выводим правила расчета асимптотической нотны (для малой абсолютной погрешности и малой относительной погрешности). [48, 77].

Формулируем основные результаты статистики интервальных данных, в том числе базовое понятие рационального объема выборки [84].

Рассчитываем асимптотическую нотну, рациональный объем выборки и доверительные интервалы при оценивании математического ожидания с помощью среднего арифметического и при оценивании дисперсии с помощью выборочной дисперсии [85].

Для управления инвестиционными проектами необходимо сравнение потоков платежей. Для этого используется чистая текущая стоимость NPV как характеристика финансового потока. Необходимо изучать устойчивость (чувствительность) выводов по отношению к отклонениям коэффициентов дисконтирования и величин платежей. Обсуждаем влияние интервальности дисконт-факторов на величину NPV. Разрабатываем и применяем алгоритм расчета погрешности NPV [86].

9.15. Основы теории классификации

Рассматриваем основные математические методы классификации [87]. Исходим из триады: построение классификаций (кластер-анализ, распознавание образов без учителя и другие синонимы) - анализ классификаций (в рамках статистики нечисловых данных) - использование классификаций (дискриминантный анализ, диагностика, распознавание образов с учителем).

Лемма Неймана-Пирсона дает оптимальный способ диагностики в случае двух классов, основанный на отношении плотностей распределения вероятностей, соответствующих этим классам. Описываем непараметрический дискриминантный анализ на основе непараметрических оценок плотности в пространствах произвольной природы. В качестве инструментов диагностики предлагаем различные варианты непараметрических оценок плотности в пространствах произвольной природы, прежде всего ядерные оценки [88 - 91].

Рассматриваем линейный дискриминантный анализ, в котором диагностика на два класса проводится с помощью «индексов» - линейных функций от координат. Обсуждаем характеристики качества алгоритмов диагностики. Демонстрируем невозможность использования такой характеристики, как «вероятность правильной классификации». Рекомендуем применять в качестве такой характеристики

«прогностическую силу». Указываем асимптотическое распределение и доверительные интервалы для прогностической силы. Даем способ статистической проверки возможности использования прогностической силы на основе проверки гипотезы о совпадении ее значений для двух критических порогов алгоритма диагностики [92].

Обсуждаем, чем схожи и чем различаются задачи группировки и кластер-анализа. Вводим агломеративные иерархические алгоритмы ближнего соседа, дальнего соседа и средней связи. Построение дендрограмм для таких алгоритмов. На примере метода *k*-средних обсуждается проблема остановки алгоритма [93].

9.16. Элементы теории рейтингов

При обсуждении элементов теории и применений рейтингов рассматриваем рейтинги, интегральные показатели, обобщенные показатели (используем эти термины как синонимы) [67, 94].

Бинарные рейтинги сводятся к задаче диагностики на два класса, для оценки различающей возможности рейтингов используем прогностическую силу [92].

Рассматриваем построение интегрального показателя в задачах принятия решений. На примере деловой игры "Таня Смирнова выбирает место работы" обсуждаем экспертные методы построения системы факторов (в том числе иерархической – единичные, групповые и обобщенный показатели), системы весов факторов, оценки объектов экспертизы по факторам [93, 95, 96].

9.17. Эконометрика как научная дисциплина

В конце курса естественно обсудить эконометрику в целом, в то время как ранее рассматривались лишь отдельные вопросы этой науки.

Кратко рассказываем об истории эконометрики (от переписи военнообязанных во времена Моисея до настоящего времени) [97, 98].

Обсуждаем структуру статистической науки (математическая статистика – прикладная статистика – статистические методы в предметных областях). Эконометрика - это статистические методы в конкретной предметной области - в экономике и управлении. Выделяем специфические черты эконометрики по сравнению с другими предметными областями (неотрицательность рассматриваемых величин, которая дает еще один довод в пользу использования непараметрических методов; большое значение методов сбора и анализа субъективных экспертных методов и др.) [37].

Выделяем четыре этапа развития теории статистики (описательная, параметрическая, непараметрическая, нечисловая), указываем характерные для того или иного этапа методы анализа данных и временные промежутки. По видам данных статистика делится на четыре области

(статистика чисел, многомерный статистический анализ, временные ряды, статистика нечисловых данных). В статистике выделяют три основные задачи (описание данных, оценивание, проверка гипотез). В настоящее время наблюдаем пять точек роста статистической науки: непараметрика, информационные технологии (бутстреп), устойчивость, статистика интервальных данных, нечисловая статистика [37]..

Отечественная научная школа в области организационно-экономического моделирования, эконометрики и статистики [5 - 7] основана на новой парадигме математических методов исследования [10 - 16], другими словами, на современной парадигме эконометрики XXI в., в отличие от распространенных учебников [8, 9], подготовленных в духе старой парадигмы середины XX в.

9.18. Контрольные работы и домашние задания второго семестра

Для контроля знаний предусмотрено 6 самостоятельно выполняемых контрольных работ:

1. Статистический приемочный контроль - анализ и синтез планов.
2. Проверка однородности связанных выборок.
3. Нечеткость и интервальность.
4. Расчет погрешности чистой текущей стоимости *NPV*.
5. Кластер-анализ с помощью агломеративного иерархического алгоритма ближнего соседа.
6. Построение интегрального показателя.

Домашние задания проводится по теме «Индекс инфляции и метод наименьших квадратов».

Домашнее задание 1 состоит в сборе данных о ценах на продуктовые товары, входящие в потребительскую корзину Института высоких статистических технологий и эконометрики МГТУ им. Н.Э. Баумана [2 - 4]. Для этого необходимо выбрать и зафиксировать места сбора информации о ценах, конкретные объекты наблюдения (марки тех конкретных товаров, мониторинг цен на которые проводится). Затем проводится сбор данных по ценам за пять моментов времени, попадающие в заданные интервалы (примерно 1 раз в 2 недели).

Домашнее задание 2 посвящено анализу собранных данных. Вначале необходимо провести расчет пяти наборов индексов инфляции по 10 товарным группам, выделенным в используемой потребительской корзине. Затем следует рассчитать пять общих индексов инфляции (по всей корзине) двумя способами: на основе теоремы сложения и как отношение стоимостей потребительской корзины, сравнить полученные результаты.

После получения набора индексов инфляции необходимо по первым четырем индексам инфляции спрогнозировать значение пятого индекса

методом наименьших квадратов. Скажем подробнее. По первым четырем наборам индексов инфляции по 10 товарным группам и четырем общим индексам инфляции методом наименьших квадратов необходимо рассчитать точечные и интервальные прогнозы (т.е. доверительные границы) для зависимости индекса инфляции от времени (т.е. для тренда) и для индивидуальных значений на пятый момент времени, предварительно выбрав модель динамики цен (обычно используют линейную модель тренда).

Рассчитав индексы инфляции по товарным группам и общие для различных начальных моментов времени, можно проверить выполнение теоремы умножения.

Подводя итоги выполнения домашнего задания, следует сравнить прогнозы с реальными индексами инфляции (по товарным группам и общим), сделать выводы о динамике индексов инфляции и о возможности их прогнозирования.

9.19. Заключительные замечания

Как показано выше, основные составляющие современной эконометрики представлены в разработанном нами учебном курсе [2 - 4]. Целесообразно именно его преподавать во многих университетах и вузах другого профиля, оставив в прошлом устаревшие учебники [8, 9] и им аналогичные. В таких учебниках из всех базовых тем современной эконометрики рассматривается лишь одна - метод наименьших квадратов (конечно, гораздо подробнее, чем в нашем курсе и чем необходимо специалистам в области экономики и управления).

Как показано в [100], современная эконометрика - неотъемлемая составляющая научного обеспечения искусственного интеллекта и цифровой экономики. Разнообразные применения эконометрических методов при решении практических задач экономики и управления рассмотрены в сотнях тысяч исследований, из которых в качестве примеров укажем на работы [101, 102].

В современных условиях эконометрика как научная, практическая и учебная дисциплина становится всё более востребованной. Констатируем справедливость этого утверждения и для системной нечеткой интервальной математики как основы современной эконометрики.

ГЛАВА 10. СИСТЕМНАЯ НЕЧЕТКАЯ ИНТЕРВАЛЬНАЯ МАТЕМАТИКА - ОСНОВА МАТЕМАТИКИ XXI ВЕКА

Уже более полувека известно [1], что вклад исследователя в фундаментальную науку измеряется числом цитирований его работ в

научных публикациях. Согласно Российскому индексу научного цитирования (РИНЦ) автор настоящей главы - второй по цитированию среди ныне живущих отечественных математиков (и десятый по цитированию среди экономистов). Признание коллег накладывает ответственность и обосновывает желание высказаться по поводу состояния и перспектив математических исследований. Исходная точка для рассуждений - работы по математических и инструментальным методам экономики (научная специальность 08.00.13). Начнем с краткой формулировки основных результатов.

Определения математики как науки менялись со временем. В XIX в. определяли математику как науку о числах и фигурах (телах). В XXI в. математика — это наука о структурах, порядке и отношениях, или короче: "математика - наука о формальных структурах". Следовательно, ее нельзя относить к естественным наукам.

Математика изучает мысленные конструкции. Идеал - построение и развитие аксиоматических теорий (в смысле Д. Гильберта). В практике математических рассуждений аксиоматические теории - это, как правило, недостижимый идеал. Практически никто из математиков не формулирует в своих работах аксиомы и правила вывода.

10.1. О структуре математики как области деятельности

Со стороны видны два направления деятельности математиков. Исследования представителей первого из них нацелены на построение и изучение моделей реального мира, на получение научных результатов, которые - прямо или опосредованно - позволяют решать практические задачи. Представители второго направления занимаются решением конкретных внутриматематических задач, как правило, трудных. Примерами таких уже решенных задач являются "великая теорема Ферма", задача пяти красок и т.п. Именно представители второго направления готовят новых математиков, руководят профессиональными объединениями. В результате такого разделения обязанностей первое направление оказывается ущемленным.

С точки зрения представителей первого направления наиболее важные области математики - это математический анализ, идущий от Ньютона и Лейбница, алгебра (линейная, высшая и др.) и геометрия (многомерная, начертательная, дифференциальная, топология и др.). Для решения прикладных задач в XX в. наиболее важными оказались теория вероятностей и математическая статистика, теория оптимизации, дифференциальные и разностные уравнения. Начиная со второй половины XX в. появились новые области математики - статистика нечисловых данных, теория нечетких множеств, автоматизированный системно-когнитивный анализ, интервальная математика. Объединяющую их системную нечеткую интервальную математику рассматриваем как основу

математики XXI века. Основная часть областей математики, разработанных представителями второго направления, в применении к решению прикладных задач оказалась, увы, бесплодной.

Необходимо различать математические, прагматические и компьютерные числа. Разработан ряд подходов к моделированию связей математических и прагматических чисел - на основе теории группировки, интервального анализа, нечетких множеств, автоматизированного системно-когнитивного анализа.

По нашей оценке, столбовая дорога" будущей математики - это системная нечеткая интервальная математика.

В конце главы рассказано о многообразии литературных источников.

10.2. Определения математики

В XIX в. определяли математику как науку о числах и фигурах (телах). Например, Ф. Энгельс писал:

«Чистая математика имеет своим объектом пространственные формы и количественные отношения действительного мира, стало быть — весьма реальный материал. Тот факт, что этот материал принимает чрезвычайно абстрактную форму, может лишь слабо затушевать его происхождение из внешнего мира. Но чтобы быть в состоянии исследовать эти формы и отношения в чистом виде, необходимо совершенно отделить их от их содержания, оставить это последнее в стороне как нечто безразличное» [2].

По традиции это определение довольно часто используется и в настоящее время. На нем основано преподавание в средней школе.

К настоящему времени определение математики изменилось. Пишут примерно так: "Математика — наука о структурах, порядке и отношениях". При этом обращают внимание на преемственность: "Математика исторически сложилась на основе операций подсчёта, измерения и описания формы объектов", т.е. на основе рассмотрения чисел и фигур.

Весьма важным является происхождение математических объектов. Они "создаются путём идеализации свойств реальных объектов и процессов или других математических объектов и записи этих свойств на формальном языке".

Можно сказать короче: "Математика - наука о формальных структурах", поскольку "порядок" и "отношения" - это также структуры, но более специального вида. Структуры, которые изучают математики, получены при моделировании (упрощении, идеализации) реальных объектов, процессов, структур, либо же построены на основе других математических структур. Но затем их начинают изучать математическими средствами совершенно независимо от реальности. В результате можно получить новое знание о реальности. А можно - уйти от реальности внутрь формальной структуры.

Поскольку математика - наука о формальных структурах, то ее нельзя относить к естественным наукам, как иногда делают. В известном термине "физико-математические науки" первая составляющая - "физика", т.е. наука о реальном мире. А вторая - "математика" - относится мысленным объектам - формальным структурам. Кандидат или доктор физико-математических наук может всю жизнь заниматься моделированием социально-экономических систем и знать о физике только то немногое, что осталось в голове от предмета "физика" в средней школе. Физика - лишь одна из областей применения математики. От термина "физико-математические науки" необходимо отказаться. Иначе надо вводить, например, медико-математические науки или экономико-математические науки. Последние, впрочем, существуют внутри экономических наук как "экономико-математические методы и модели", или "математические и инструментальные методы экономики" (научная специальность 08.00.13).

10.3. Аксиоматические теории

Математика изучает мысленные конструкции. Идеал - построение и развитие аксиоматических теорий (в смысле Д. Гильберта). Для построения математической теории вводится список изучаемых объектов (например, точка, прямая, ...), формулируются некоторые утверждения, принимаемые без доказательства (аксиомы), а все остальные утверждения выводятся из них чисто логическим путем по фиксированным правилам. При таком подходе из математической теории изгоняются интуиция, наглядные представления из реального мира (геометрические, физические и т.п.), индуктивные рассуждения и т.д. Таким образом, аксиоматический метод позволяет выяснить, что именно вытекает из аксиом, очищает рассуждения от осознанных или неосознанных следствий из свойств реального мира.

В практике математических рассуждений аксиоматические теории - это, как правило, недостижимый идеал, к которому, по общему мнению, надо стремиться. Строгость доказательств теорем проверяется практикой математических рассуждений в конкретный момент времени и в конкретном месте (местах). По мере выявления противоречий вносятся изменения в математические традиции. Однако при проведении конкретных доказательств специалисту обычно однозначно ясно, какие рассуждения являются строгими, а какие нет.

В реальной работе математика крайне редко реализуется идеал построения аксиоматической теории. Обычно рассуждения проводятся на общепринятом в конкретное время и конкретной области исследования уровне строгости. И между специалистами, как правило, не возникает споров по поводу того, доказано то или иное утверждение или нет. Предполагается, что можно довести рассуждение до уровня

аксиоматической теории. Но практически никто из реально работающих математиков это не делает. Любопытно, что такая позиция не приводит к проблемам ни теоретической работе, ни в практической деятельности.

10.4. Два направления в математике

Что основное в деятельности математиков? Они ведут научные исследования, в результате которых доказывают новые теоремы. Конечно, они занимаются и другими видами деятельности - общаются с другими математиками и представителями внематематического мира (инженерами, экономистами, управленцами и др.), преподают, занимаются административной работой, пишут статьи и книги и т.п. Но основное в их деятельности - доказательство новых теорем. Именно этим они отличаются от представителей других видов деятельности.

Видны два направления деятельности математиков. Исследования представителей первого из них нацелены на построение и изучение моделей реального мира, на получение научных результатов, которые - прямо или опосредованно - позволяют решать практические задачи. Традиционная схема исследования такова. Для той или иной области реального мира формируется математическая модель. Затем происходит мысленный отрыв от реального мира, переход внутрь математической структуры. Математическими средствами проводится исследования. На третьем этапе полученные математические результаты "спускаются" в реальный мир, интерпретируются в терминах соответствующей прикладной области.

Представители второго направления не думают о проблемах реального мира. Они занимаются решением конкретных трудных задач. Примерами являются "великая теорема Ферма", задача пяти красок и т.п. В XXI в. известна гипотеза Пуанкаре, которую доказал Г. Перельман.

Обычно, но не всегда, не видно пользы от работ представителей второго направления для внематематических областей. После решения очередной трудной задачи про неё забывают, переходят к решению следующих.

Между двумя направлениями есть промежуточная область. При изучении моделей реального мира возникают новые математические задачи. Те, кто ими занимается, могут работать внутри математики, не обращаясь к рассмотрению проблем внешнего мира, и в этом смысле действовать аналогично представителям второго направления.

Представители первого направления часто работают вместе с учеными других областей науки и техники, обычно в различных прикладных научных структурах. Представителям второго направления целесообразно отгородиться от внешнего мира. Они сосредотачиваются в математических институтах и на профильных факультетах. К сожалению, именно они готовят новых математиков, руководят профессиональными

объединениями. В результате первое направление оказывается ущемленным по сравнению со вторым, как в новых кадрах, так и в признании научных результатов представителей первого направления в профессиональных объединениях математиков, например, в отделении математики РАН. Сложившаяся ситуация, очевидно, тормозит развитие математики, отрывает начинающих исследователей от направлений, нацеленных на участие в практической деятельности.

10.5. Области математики

С точки зрения представителей первого направления наиболее важные области математики - это математический анализ, идущий от Ньютона и Лейбница, алгебра (линейная, высшая и др.) и геометрия (многомерная, начертательная, дифференциальная, топология и др.). Для решения прикладных задач в XX в. наиболее важными оказались теория вероятностей и математическая статистика, теория оптимизации, дифференциальные и разностные уравнения. Начиная со второй половины XX в. появились новые области математики - статистика нечисловых данных, теория нечетких множеств, автоматизированный системно-когнитивный анализ, интервальная математика. Объединяющую их системную нечеткую интервальную математику [3] рассматриваем как основу математики XXI века.

Основная часть областей математики, разработанных представителями второго направления, в применении к решению прикладных задач оказалась, увы, бесплодной. Не для этого они разрабатывались. Опыт двадцати лет XXI в. подтверждает сказанное. Печально глядеть на длинные ряды математических журналов на библиотечных полках, понимая, что пользы для человечества от опубликованных в них теорем нет и почти наверняка не будет.

Математиков второго направления можно сравнить с шахматистами. Они играют, сражаются за первые места, их партии зачастую можно рассматривать как произведения искусства. Но целесообразно ли давать им государственную поддержку, открывать факультеты шахмат в ведущих университетах? В настоящее время ответ общества - нет, нецелесообразно. Математиков второго направления также вряд ли нужно готовить в государственных вузах и размещать в государственных научно-исследовательских организациях.

Некогда популярные области математики хиреют. Примером является элементарная геометрия, изучающая точки, прямые, треугольники, окружности. Исследования в этой области начались в Древней Греции, Сводка полученных результатов дана в знаменитых "Началах" Евклида. Однако основной массив теорем был получен учителями математики в гимназиях XIX в. В следующем веке поток новых результатов иссяк. Тем не менее до сих пор в средней школе элементарную

геометрию изучают в большом объеме, включают в программы различных экзаменов. На эту устаревшую область математики излишне тратят силы преподаватели и учащиеся.

За полвека автору этой главы никто не смог привести примеры практических задач, в которых была бы полезна теорема о том, что три перпендикуляра, восстановленные в серединах сторон треугольника, пересекаются в одной точке. Утверждение, конечно, красивое, но бесполезное для практики. Нужно ли обучать доказательству этой теоремы? Не лучше ли рассмотреть математические теории, полезные для практики?

Аналогичное увядание наблюдаем для параметрической математической статистики. Но есть и нюансы. Преподавание этой области ущербно. До сих пор в учебных курсах рассказывают об оценках максимального правдоподобия, хотя продемонстрированы преимущества перед ними одношаговых оценок.

Сказанное обосновывает необходимость рассуждений о направлениях будущего развития математики с целью выделения наиболее перспективных. Исходим из проанализированных нами потребностей научной специальности 08.00.13 "математические и инструментальные методы в экономике".

10.6. Математические, прагматические и компьютерные числа

Для описания фактов реальности часто используют числа. В математике выделяют натуральные, рациональные, действительные (вещественные) числа. Обсудим некоторые их свойства, оставив без внимания комплексные числа, кватернионы, трансфинитные числа.

Еще в Древней Греции была установлено, что натуральных чисел бесконечно много. С теоретической точки зрения ясно, что дробей и вещественных чисел - бесконечно много. Но это в математике. А на практике мы пользуемся всего лишь такими числами, в которых значащих десятичных цифр - конечное число. Более того, обычно значащих десятичных цифр совсем немного - пять, семь, не более десяти. Таких чисел - конечное число, хотя и довольно большое - миллионы.

Таким образом, математических чисел (имеющихся в математических теоретических системах) - бесконечно много, а прагматических (которые мы применяем в практических расчетах) - конечное число. Этот разрыв между математикой и практикой имеет разнообразные последствия.

Прагматические числа записываются конечным (не более 10) набором значащих цифр не только из-за сложности записи, но и потому, что ограничена точность измерений (наблюдений, испытаний, анализов, опытов, обследований).

Нельзя записать с помощью конечного набора цифр постоянно используемые в различных разделах математики трансцендентные числа (отношение длины окружности к диаметру, основание натуральных логарифмов и др.). Нельзя записать и иррациональные числа, например, длину диагонали квадрата е единичным основанием.

Числа, используемые в компьютерных расчетах, также отличаются от математических. Компьютерные числа примыкают к прагматическим, хотя могут использовать большее число бинарных разрядов. Принципиально важным является наличие "машинного нуля" - положительной границы, такого числа, что все положительные результаты расчетов, меньшие машинного нуля, считаются равными 0. Как следствие, бесконечный ряд, слагаемые которого - обратные величины натуральных чисел, в математике имеет бесконечную сумму, а при вычислении на компьютере - конечную, поскольку все слагаемые, начинающиеся с некоторого, обнуляются.

Как преодолеть разрыв? Необходима разработка новой математической теории. Назовем ее теорией прагматических чисел.

10.7. Моделирование связей математических и прагматических чисел

Есть два подхода. Во-первых, прагматические числа можно моделировать дискретными математическими моделями. В частности, использовать таблицы сопряженности, теорию информации, теорию систем, системно-когнитивный анализ. При таком подходе считается, что исходные данные взяты из заданного конечного множества. В рамках рассматриваемого подхода разработано большое число методов анализа данных.

Во-вторых, можно моделировать связи между прагматическими и математическими числами с целью использовать аппарат непрерывных и дифференцируемых величин. В рамках второго подхода рассмотрим ряд моделей, в которых прагматические числа рассматриваются как приближенные значения математических.

В модели группировки значения дискретной переменной порождаются в результате группировки значений непрерывной переменной. Например, фиксируем температуру 15^0, если значения непрерывной переменной больше $14{,}5^0$ и не превосходит $15{,}5^0$. Здесь границы между интервалами группировки заранее заданы и не зависят от значения непрерывной переменной. В математической статистике такие модели рассматриваются с XIX в. (поправки Шеппарда).

В моделях интервального анализа, прежде всего в статистике интервальных данных, значения прагматического и математического чисел различаются не более чем на малое заданное число. При этом границы между интервалами группировки зависят от значения непрерывной

переменной. Статистика интервальных данных развивается с 1980-х годов. Она принципиально отличается от математической статистики первой половины XX в. В частности, в статистике интервальных данных отсутствуют состоятельные статистические оценки, введено понятие рационального объема выборки, превышать который нерационально. Связано это с тем, максимально возможное расхождение значений статистик, рассчитанных по прагматическим и математическим числам (т.н. нотна - одно из основных понятий статистики интервальных данных) не стремится к 0 при росте объема выборки.

Третий тип моделей строится на основе теории нечетких множеств, математический аппарат которой активно развивается с 1960-х годов. Расхождения между функциями от прагматических и математических чисел изучаются как нечеткие объекты.

Иногда утверждают, что теория вероятностей и теория нечетких множеств - две разные области математики. Это не так. Еще в 1970-х годах установлено, что теория нечетких множеств в некотором смысле сводится к теории случайных множеств и тем самым к теории вероятностей. Однако этот фундаментальный факт мало влияет на алгоритмы решений практических задач - они остаются различными для применений теории нечеткости и для применений вероятностно-статистических моделей и методов.

10.8. Системная нечеткая интервальная математика в математике XXI века

Моделированию связей математических и прагматических чисел посвящена монография "Системная нечеткая интервальная математика" 2014 г., подготовленная нами совместно с проф. Е.В. Луценко [3]. Название монографии констатирует выделение основного (на современном этапе) стержня математики как развивающейся науки. В настоящей главе мы рассматриваем системную нечеткую интервальную математику как основу математики XXI века. Она на новом уровне и в новом направлении развивает основные концепции математики предыдущего тысячелетия. Нечеткие и интервальные числа - основа системной нечеткой интервальной математики. Обсудим такое базовое понятие для математики XXI века, как система.

В переводе с древнегреческого система - это некое целое, составленное из частей; соединение. Другими словами, система — это множество элементов, находящихся в отношениях и связях друг с другом, которое образует определённую целостность, единство. Термин «система» целесообразно использовать в тех случаях, когда нужно подчеркнуть, что *что-то* является большим, сложным, не полностью сразу понятным, при этом целым, единым. В отличие от понятий «множество», «совокупность» понятие системы подчёркивает упорядоченность, целостность, наличие

закономерностей построения, функционирования и развития. Свойства системы не сводятся к сумме свойств ее элементов. Используют специальный термин *эмерджентность* для обозначения появления у системы свойств, не присущих её элементам в отдельности; несводимость свойств системы к сумме свойств её компонентов.

Такие термины, как анализ систем, системная математика, системный анализ, в частности, системный анализ данных, имеют практически совпадающее содержание, их, по нашему мнению, можно рассматривать как синонимы. Близки к ним системотехника и системное проектирование. Часть этого направления - теория принятия решений.

Современный этап развития этого научного направления - это автоматизированный системно-когнитивный анализ. Приведем часть аннотации к базовой публикации по этой стержневой области системного анализа.

"Системный анализ представляет собой современный метод научного познания, общепризнанный метод решения проблем. Однако возможности практического применения системного анализа ограничиваются отсутствием программного инструментария, обеспечивающего его автоматизацию. Существуют разнородные программные системы, автоматизирующие отельные этапы или функции системного анализа в различных конкретных предметных областях.

Автоматизированный системно-когнитивный анализ (АСК-анализ) представляет собой системный анализ, структурированный по базовым когнитивным операциям, благодаря чему удалось разработать для него математическую модель, методику численных расчетов (структуры данных и алгоритмы их обработки), а также реализующую их программную систему – систему «Эйдос». Система «Эйдос» разработана в постановке, не зависящей от предметной области, и имеет ряд программных интерфейсов с внешними данными различных типов. АСК-анализ может быть применен как инструмент, многократно усиливающий возможности естественного интеллекта во всех областях, где используется естественный интеллект. АСК-анализ был успешно применен для решения задач идентификации, прогнозирования, принятия решений и исследования моделируемого объекта путем исследования его модели во многих предметных областях, в частности в экономике, технике, социологии, педагогике, психологии, медицине, экологии, ампелографии, геофизике, энтомологии, криминалистике и др." [4].

Отметим, что методы анализа данных могут быть развиты на основе теории информации. Это утверждение продемонстрировано в подходе Кульбака [5], который в свое время высоко оценил А.Н. Колмогоров.

10.9. Некоторые распространенные заблуждения

Верно ли, что любая математическая теория строится на определенной аксиоматике? Достаточно просмотреть несколько математических работ, чтобы убедиться в том, что практически никто из их авторов не формулирует аксиомы и правила вывода. Но при этом неявно предполагается, что привычные математические теории - например, дифференциальное и интегральное исчисления или предельные теоремы теории вероятностей, - не содержат противоречий. Т.е. аксиоматика - где-то вдалеке. В частности, из-за того, что абсолютно строгие рассуждения являются крайне длинными. Например, для строгого введения понятия "один" Н. Бурбаки понадобился целый том "Элементов математики".

Некоторые математические теории по традиции исходят из нерациональных предпосылок. Например, базовым понятием теории вероятностей является понятие вероятностного пространства, состоящего из пространства элементарных событий, сигма-алгебры измеримых множеств (событий) в нём и вероятностной меры, определенной на элементах этой сигма-алгебры. При использовании этого понятия приходится держать в уме возможность появления неизмеримых множеств на различных этапах рассуждений. Вместе с тем в случае, когда пространство элементарных событий состоит из конечного числа элементов, можно считать все его подмножества измеримыми. Как следствие, можно избавиться от опасности появления неизмеримых множеств. Так следует ли заниматься вопросами измеримости? На наш взгляд, от них можно избавиться на этапе выбора изучаемой модели, приняв, что используются конечные множества. Переходить к бесконечным множествам имеет смысл только тогда, когда такой переход облегчает рассуждения (как при переходе от сумм к интегралам). Примерно так говорил автору настоящей главы А.Н. Колмогоров полвека назад.

Обсудим соотношение схем и теорем. Теорема отличается от схемы рассуждений добавлением условий, при которых теорема верна. Примером схемы является центральное утверждение теории вероятностей: распределение центрированной и нормированной суммы независимых случайных величин приближается стандартным нормальным распределением при увеличении числа слагаемых. Добавляя те или иные условия, получаем различные варианты Центральной Предельной Теоремы (ЦПТ) теории вероятностей. На протяжении нескольких веков условия справедливости ЦПТ совершенствовались [6]. Надо подчеркнуть важность формирований перспективных схем рассуждений, указывающих путь дальнейшим исследованиям по выявлению условий, при которых теорема верна. Формирование перспективной схемы рассуждений - не менее важное достижение, чем доказательство теоремы при тех или иных условиях.

Вместо метрических пространств естественно применять пространства с естественными показателями близости, поскольку неравенство треугольника, как правило, не является необходимым для проведения рассуждений.

10.10. Организационные вопросы развития математики

Мы показали, что "столбовая дорога" будущей математики - это системная нечеткая интервальная математика. Она активно развивается многими исследователями.

Однако нельзя не сказать о том, что новое развивается в борьбе со старым. Традиционный подход оторванной от потребностей практики чистой математики в настоящее время господствует в учебных заведениях и немногочисленных научных учреждениях. Специалисты, занимающиеся перспективными исследованиями, обычно выдавливаются из окружающей их инертной среды и переходят в организации практической направленности, в которых реализуют свои идеи. На примере элементарной геометрии (геометрии прямых и окружностей) мы видим, как традиция поддерживает отжившие области математики. Нельзя ожидать быстрого отмирания устаревших отраслей математики, поскольку за них будут держаться их адепты, неспособные перестроиться. Через несколько десятилетий всё будет ясно, но в течение этого времени необходимо действовать в новых направлениях. Надо продолжать активно развивать центральную область математики XXI в. - системную нечеткую интервальную математику.

10.11. Кратко о многообразии литературных источников

По рассмотренным выше вопросам опубликовано довольно много статей и книг. Исходя из нужд читателей, укажем некоторые из них.

Тематика статистики нечисловых данных и статистики интервальных данных раскрыта в монографиях [7 - 9]. Современное состояние этих научных, практических и учебных дисциплин отражено в статьях [10] и [11, 12] соответственно. Проблемам упомянутой выше теории принятия решений посвящены монографии [8, 13, 14]. Использование современных математических методов при решении различных прикладных задач рассмотрено в монографиях [15 - 19]. Речь идет о перспективных математических и инструментальных методах контроллинга, организационно-экономическом, математическом и программном обеспечении контроллинга, инноваций и менеджмента, современным подходам в наукометрии, современной цифровой экономике, высоких статистических технологиях и системно-когнитивном моделировании в экологии. Многие из более чем 140 статей, опубликованных нами в "Научном журнале КубГАУ", посвящены тематике, обобщенной в настоящей главе.

ЧАСТЬ 2-Я. АВТОМАТИЗИРОВАННЫЙ СИСТЕМНО-КОГНИТИВНЫЙ АНАЛИЗ КАК МЕТОД ПРЕОБРАЗОВАНИЯ ДАННЫХ В ИНФОРМАЦИЮ, А ЕЕ В ЗНАНИЯ И ПРИМЕНЕНИЯ ЭТИХ ЗНАНИЙ ДЛЯ РЕШЕНИЯ ЗАДАЧ В РАЗЛИЧНЫХ ПРЕДМЕТНЫХ ОБЛАСТЯХ

ГЛАВА 11. ПОНЯТИЯ ДАННЫХ, ИНФОРМАЦИИ И ЗНАНИЙ, СХОДСТВО И РАЗЛИЧИЯ МЕЖДУ НИМИ

11.1. Данные, подходы к определению

Традиционное определение понятия данных: данные – это информация, записанная на носителях в определённой системе кодирования или на определенном языке.

Продемонстрируем, что традиционное определение данных является ложным и абсурдным с логической точки зрения, применяя логический метод «Ложного основания». Это метод говорит о том, что если логическое следствие из некоторых исходных положений, является ложным, то и сами эти исходные положения также являются ложными.

Для этого попробуем дать определение понятия информации, основываясь на традиционном определении понятия данных и используя традиционный подход к научным определениям (дефинициям) через подведение определяемого понятия или термина под более общее понятие и выделение одного или нескольких специфических признаков.

При этом будем считать, как это по сути принято в традиционном определении (как мы это видели выше), что информация – это более общее понятие, чем данные, а специфическим признаком данных является то, что они записаны на носителе на определенном языке (это и есть ложное основание).

Эта попытка аналогична попытке дать определение более общего понятия «животное» на основе частного понятия «млекопитающее».

Исходное определение частного понятия: млекопитающее – это животное (более общее понятие), выкармливающее своих детенышей молоком (специфический признак).

Животное – это такое млекопитающее, которое:

– выкармливает своих детенышей молоком;

– или выкармливает своих детенышей не молоком.

По сути это означает, что животные это млекопитающие, но не только млекопитающие.

Информация – это такие данные, которые:

записаны на носителе на определенном языке,

или записаны на носителе не на определенном языке;

или не записаны на носителе на определенном языке,

или не записаны на носителе не на определенном языке.

Таким образом на основе традиционного определения понятия «данные» ***мы приходим к явно абсурдному результату***, т.к. информацию или данные, не записанные на каком-либо носителе на каком-либо языке или системе кодирования невозможно представить даже теоретически. *Это и означает, что исходное традиционное определение данных, однозначным логическим следствием из которого является этот абсурдный результат, является таким же абсурдным, как и следствие из него, т.е. является ложным, неверным, что и т.д.*

Следовательно, традиционное определение понятия данных является некорректным.

Но как же тогда корректно определить это понятие?

Однако, сделать это очень не просто по ряду причин:

Во-первых, по-видимому, *понятие данных относится к числу наиболее общих понятий, выработанных человечеством*. Это ясно из того, что о чем бы мы не рассуждали или не рассуждали, о каких бы объектах, процессах и явлениях внешнего и внутреннего мира, о природе обществе и человеке, мы все равно можем делать это только основываясь на каких-то данных об объекте рассуждения или познания. *Поэтому возникают принципиально неразрешимая проблема поиска более общего понятия, чем понятие данных.* Аналогичная неразрешимая проблема возникает при попытке определить традиционным путем (подведением под более общее понятие и выделение специфических признаков) другие предельно общие понятия, такие как бытие и небытие, материя и сознание, Бог, Вселенная и т.п. Впрочем, подобные понятия можно пересчитать по пальцам одной руки.

Во-вторых, но даже если бы такое более общее понятие, чем понятие данных, удалось найти, все равно возникла бы проблема выделения таких специфических признаков, которые в этом более общем понятии позволяют выделить подмножество, соответствующее определяемому понятию: «Данные».

Поэтому в нашем распоряжении остается один вариант: описывать различные конкретные примеры данных всеми возможными признаками, а затем на основе этих описаний сформировать обобщенный образ данных, *включив* в него, признаки, вероятность наблюдения которых в данных намного выше, или на много ниже, чем в других понятиях, и *исключив* из него признаки, вероятность встречи которых в данных мало отличается от

средней вероятности их встречи по всем формируемым обобщенным понятиям.

Такой ход рассуждений называется абдукция. Например: «Сократ смертен, Сократ человек, следовательно, человек смертен». Таким образом, мы что-то узнали об обобщенной категории «Человек». Понятно, что такой ход рассуждений является *правдоподобным*, но не гарантирует истинного результата, хотя в данном примере результат и получился истинный. В отличие от этого рассуждение от общего к частному всегда дает *истинный* результат, например: «Люди смертны, Сократ – человек, следовательно, Сократ смертен». Отметим, что степень правдоподобности результатов абдукции возрастает, при увеличении числа примеров объектов, относящихся к различным обобщенным образам, и увеличении числа признаков, описывающих эти примеры.

Этот подход позволяет дать описательное или операциональное определение данных.

Однако этот подход предполагает рассмотрение не только понятия «Данные», но и других связанных с ним понятий, таких как «Информация» и «Знания», тем более что многие, вообще не видят между ними особых различий. Поэтому даже операциональное определение данных можно дать только в сопоставлении его с понятием информация, что мы и сделаем в следующем разделе.

Наиболее фундаментальным свойством данных является то, что они констатируют <u>различие</u> чего-либо.

11.2. Информация и данные

<u>Информация</u> определяется как осмысленные данные.

Данные – это более общее понятие, чем информация. Информация тоже является данными, но не каким угодно, а только осмысленными. ***Следовательно, данные, – это такая информация, которая может быть как осмысленной, так и не осмысленной.***

Это корректная попытка дать определение более общего понятия «Данные», через частный случай данных: понятие «Информация» и специфический признак информации: *осмысленность*.

Смысл данных, в соответствии с концепцией смысла Шенка-Абельсона [18], состоит в том, что известны причинно-следственные зависимости между событиями, которые описываются этими данными. Понятие причинно-следственных связей относится к реальной области. Данные же являются лишь моделью, с определенной степенью адекватности *отражающей* реальную предметной область. Поэтому в данных никаких причинно-следственных связей нет и выявить их в данных невозможно.

Но причинно-следственные связи вполне возможно выявить между *событиями*, отражаемыми этими данными. Но для этого нужно предварительно *преобразовать базу исходных данных в базу событий*.

Операция выявления причинно-следственных связей между событиями, отраженными в данных, называется «Анализ данных». По сути, анализ данных представляет собой их осмысление и преобразование в информацию.

Например, анализируя временные ряды, отражающие события на фондовом рынке, мы начинаем замечать, что если вырос спрос на какую-либо валюту, то за этим обычно следует повышение ее курса.

Анализ данных включает следующие этапы:

1. Выявление событий в данных:

– разработка классификационных и описательных шкал и градаций;

– преобразование исходных в базу событий – эвентологическую базу, путем кодирования исходных данных с применением классификационных и описательных шкал и градаций, т. е. по сути, путем нормализации исходных данных.

2. Выявление причинно-следственных зависимостей между событиями в эвентологической базе данных.

В случае систем управления, событиями в данных являются совпадения определенных значений входных факторов и выходных параметров объекта управления, т. е. по сути, случаи перехода объекта управления в определенные будущие состояния, соответствующие классам, под действием определенных сочетаний значений управляющих факторов. *Качественные* значения входных факторов и выходных параметров естественно формализовать в форме лингвистических переменных. Если же входные факторы и выходные параметры являются *числовыми*, то их значения измеряются с некоторой погрешностью и фактически представляют собой *интервальные числовые значения*, которые также могут быть представлены или формализованы в форме порядковых лингвистических переменных (типа: «малые», «средние», «большие» значения показателей).

Какие же ***математические меры*** могут быть использованы для количественного измерения силы и направления причинно-следственных зависимостей?

Наиболее очевидным ответом на этот вопрос, который обычно первым всем приходит на ум, является: «Корреляция». Однако, в статистике хорошо известно, что это совершенно не так, т. к. для выявления причинно-следственных связей в соответствии с методом научной индукции (Ф. Бэкон, Дж. Милль) необходимо сравнивать результаты, по крайней мере, в двух группах, в одной из которых фактор действовал, а в другой нет.

Например, на плакате, выпущенном полицией[3], написано: «*По статистике, порядка 7,5-8 % аварий в России ежегодно совершается по*

[3] Автор такой плакат видел, когда проходил медосмотр перед получением прав нового образца.

вине водителей, находящихся в состоянии алкогольного опьянения»[4]. Все. Точка. Больше ничего не написано. Однако, чтобы понять, является ли состояние алкогольного опьянения фактором, увеличивающим риск совершения ДТП или его тяжесть, этой информации недостаточно. Для этого обязательно необходима также информация о том, *сколько процентов аварий в России ежегодно совершается по вине трезвых водителей. Но эта информация не приводится*, поэтому формально здесь возможно три варианта:

1) по вине трезвых водителей аварий совершается меньше, чем по вине пьяных;

2) по вине трезвых водителей аварий совершается столько же, сколько по вине пьяных;

3) по вине трезвых водителей аварий совершается больше, чем по вине пьяных.

Первый вариант содержит информацию о том, что опьянение – это фактор риска совершения ДТП, второй – что это никак не влияет на риск совершения ДТП, а третий – что опьянение уменьшает его. Конечно, все понимают, что в жизни реализуется 1-ф вариант. Но об этом ведь в данном плакате нет прямых статистических данных. Таким образом, знак разности этих процентов определяет направление влияния этого фактора, а модуль этой разности силу его влияния, что и используется как один из частных критериев знаний в АСК-анализе и системе «Эйдос» [14, 20, 21].

Для преобразования исходных данных в информацию необходимо не только выявить события в этих данных, но и найти причинно-следственные связи между этими событиями. В АСК-анализе предлагается 7 количественных мер причинно-следственных связей, основной из которых является семантическая мера целесообразности информации по А.Харкевичу. Все эти меры причинно-следственных связей основаны на ***сравнении*** условных вероятностей встречи различных значений факторов при переходе объекта моделирования в различные состояния с безусловной вероятностью их встречи по всей выборке.

11.3. Знания и информация

Знания – это информация, полезная для достижения целей, т. е. для управления.

Значит для преобразования информации в знания необходимо:

1. ***Поставить цель*** *(классифицировать будущие состояния моделируемого объекта на целевые и нежелательные в какой-то шкале, лучше всего в порядковой или числовой).*

2. Оценить полезность информации для достижения этой цели (знак и силу влияния).

[4] См., например: https://cnev.ru/polezno/stati/osnovnye-prichiny-dtp-pyanstvo-za-rulem

Второй пункт, по сути, выполнен при преобразовании данных в информацию. Поэтому остается выполнить только первый пункт, т. к. классифицировать будущие состояния объекта управления как желательные (целевые) и нежелательные.

Знания могут быть представлены в различных формах, характеризующихся различной степенью формализации:

– вообще неформализованные знания, т. е. знания в своей собственной форме, ноу-хау (мышление без вербализации есть медитация);

– знания, формализованные в естественном вербальном языке;

– знания, формализованные в виде различных методик, схем, алгоритмов, планов, таблиц и отношений между ними (базы данных);

– знания в форме технологий, организационных, производственных, социально-экономических и политических структур;

– знания, формализованные в виде математических моделей и методов представления знаний в автоматизированных интеллектуальных системах (логическая, фреймовая, сетевая, продукционная, нейросетевая, нечеткая и другие).

Таким образом, для решения сформулированной проблемы необходимо осознанно и целенаправленно последовательно повышать степень формализации исходных данных до уровня, который позволяет ввести исходные данные в интеллектуальную систему, а затем:

– преобразовать исходные данные в информацию;

– преобразовать информацию в знания;

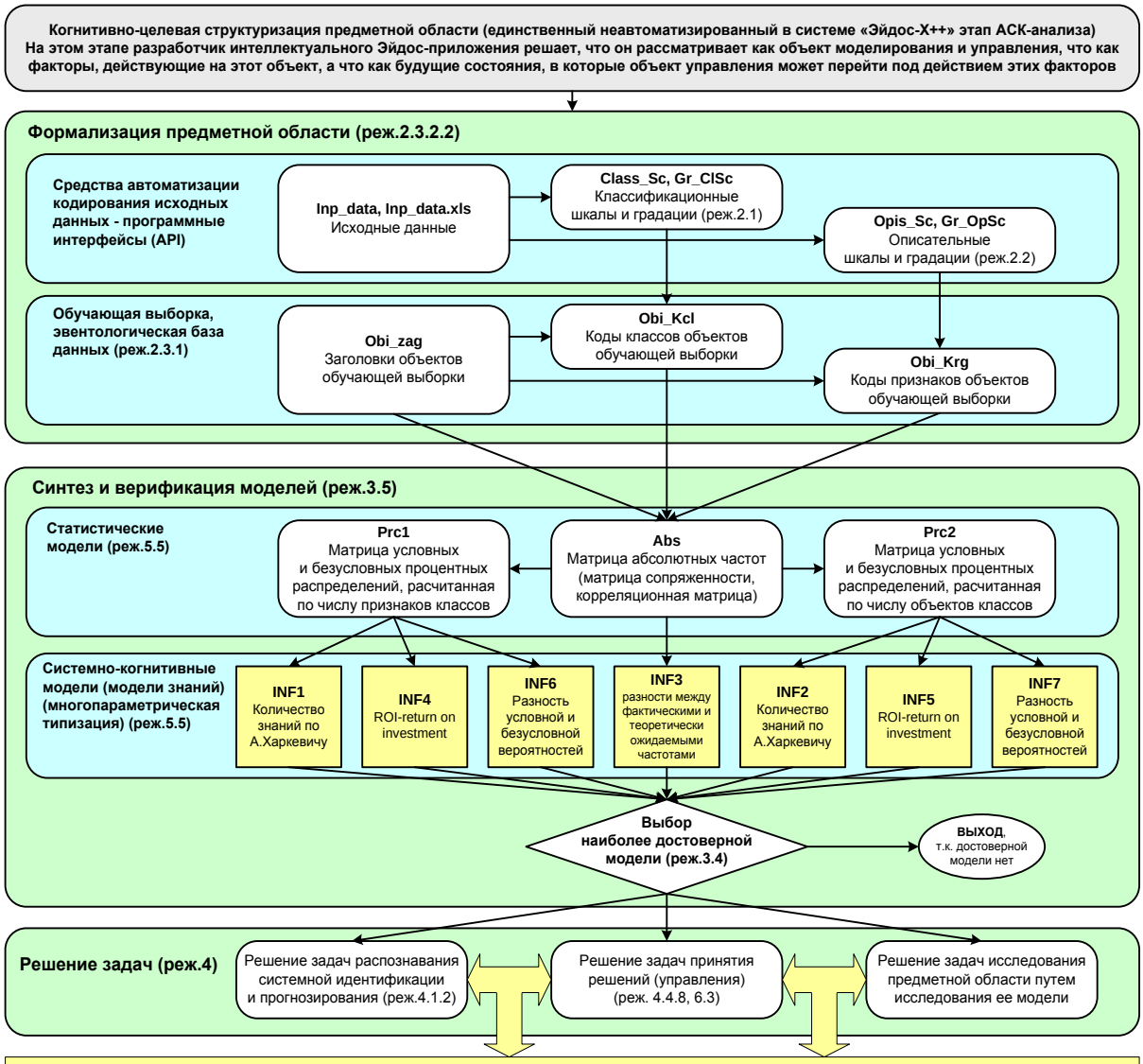
– использовать знания для решения задач управления, принятия решений и исследования предметной области.

Процесс преобразования данных в информацию, а ее в знания называется анализ. Основные его этапы приведены на рисунке 1.

Рисунок 1. Преобразование данных в информацию, а ее знания

В системе «Эйдос» этот процесс осуществляется в следующей последовательности (рисунок 2).

Последовательность обработки данных, информации и знаний в системе «Эйдос», повышение уровня системности данных, информации и знаний, повышение уровня системности моделей



Если модель адекватна, **достоверна**, т.е. соответствует действительности, то и результаты решения задач в этой модели также соответствуют действительности. Это значит, что если достоверность модели низка или неизвестна, то применять ее для решения реальных задач нельзя. Если же это делается, то является авантюризмом и профанацией науки.

О соотношении задач.

- распознавание, классификация, идентификация и диагностика (это одно и тоже, т.е. синонимы). При решении этих задач определяется степень сходства/различия образа конкретного объекта с обобщенными образами классов.
- идентификация и прогнозирование (при идентификации значения свойств и принадлежность объекта к классу относятся к одному моменту времени, а при прогнозировании значения факторов относятся к прошлому, а переход объекта под действием этих факторов в состояние, соответствующее классу относится к будущему, по сути, **прогнозирование - это идентификация будущих состояний, т.е. это тоже идентификация, но не в пространстве (настоящем), а в пространстве-времени**;
- прогнозирование и принятие решений (при прогнозировании по значениям факторов, действующих на объект моделирования, определяется в какое будущее состояние он перейдет под их действием. При принятии решений, наоборот, по будущему целевому состоянию объекта моделирования определяются значения факторов, которые обуславливают его переход в это будущее целевое состояние. Таким образом **задача принятия решений является обратной по отношению к задаче прогнозирования**);
- принятие решений путем многократного многовариантного прогнозирования при различных сочетаниях значений факторов невозможно из-за комбинаторного взрыва. Прогнозирование может быть **элементом** принятия решения, т.е. применено для оценки **адекватности** рассматриваемого уже ранее сформированного другим методом варианта решения, но оно в реальных случаях, т.е. когда много факторов, не может быть применено для выработки самого варианта решения;
- принятие решений и исследование моделируемой предметной области (задача принятия решений является обратной по отношению к задаче прогнозирования только в простейшем случае: в случае использования SWOT-анализа. Однако SWOT-анализ имеет свои ограничения: может быть задано только **одно** будущее целевое состояние, некоторые рекомендуемые факторы может **не быть** технологической и финансовой **возможности использовать**. Поэтому в АСК-анализе и системе «Эйдос» реализован развитый алгоритм принятия решений п.6.3 в котором кроме SWOT-анализа используются также результаты решения задачи прогнозирования и результаты кластерно-конструктивного анализа классов и значений факторов, т.е. некоторые результаты решения задачи исследования предметной области.)

Рисунок 2. Преобразование данных в информацию, а ее знания в системе «Эйдос»

Основные публикации автора по вопросам выявления, представления и использования знаний [13, 17].

Из вышеизложенного можно сделать обоснованный вывод о том, что *АСК-анализ и система «Эйдос» обеспечивают движение познания от эмпирических данных к информации, а от нее к знаниям. По сути, это движение от феноменологических моделей, описывающих явления внешне, к содержательным теоретическим моделям [17].*

Появляется все больше сайтов, посвященных искусственному интеллекту, в открытом доступе появляются базы данных для машинного обучения (UCI[5], Kaggle[6] и другие) и даже on-line интеллектуальные приложения, совершенствуется и интерфейсы, применяемые в Internet.

В этом смысле показательно приобретение разработчиком одной из первых и наиболее популярный по сегодняшний день глобальных социальных сетей Facebook Марком Цукербергом фирмы Oculus, являющейся ведущим в мире разработчиком и производителем амуниции виртуальной реальности.

Однако учащиеся и ученые до сих пор практически не замечают, что уже давно существует и действует открытая масштабируемая интерактивная интеллектуальная on-line среда для обучения и научных исследований, основанная на автоматизированном системно-когнитивном анализе (АСК-анализ) и его программном инструментарии – интеллектуальной системе «Эйдос», а также сайте автора.

Соотношение между содержанием понятий: «Данные», «Информация» и «Знания» наглядно показаны на рисунке 1. На рисунке 2 приведена схема преобразования данных в информацию, а ее в знания и решения на этой основе ряда задач в АСК-анализе и интеллектуальной системе «Эйдос».

Ниже рассмотрим основные компоненты этой среды подробнее.

11.4. От больших данных к большой информации, а от нее к большим знаниям

Internet постепенно интеллектуализируется и превращается из нелокального хранилища больших данных (***big data***) в информационное пространство, содержащее осмысленные большие данные, т. е. «*большую информацию*» (***great info***), а затем в пространство знаний или «когнитивное пространство», в котором большая информация активно используется для достижения целей (управления) и тем самым превращается в «*большие знания*» (***great knowledge***).

Рекомендуемая литература

[5] http://archive.ics.uci.edu/ml/datasets.html

[6] https://www.kaggle.com/datasets

Луценко Е.В. Когнитивная ветеринария – ветеринария цифрового общества: дефиниция базовых понятий / Е.В. Луценко, Е.К. Печурина, А.Э. Сергеев // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2019. – №08(152). С. 141 – 199. – IDA [article ID]: 1521908015. – Режим доступа: http://ej.kubagro.ru/2019/08/pdf/15.pdf, 3,688 у.п.л.

11.5. Основные термины баз данных, информационных и интеллектуальных систем

Банк данных – это базы данных плюс система *управления базами данных* (СУБД) (стандартные термины). СУБД – это, по сути, *система управления данными*.

Информационный банк – это информационные базы плюс информационные системы (предлагается стандартизировать эти термины). Информационная система – это, по сути, *система управления информацией*.

Банк знаний – это базы знаний плюс интеллектуальные системы (стандартные термины). Интеллектуальная система – это, по сути, *система управления знаниями*.

Существует очевидная параллель между терминами и понятиями, связанными с данными, информацией и знаниями, наглядно представленная в таблице 1.

Таблица 1 – ПАРАЛЛЕЛЬ МЕЖДУ ПОНЯТИЯМИ И ТЕРМИНАМИ, КАСАЮЩИМИСЯ ДАННЫХ, ИНФОРМАЦИИ И ЗНАНИЙ

Объект	**Субъект**	**Система**
База данных (БД)	Система управления базами данных (СУБД)	Банк данных = БД+СУБД
Информационная база (ИБ)	Информационная (аналитическая) система ***(система управления информационными базами – СУИБ)***	***Информационный банк = ИБ+СУИБ***
База знаний (БЗ)	Интеллектуальная система ***(система управления базами знаний – СУБЗ)***	***Банк знаний = БЗ+СУБЗ***

Поэтому неверно говорить о базах данных, понимая под этим по существу банки данных, т.к. база данных – это просто данные на носителях, а банк данных включает кроме самой базы данных еще и программную систему управления этими базами данных.

Автор предлагает «узаконить», т.е. стандартизировать термины, отмеченные в таблице 2 красным цветом. Это позволит упорядочить все эти термины в единой стройной системе, построенной на основе соотношения содержания понятий «данные», «информация» и «знания».

Это актуально, т.к. в настоящее время существуют явная путаница в использовании этих понятий, встречающая даже в названиях соответствующих дисциплин: «Управление знаниями»,

«Интеллектуальные информационные системы», «Представление знаний в информационных системах». Например, дисциплина «Управление знаниями» является *гуманитарной* и в ней изучаются слабо формализованные, не основанные на применении автоматизированных интеллектуальных систем, этапы, формы и методы управления знаниями[7]. Вместе с тем название этой дисциплины явно соотносится с названием дисциплины «Управление данными». Интеллектуальные системы часто некорректно называются интеллектуальными информационными системами, с тем же успехом их можно было бы называть: «Интеллектуальные СУБД», но лучше и правильнее было бы называть их как предложено: «Системы управления базами знаний». Дисциплина «Алгоритмы и структуры данных» соотносится с дисциплиной «Представление знаний в информационных системах», хотя ясно, что они представляются не в информационных, а в интеллектуальных системах, т.к. в информационных системах представляется информация, а не знания, а знания представляются в интеллектуальных системах. В настоящее время дисциплина «Интеллектуальные информационные системы» по своему содержанию включает «Представление знаний в информационных системах», тогда как из вышеизложенного ясно, что они должны соотносится по своему содержанию также, как СУБД и «Модели баз данных» (в которых обычно подробно преподается лишь одна реляционная модель).

Отметим также, что если применить данное выше определение знаний к моделям, описываемым в дисциплине «Представление знаний в информационных системах», то обнаруживается, что иногда в ней описываются не модели баз знаний, а модели баз данных или информационные модели. В частности это видно на примере семантических сетей, которые, по сути, представляют собой не более чем инфологическую модель реляционной базы данных.

Дисциплины «Управление знаниями» и «Представление знаний в интеллектуальных системах» по сути, представляют собой две части одной дисциплины и должны отражать не способы управления знаниями различной степени формализации (как в настоящее время), а описание автоматизированных интеллектуальных систем и баз знаний.

Существует дисциплина: «Алгоритмы и структуры данных». Предлагается ввести аналогичные дисциплины: «Алгоритмы и информационные структуры» (в АСК-анализе – это формализация предметной области и синтез модели) и «Алгоритмы структурирования

[7] Типичные вопросы, изучаемые в этой дисциплине: стратегия управления знаниями предприятия; организационная культура в контексте управления знаниями; измерение интеллектуального капитала; корпоративные знания: как ими управлять; интеграция знаний предприятия; бизнес держится на знаниях, сам того не зная; новые программы корпоративного обучения в среде управления знаниями: опыт зарубежных компаний; менеджмент знаний: подход к внедрению; общепринятых заблуждений об управлении знаниями (knowledge management)

знаний» (по содержанию близко к когнитологии, инженерии знаний, представлению знаний)».

Конечно, интеллектуальные системы являются и информационными (аналитическими) системами, и системами управления базами данных (рисунок 1). Информационные (аналитические) системы, являются системами управления базами данных (рисунок 1). Но не всякая система управления базами данных является информационной (аналитической) системой, а лишь такая, в которой в результате анализа данных выявляется их смысл и они преобразуются в информацию. И не всякая информационная (аналитическая) система является интеллектуальной, а лишь такая, в которой в результате постановки цели и решения задачи управления (т.е. достижения цели) информация преобразуется в знание. Поэтому об авторах образовательных стандартов, в которых предлагается вести дисциплину: «Информационные интеллектуальные системы» можно сказать, что они не вполне понимают, о чем говорят.

Факт наличия причинно-следственных зависимостей может быть установлен методом хи-квадрат, а ее вид – многофакторным анализом. Однако факторный анализ позволяет обрабатывать данные лишь очень небольших размерностей (по числу факторов) и предъявляет чрезвычайно жесткие требования к наличию полных повторностей всех вариантов сочетаний факторов в исходных данных (т.е. данные не должны быть фрагментарными), что на практике выполнить удается крайне редко.

Поэтому большой интерес представляют другие подходы, обеспечивающие в различных предметных областях, в частности в ветеринарии, ***применение информационных и когнитивных технологий для выявления силы и направления причинно-следственных зависимостей в эмпирических данных.***

При этом будут возникать новые направления науки. Широкое применение математических методов в экономике привело к возникновению такого направления науки и специальности ВАК РФ, как 08.00.13 - Математические и инструментальные методы экономики». Аналогично, ветеринария, широко применяющая информационные и когнитивные технологии для выявление силы и направления причинно-следственных зависимостей в эмпирических данных, станет ***когнитивной ветеринарией***.

Рекомендуемая литература

Луценко Е.В. Когнитивная ветеринария – ветеринария цифрового общества: дефиниция базовых понятий / Е.В. Луценко, Е.К. Печурина, А.Э. Сергеев // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2019. – №08(152). С. 141 – 199. – IDA [article ID]: 1521908015. – Режим доступа: http://ej.kubagro.ru/2019/08/pdf/15.pdf, 3,688 у.п.л.

11.6. Критерии идентификации банков данных, информационных и интеллектуальных систем

Эти критерии очевидны из предыдущего изложения.

Отметим, что довольно часто даже специалисты не проводят особых различий между базами данных и банками данных, называя банки данных базами данных. Обычно при этом подразумевается, что это не просто данные на носителях, но и программное обеспечение манипулирования ими. Таким образом, научная терминология используется неправильно, некорректно.

Аналогично часто банки данных, т.е. СУБД с базами данных, довольно часто называют информационными системами, подразумевая при этом, что в базах данных содержится информация, а не данные, т.е. не понимая, чем отличается информация от данных.

Более того, даже в учебниках по моделям представления знаний в базах знаний интеллектуальных систем часто фактически описываются не базы знаний, а информационные базы, или даже просто базы данных.

Поэтому вопрос о критериях идентификации банков данных, информационных и интеллектуальных систем является весьма актуальным. Если его решить то терминологическая путаница в головах учащихся и специалистов, а также в соответствующей литературе, может несколько уменьшится. Хотя если оставаться на реальной почве, то можно признать, что надежд на это немного.

Опираясь на таблицу 2 мы можем сказать, что если в программной системе, работающей с базами данных, есть классификационные и описательные шкалы и градации; база событий (эвентологическая база), базы причинно-следственных зависимостей между событиями в эвентологической базе данных, т.е. система обеспечивает выявление смысла данных и его использование для решение различных задач, то есть все основания называть эту систему информационной или аналитической системой.

Опираясь на таблицу 2 мы можем сказать, что если в программной системе, работающей с базами данных, количественно выявляется сходство/различие между: 1) образами конкретных объектов и обобщенными образами классов; 2) образами классов; 3) значениями факторов; 4) количество информации в значениях факторов о достижении цели и это используется для решения задач идентификации, прогнозирования, классификации, поддержки принятия решений по достижению поставленной цели, исследования моделируемой предметной области путем исследования ее модели, то есть все основания называть эту систему интеллектуальной системой или системой искусственного интеллекта (СИИ).

Надо признать, что эти требования довольно жесткие и ***многие из систем фактически не оправдывают своих названий***, т.е. в названии

системы заявляются более развитые функции, чем система поддерживает фактически.

С другой стороны вполне понятно желание разработчиков назвать свою систему красивым и модным сочетанием слов: «Информационная система», «Аналитическая система» или даже «система искусственного интеллекта», «Интеллектуальная система».

Но надо понимать, что когда разработчик делает это по сути не понимая смысла используемых им терминов, то он часто поддается соблазну выдать желаемое за действительное и ***вводит потенциальных пользователей в заблуждение, т.е. попросту обманывает их***.

В этом нет ничего удивительного, специалисты по рекламе практически всегда так и делают[8], но это не может оправдать ни разработчиков программных систем, ни специалистов по рекламе.

Рекомендуемая литература

Луценко Е.В. Когнитивная ветеринария – ветеринария цифрового общества: дефиниция базовых понятий / Е.В. Луценко, Е.К. Печурина, А.Э. Сергеев // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2019. – №08(152). С. 141 – 199. – IDA [article ID]: 1521908015. – Режим доступа: http://ej.kubagro.ru/2019/08/pdf/15.pdf, 3,688 у.п.л.

ГЛАВА 12. БАЗОВЫЙ АВТОМАТИЗИРОВАННЫЙ СИСТЕМНО-КОГНИТИВНЫЙ АНАЛИЗ И СИСТЕМА ЭЙДОС КАК МЕТОД И ИНСТРУМЕНТАРИЙ РЕШЕНИЯ ЗАДАЧ

12.1. Очень кратко об АСК-анализе

Автоматизированный системно-когнитивный анализ (АСК-анализ) предложен автором в 2002 году в ряде статей и фундаментальной монографии [1]. *Сам термин: «Автоматизированный системно-когнитивный анализ (АСК-анализ)»* ***был предложен профессором Е.В.Луценко в ряде работ в 2001-2002 годах***. *На тот момент он вообще не встречался в Internet.* Сегодня по соответствующему запросу в Яндексе находится 9 миллионов сайтов с этим сочетанием слов[9].

АСК-анализ включает:

[8] Например, показывают рекламный ролик с молодой женщиной, которой не более 25 лет, и вдруг она заявляет: «мне 40», и улыбается, а мы уже сами догадываемся, что а выглядит она так молодо потому, что использует рекламируемый ею чудодейственный крем, омолаживающий кожу лица.

[9] https://yandex.ru/search/?lr=35&clid=2327117-18&win=360&text=%20360&text=Автоматизированный+системно-когнитивный+анализ+(АСК-анализ)

– теоретические основы, в частности базовую формализуемую когнитивную концепцию;

– математическую модель, основанную на системном обобщении теории информации (СТИ);

– методику численных расчетов (структуры баз данных и алгоритмы их обработки);

– программный инструментарий, в качестве которого в настоящее время выступает универсальная когнитивная аналитическая система «Эйдос» (интеллектуальная система «Эйдос»).

Более подробно АСК-анализ описан в работах [2, 3] и ряде других. Около половины из 655 опубликованных автором научных работ посвящены теоретическим основам АСК-анализа и его практическим применениям в ряде предметных областей. На момент написания данной работы автором опубликовано более 39 монографий, 27 учебных пособий, в т.ч. 3 учебных пособия с грифами УМО и Министерства, получен 31 патент РФ на системы искусственного интеллекта, 335 публикации в изданиях, входящих в перечень ВАК РФ и приравненных им (по данным РИНЦ), 6 статей в журналах, входящих в WoS, 5 публикаций в журналах, входящих в Скопус[10].

Три монографии включены в фонды библиотеки конгресса США[11].

АСК-анализ и система "Эйдос" были успешно применены в 9 докторских и 8 кандидатских диссертациях по экономическим, техническим, биологическим, психологическими и медицинским наукам, еще несколько докторских диссертаций с применением АСК-анализа в стадии выхода на защиту. Автор является основателем междисциплинарной научной школы: «Автоматизированный системно-когнитивный анализ»[12]. Научная школа: "Автоматизированный системно-когнитивный анализ" является междисциплинарным научным направлением на пересечении по крайней мере трех научных специальностей (согласно недавно утвержденной новой номенклатуры научных специальностей ВАК РФ[13]). Основные научные специальности, которым соответствует научная школа:

– 5.12.4. Когнитивное моделирование;

– 1.2.1. Искусственный интеллект и машинное обучение;

– 2.3.1. Системный анализ, управление и обработка информации.

Научная школа: "Автоматизированный системно-когнитивный анализ" включает следующие междисциплинарные научные направления:

– Автоматизированный системно-когнитивный анализ числовых и текстовых табличных данных;

---

[10] http://lc.kubagro.ru/aidos/Sprab0802.pdf

[11] https://catalog.loc.gov/vwebv/search?searchArg=Lutsenko+E.V. (и кликнуть: “**Search**”)

[12] https://www.famous-scientists.ru/school/1608

[13] https://www.garant.ru/products/ipo/prime/doc/400450248/

– Автоматизированный системно-когнитивный анализ текстовых данных;

– Спектральный и контурный автоматизированный системно-когнитивный анализ изображений;

– Сценарный автоматизированный системно-когнитивный анализ временных и динамических рядов.

Приводить здесь ссылки на все эти работы вряд ли целесообразно. Отметим лишь, что у автора есть личный сайт [4] и страничка в РесечГейт [5], на которых можно получить более полную информацию о методе АСК-анализа. Краткая информация об АСК-анализе и системе «Эйдос» есть в материале: http://lc.kubagro.ru/aidos/Presentation_Aidos-online.pdf.

12.2. Очень кратко о системе «Эйдос»

Существует много систем искусственного интеллекта. Универсальная когнитивная аналитическая система «Эйдос» отличается от них следующими параметрами:

– является универсальной и может быть применена во многих предметных областях, т.к. разработана в универсальной постановке, не зависящей от предметной области (http://lc.kubagro.ru/aidos/index.htm). Система «Эйдос» является автоматизированной системой, т.е. предполагает непосредственное участие человека в реальном времени при решении задач идентификации, прогнозирования, приятия решений и исследования предметной области (автоматические системы работают без такого участия человека);

– находится в полном открытом бесплатном доступе (http://lc.kubagro.ru/aidos/_Aidos-X.htm), причем с актуальными исходными текстами (http://lc.kubagro.ru/__AidosALL.txt): открытая лицензия: CC BY-SA 4.0 (https://creativecommons.org/licenses/by-sa/4.0/), и это означает, что ей могут пользоваться все, кто пожелает, без какого-либо дополнительного разрешения со стороны первичного правообладателя – автора системы «Эйдос» проф. Е.В.Луценко (отметим, что система «Эйдос» создана полностью с использованием только лицензионного инструментального программного обеспечения и на нее имеется 31 свидетельство РосПатента РФ);

– является одной из первых отечественных систем искусственного интеллекта персонального уровня, т.е. не требует от пользователя специальной подготовки в области технологий искусственного интеллекта: «имеет нулевой порог входа» (есть акт внедрения системы «Эйдос» 1987 года) (http://lc.kubagro.ru/aidos/aidos02/PR-4.htm);

– реально работает, обеспечивает устойчивое выявление в сопоставимой форме силы и направления причинно-следственных зависимостей в неполных зашумленных взаимозависимых (нелинейных) данных очень большой размерности числовой и не числовой природы, измеряемых в различных типах шкал (номинальных, порядковых и

числовых) и в различных единицах измерения (т.е. не предъявляет жестких требований к данным, которые невозможно выполнить, а обрабатывает те данные, которые есть);

– имеет «нулевой порог входа», содержит большое количество локальных (поставляемых с инсталляцией) и облачных учебных и научных Эйдос-приложений (в настоящее время их 31 и более 300, соответственно: http://aidos.byethost5.com/Source_data_applications/WebAppls.htm) (http://lc.kubagro.ru/aidos/Presentation_Aidos-online.pdf);

– поддерживает on-line среду накопления знаний и обмена ими, широко используется во всем мире (http://aidos.byethost5.com/map5.php);

– обеспечивает мультиязычную поддержку интерфейса на 51 языке. Языковые базы входят в инсталляцию и могут пополняться в автоматическом режиме;

– наиболее трудоемкие в вычислительном отношении операции синтеза моделей и распознавания реализует с помощью графического процессора (GPU), что на некоторых задачах обеспечивает ускорение решение этих задач в несколько тысяч раз, что реально обеспечивает интеллектуальную обработку больших данных, большой информации и больших знаний (графический процессор должен быть на чипсете NVIDIA);

– обеспечивает преобразование исходных эмпирических данных в информацию, а ее в знания и решение с использованием этих знаний задач классификации, поддержки принятия решений и исследования предметной области путем исследования ее системно-когнитивной модели, генерируя при этом очень большое количество табличных и графических выходных форм (развития когнитивная графика), у многих из которых нет никаких аналогов в других системах (примеры форм можно посмотреть в работе: http://lc.kubagro.ru/aidos/aidos18_LLS/aidos18_LLS.pdf);

– хорошо имитирует человеческий стиль мышления: дает результаты анализа, понятные экспертам на основе их опыта, интуиции и профессиональной компетенции;

– вместо того, чтобы предъявлять к исходным данным практически неосуществимые требования (вроде нормальности распределения, абсолютной точности и полных повторностей всех сочетаний значений факторов и их полной независимости и аддитивности) автоматизированный системно-когнитивный анализ (АСК-анализ) предлагает без какой-либо предварительной обработки осмыслить эти данные и тем самым преобразовать их в информацию, а затем преобразовать эту информацию в знания путем ее применения для достижения целей (т.е. для управления) и решения задач классификации, поддержки принятия решений и содержательного эмпирического исследования моделируемой предметной области.

В чем сила подхода, реализованного в системе Эйдос? В том, что она реализует подход, эффективность которого не зависит от того, что мы думаем о предметной области и думаем ли вообще. Она формирует модели непосредственно на основе эмпирических данных, а не на основе наших представлений о механизмах реализации закономерностей в этих данных. Именно поэтому Эйдос-модели эффективны даже если наши представления о предметной области ошибочны или вообще отсутствуют.

В этом и слабость этого подхода, реализованного в системе Эйдос. Модели системы Эйдос - это феноменологические модели, отражающие эмпирические закономерности в фактах обучающей выборки, т.е. они не отражают причинно-следственного механизма детерминации, а только сам факт и характер детерминации. Содержательное объяснение этих эмпирических закономерностей формулируется уже экспертами на теоретическом уровне познания в содержательных научных законах14.

В разработке системы «Эйдос» были следующие этапы:

1-й этап, «подготовительный»: 1979-1992 годы. Математическая модель системы "Эйдос" разработана в 1979 и впервые прошла экспериментальную апробацию в 1981 году (первый расчет на компьютере на основе модели). С 1981 по 1992 система "Эйдос" неоднократно реализовалась на платформе Wang (на компьютерах Wang-2200C). В 1987 году впервые получен акт внедрения[15] на одну из ранних версий системы «Эйдос», реализованную в среде персональной технологической системы «Вега-М» разработки автора (см.2-й акт).

2-й этап, «эра IBM PC и MS DOS»: 1992-2012 годы. Для IBM-совместимых персональных компьютеров система "Эйдос" впервые реализована на языках CLIPPER-87 и CLIPPER-5.01 (5.02) в 1992 году, а в 1994 году уже были получены свидетельства РосПатента[16], первые в Краснодарском крае и, возможно, в России на системы искусственного интеллекта. С тех пор и до настоящего времени система непрерывно совершенствуется на IBM PC.

3-й этап, «эра MS Windows xp, 8, 7: 2012-2020 годы. С июня 2012 по 14.12.2020 система «Эйдос» развивалась на языке Аляска-1.9 + Экспресс++ + библиотека для работы с Internet xb2net. Система «Эйдос-X1.9» хорошо работала на всех версиях MS Windows кроме Windows-10, которая требовала специальной настройки. Наиболее трудоемкие в вычислительном отношении операции синтеза моделей и распознавания реализует с помощью графического процессора (GPU), что на некоторых задачах обеспечивает ускорение решение этих задач в несколько тысяч раз, что реально обеспечивает интеллектуальную обработку больших данных,

[14] Ссылка на это краткое описание системы «Эйдос» на английском языке: http://lc.kubagro.ru/aidos/The_Eidos_en.htm

[15] http://lc.kubagro.ru/aidos/aidos02/PR-4.htm

[16] http://lc.kubagro.ru/aidos/index.htm

большой информации и больших знаний (графический процессор должен быть на чипсете NVIDIA).

4-й этап, «эра MS Windows-10: с 2020 года по настоящее время. С 13.12.2020 года по настоящее время система «Эйдос» развивается на языке Аляска-2.0 + Экспресс++. Библиотека xb2net в ней больше не используется, т.к. все возможности работы с Internet входят в базовые возможности языка программирования.

На рисунке 3 приведена титульная видеограмма DOS-версии системы «Эйдос», а на рисунке 4 – текущей версии системы «Эйдос»:

Рисунок 3. Титульная видеограмма DOS-версии системы «Эйдос» (до 2012 года)[17]

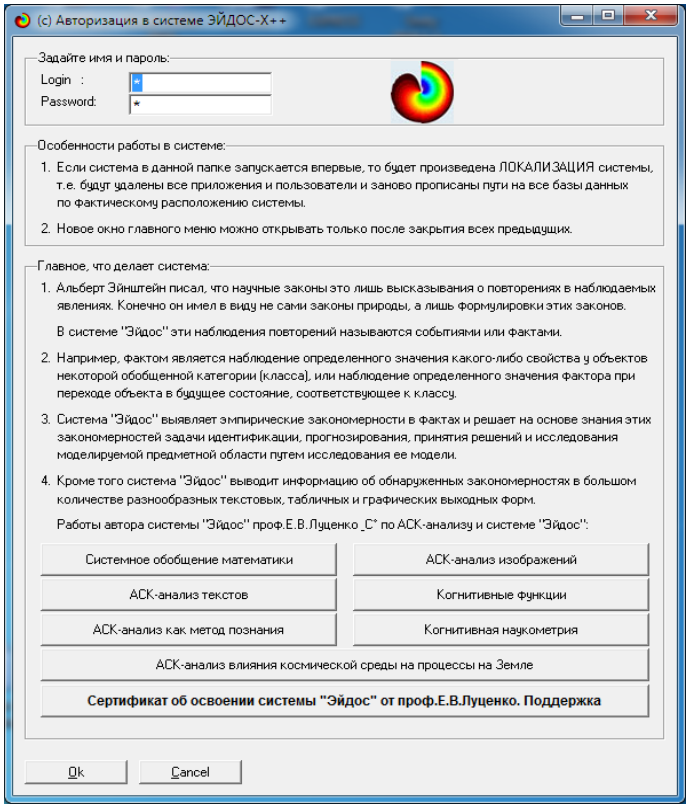


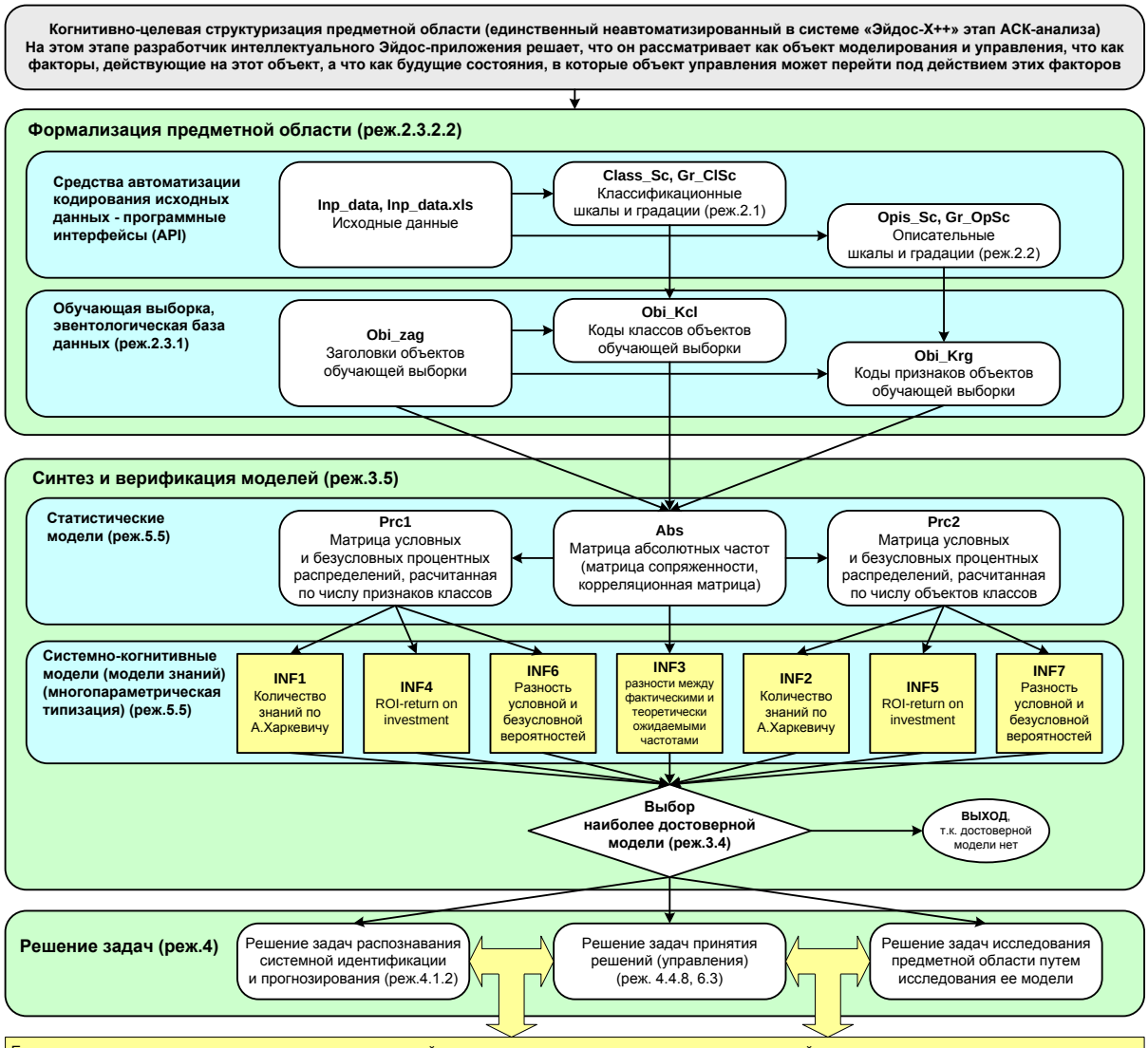
Рисунок 4. Титульная видеограмма текущей версии системы «Эйдос»

[17] http://lc.kubagro.ru/pic/aidos_titul.jpg

12.3. Немного подробнее об этапах АСК-анализа

Весь процесс создания моделей в АСК-анализе и использования этих моделей для решения подготовительных и реальных задач осуществляется в системе «Эйдос» и предусматривает следующие этапы, суть которых рассмотрена ниже (рисунок 5):

Последовательность обработки данных, информации и знаний в системе «Эйдос», повышение уровня системности данных, информации и знаний, повышение уровня системности моделей



Если модель адекватна, **достоверна**, т.е. соответствует действительности, то и результаты решения задач в этой модели также соответствуют действительности. Это значит, что если достоверность модели низка или неизвестна, то применять ее для решения реальных задач нельзя. Если же это делается, то является авантюризмом и профанацией науки.

О соотношении задач.

- распознавание, классификация, идентификация и диагностика (это одно и тоже, т.е. синонимы). При решении этих задач определяется степень сходства/различия образа конкретного объекта с обобщенными образами классов.
- идентификация и прогнозирование (при идентификации значения свойств и принадлежность объекта к классу относятся к одному моменту времени, а при прогнозировании значения факторов относятся к прошлому, а переход объекта под действием этих факторов в состояние, соответствующее классу относится к будущему, по сути, **прогнозирование - это идентификация будущих состояний, т.е. это тоже идентификация, но не в пространстве (настоящем), а в пространстве-времени**;
- прогнозирование и принятие решений (при прогнозировании по значениям факторов, действующих на объект моделирования, определяется в какое будущее состояние он перейдет под их действием. При принятии решений, наоборот, по будущему целевому состоянию объекта моделирования определяются значения факторов, которые обуславливают его переход в это будущее целевое состояние. Таким образом **задача принятия решений является обратной по отношению к задаче прогнозирования**);
- принятие решений путем многократного многовариантного прогнозирования при различных сочетаниях значений факторов невозможно из-за комбинаторного взрыва. Прогнозирование может быть **элементом** принятия решения, т.е. применено для оценки **адекватности** рассматриваемого уже ранее сформированного другим методом варианта решения, но оно в реальных случаях, т.е. когда много факторов, не может быть применено для выработки самого варианта решения;
- принятие решений и исследование моделируемой предметной области (задача принятия решений является обратной по отношению к задаче прогнозирования только в простейшем случае: в случае использования SWOT-анализа. Однако SWOT-анализ имеет свои ограничения: может быть задано только **одно** будущее целевое состояние, некоторые рекомендуемые факторы может **не быть** технологической и финансовой **возможности использовать**. Поэтому в АСК-анализе и системе «Эйдос» реализован развитый алгоритм принятия решений п.6.3 в котором кроме SWOT-анализа используются также результаты решения задачи прогнозирования и результаты кластерно-конструктивного анализа классов и значений факторов, т.е. некоторые результаты решения задачи исследования предметной области.)

Рисунок 5. Порядок обработки данных, информации и знаний в системы «Эйдос»

Ниже приведен краткий экскурс в возможности метода и его программного инструментария. Отметим, что в данной работе мы используем далеко не все эти возможности, а лишь некоторые из них, достаточные для решения поставленной в работе задачи.

12.3.1. Когнитивная структуризация предметной области. Две интерпретации классификационных и описательных шкал и градаций

На этапе когнитивно-целевой структуризации предметной области мы неформализуемым путем решаем на качественном уровне, что будем рассматривать в качестве факторов, действующих на моделируемый объект (причин), а что в качестве результатов действия этих факторов (последствий). По сути это постановка решаемой проблемы.

Описательные шкалы служат для формального описания факторов, а классификационные – результатов их действия на объект моделирования. Шкалы могут быть числовые и текстовые. Текстовые шкалы могут быть номинальные и порядковые.

Когнитивная структуризация предметной области является первым и единственным неавтоматизированном в системе «Эйдос» этапом АСК-анализа, т.е. все последующие этапы АСК анализа в ней полностью автоматизированы.

В АСК-анализе и системе «Эйдос» применяется две интерпретации классификационных и описательных шкал и градаций: ***статичная и динамичная*** и соответствующая терминология (обобщающая, статичная и динамичная).

Статичная интерпретация и терминология:

– градации классификационных шкал – это обобщающие категории видов объектов (классы);

– описательные шкалы – свойства объектов, градации описательных шкал – значения свойств (признаки) объектов.

Динамичная интерпретация и терминология:

– градации классификационных шкал – это обобщающие категории будущих состояний объекта моделирования (классы);

– описательные шкалы – факторы, действующие на объект моделирования, градации описательных шкал – значения факторов, действующие на объект моделирования.

Обобщающая терминология:

– классификационные шкалы и градации;

– описательные шкалы и градации;

12.3.2. Формализация предметной области

На этапе формализации предметной области разрабатываются классификационные и описательные шкалы и градации, а затем исходные

данные кодируются с их использованием, в результате чего получается обучающая выборка. Обучающая выборка, по сути, представляет собой исходные данные, *нормализованные* с помощью классификационных и описательных шкал и градаций.

В системе «Эйдос» имеется большое количество разнообразных автоматизированных программных интерфейсов (API), обеспечивающих ввод в систему внешних данных различных типов: текстовых, табличных и графических, а также других, которые могут быть представлены в этом виде, например аудио или данные электроэнцефалограммы (ЭЦГ) или кардиограммы (ЭКГ).

Этим обеспечивается возможность комфортного для пользователя применения системы «Эйдос» для проведения научных исследований в самых различных направлениях науки и решения практических задач в самых различных предметных областях, практически почти везде, где человек применяет естественный интеллект.

12.3.3. Синтез статистических и системно-когнитивных моделей (многопараметрическая типизация), частные критерии знаний

Синтез и верификация статистических и системно-когнитивных моделей (СК-моделей) моделей осуществляется в режиме 3.5 системы «Эйдос». Математические модели, на основе которых рассчитываются статистические и СК-модели, подробно описаны в ряде монографий и статей автора. Поэтому в данной работе мы рассмотрим эти вопросы очень кратко. Отметим лишь, что модели системы «Эйдос» основаны на матрице абсолютных частот, отражающей число встреч градаций описательных шкал по градациям классификационных шкал (фактов). Но для решения всех задач используется не непосредственно сама эта матрица, а матрицы условных и безусловных процентных распределений и системно-когнитивные модели, которые рассчитываются на ее основе и отражают какое количество информации содержится в факте наблюдения определенной градации описательной шкалы о том, что объект моделирования перейдет в состояние, соответствующее определенной градации классификационной шкалы (классу).

Математическая модель АСК-анализа и системы «Эйдос» основана на системной нечеткой интервальной математике и обеспечивает сопоставимую обработку больших объемов фрагментированных и зашумленных взаимозависимых данных, представленных в различных типах шкал (номинальных, порядковых и числовых) и различных единицах измерения.

Суть математической модели АСК-анализа состоит в следующем.

Непосредственно на основе эмпирических данных (см. Help режима 2.3.2.2) рассчитывается матрица абсолютных частот (таблица 1).

Таблица 2 – Матрица абсолютных частот (статистическая модель ABS)

		Классы					Сумма
		1	*...*	*j*	*...*	*W*	
Значения факторов	*1*	N_{11}		N_{1j}		N_{1W}	
	...						
	i	N_{i1}		N_{ij}		N_{iW}	$N_{i\Sigma} = \sum_{j=1}^{W} N_{ij}$
	...						
	M	N_{M1}		N_{Mj}		N_{MW}	
Суммарное количество признаков по классу				$N_{\Sigma j} = \sum_{i=1}^{M} N_{ij}$			$N_{\Sigma\Sigma} = \sum_{i=1}^{W}\sum_{j=1}^{M} N_{ij}$
Суммарное количество объектов обучающей выборки по классу				$N_{\Sigma j}$			$N_{\Sigma\Sigma} = \sum_{j=1}^{W} N_{\Sigma j}$

На ее основе рассчитываются матрицы условных и безусловных процентных распределений (таблица 2).

Таблица 3 – Матрица условных и безусловных процентных распределений (статистические модели PRC1 и PRC2)

		Классы					Безусловная вероятность признака
		1	*...*	*j*	*...*	*W*	
Значения факторов	*1*	P_{11}		P_{1j}		P_{1W}	
	...						
	i	P_{i1}		$P_{ij} = \frac{N_{ij}}{N_{\Sigma j}}$		P_{iW}	$P_{i\Sigma} = \frac{N_{i\Sigma}}{N_{\Sigma\Sigma}}$
	...						
	M	P_{M1}		P_{Mj}		P_{MW}	
Безусловная вероятность класса				$P_{\Sigma j}$			

Отметим, что в АСК-анализе и его программном инструментарии интеллектуальной системе «Эйдос» используется два способа расчета матриц условных и безусловных процентных распределений:

1-й способ: в качестве $N_{\Sigma j}$ используется суммарное количество признаков по классу;

2-й способ: в качестве $N_{\Sigma j}$ используется суммарное количество объектов обучающей выборки по классу.

На практике часто встречается существенная несбалансированность данных, под которой понимается сильно отличающееся количество объектов обучающейся выборки, относящихся к различным классам. Поэтому решать задачу на основе непосредственно матрицы абсолютных частот (таблица 1) было бы очень неразумно и переход от абсолютных частот к условным и безусловным относительным частотам (частостям) является весьма обоснованным и логичным.

Этот переход полностью снимает проблему несбалансированности данных, т.к. в последующем анализе используется не матрица абсолютных частот, а матрицы условных и безусловных процентных распределений и матрицы системно-когнитивных моделей (СК-модели, таблица 4), в частности матрица информативностей.

Этот подход снимает также проблему обеспечения сопоставимости обработки в одной модели исходных данных, представленных в различных видах шкал (номинальных, порядковых и числовых) и в разных единицах измерения [1].

В системе «Эйдос» это осуществляется всегда при решении любых задач.

Затем на основе таблицы 2 с использованием частных критериев, знаний приведенных таблице 3, рассчитываются матрицы системно-когнитивных моделей (таблица 4).

В таблице 3 приведены формулы:

– для сравнения **фактических и теоретических абсолютных частот;**

– для сравнения **условных и безусловных относительных частот** («вероятностей»).

И это сравнение в таблице 3 осуществляется двумя возможными способами: путем **вычитания** и путем **деления.**

Когда мы сравниваем фактические и теоретические абсолютные частоты путем вычитания у нас получается частный критерий знаний: «хи-квадрат» (СК-модель INF3), когда же мы сравниваем их путем деления, то у нас получается частный критерий: «количество информации по А.Харкевичу» (СК-модели INF1, INF2) или «коэффициент возврата инвестиций ROI» - Return On Investment (СК-модели INF4, INF5) в зависимости от способа нормировки.

Таблица 4– Различные аналитические формы частных критериев знаний, применяемые в АСК-анализе и системе «Эйдос»

Наименование модели знаний и частный критерий	Выражение для частного критерия	
	через относительные частоты	через абсолютные частоты
ABS, матрица абсолютных частот, N_{ij} - фактическое число встреч *i-го* признака у объектов *j-го* класса; $\overline{N}_{ij}$ - теоретическое число встреч *i-го* признака у объектов *j-го* класса; N_i – суммарное количество признаков в *i-й* строке; N_j – суммарное количество признаков или объектов обучающей выборки в *j-м* классе; N – суммарное количество признаков по всей выборке (таблица 1)	$N_i = \sum_{j=1}^{W} N_{ij};\ N_j = \sum_{i=1}^{M} N_{ij};\ N = \sum_{i=1}^{W}\sum_{j=1}^{M} N_{ij};$ $N_{ij} - \textit{фактическая частота};$ $\overline{N}_{ij} = \frac{N_i N_j}{N} - \textit{теоретическая частота}.$	
PRC1, матрица условных P_{ij} и безусловных P_i процентных распределений, в качестве N_j используется суммарное количество признаков по классу	---	$P_{ij} = \frac{N_{ij}}{N_j};\ P_i = \frac{N_i}{N}$
PRC2, матрица условных P_{ij} и безусловных P_i процентных распределений, в качестве N_j используется суммарное количество объектов обучающей выборки по классу		
INF1, частный критерий: количество знаний по А.Харкевичу, 1-й вариант расчета вероятностей: *Nj* – суммарное количество признаков по *j*-му классу. Вероятность того, что если у объекта j-го класса обнаружен признак, то это i-й признак	$I_{ij} = \Psi \times Log_2 \frac{P_{ij}}{P_i}$	$I_{ij} = \Psi \times Log_2 \frac{N_{ij}}{\overline{N}_{ij}} = \Psi \times Log_2 \frac{N_{ij} N}{N_i N_j}$
INF2, частный критерий: количество знаний по А.Харкевичу, 2-й вариант расчета вероятностей: *Nj* – суммарное количество объектов по *j*-му классу. Вероятность того, что если предъявлен объект j-го класса, то у него будет обнаружен i-й признак.		
INF3, частный критерий: **Хи-квадрат:** разности между фактическими и теоретически ожидаемыми абсолютными частотами	---	$I_{ij} = N_{ij} - \overline{N}_{ij} = N_{ij} - \frac{N_i N_j}{N}$
INF4, частный критерий: ROI - Return On Investment, 1-й вариант расчета вероятностей: *Nj* – суммарное количество признаков по *j*-му классу	$I_{ij} = \frac{P_{ij}}{P_i} - 1 = \frac{P_{ij} - P_i}{P_i}$	$I_{ij} = \frac{N_{ij}}{\overline{N}_{ij}} - 1 = \frac{N_{ij} N}{N_i N_j} - 1$
INF5, частный критерий: ROI - Return On Investment, 2-й вариант расчета вероятностей: *Nj* – суммарное количество объектов по *j*-му классу		
INF6, частный критерий: разность условной и безусловной вероятностей, 1-й вариант расчета вероятностей: *Nj* – суммарное количество признаков по *j*-му классу	$I_{ij} = P_{ij} - P_i$	$I_{ij} = \frac{N_{ij}}{N_j} - \frac{N_i}{N}$
INF7, частный критерий: разность условной и безусловной вероятностей, 2-й вариант расчета вероятностей: *Nj* – суммарное количество объектов по *j*-му классу		

Обозначения к таблице 3:
i – значение прошлого параметра;
j - значение будущего параметра;
N_{ij} – количество встреч j-го значения будущего параметра при i-м значении прошлого параметра;
M – суммарное число значений всех прошлых параметров;
W - суммарное число значений всех будущих параметров.
N_i – количество встреч i-м значения прошлого параметра по всей выборке;
N_j – количество встреч j-го значения будущего параметра по всей выборке;
N – количество встреч j-го значения будущего параметра при i-м значении прошлого параметра по всей выборке.
I_{ij} – частный критерий знаний: количество знаний в факте наблюдения i-го значения прошлого параметра о том, что объект перейдет в состояние, соответствующее j-му значению будущего параметра;
Ψ – нормировочный коэффициент (Е.В.Луценко, 2002), преобразующий количество информации в формуле А.Харкевича в биты и обеспечивающий для нее соблюдение принципа соответствия с формулой Р.Хартли;
P_i – безусловная относительная частота встречи i-го значения прошлого параметра в обучающей выборке;
P_{ij} – условная относительная частота встречи i-го значения прошлого параметра при j-м значении будущего параметра.

Когда мы сравниваем условные и безусловные относительные частоты путем вычитания у нас получается частный критерий знаний: «коэффициент взаимосвязи» (СК-модели INF6, INF7), когда же мы сравниваем их путем деления, то у нас получается частный критерий: «количество информации по А.Харкевичу» (СК-модели INF1, INF2).

Таким образом, мы видим, что **все частные критерии знаний тесно взаимосвязаны друг с другом**. Особенно интересна связь знаменитого критерия хи-квадрат Пирсона с замечательной мерой количества информации А.Харкевича и с известным в экономике коэффициентом ROI.

Вероятность рассматривается как предел, к которому стремится относительная частота (отношение количества благоприятных исходов к числу испытаний) при **неограниченном** увеличении количества испытаний. Ясно, что вероятность – это математическая абстракция, которая никогда не встречается на практике (также как и другие математические и физические абстракции, типа математической точки, материальной точки, бесконечно малой и т.п.). На практике встречается только относительная частота. Но она может быть весьма близкой к вероятности. Например, при 480 наблюдений различие между относительной частой и вероятностью (погрешность) составляет около 5%, при 1250 наблюдениях – около 2.5%, при 10000 наблюдениях – 1%.

Суть этих методов в том, что вычисляется количество информации в значении фактора о том, что объект моделирования перейдет под его действием в определенное состояние, соответствующее классу. Это позволяет сопоставимо и корректно обрабатывать разнородную

информацию о наблюдениях объекта моделирования, представленную в различных типах измерительных шкал и различных единицах измерения [1].

На основе системно-когнитивных моделей, представленных в таблице 9 (отличаются частыми критериями, приведенными в таблице 8), решаются задачи идентификации (классификации, распознавания, диагностики, прогнозирования), поддержки принятия решений (обратная задача прогнозирования), а также задача исследования моделируемой предметной области путем исследования ее системно-когнитивной модели [10-64].

Таблица 5 – Матрица системно-когнитивной модели

		Классы					Значимость фактора
		1	...	j	...	W	
Значения факторов	1	I_{11}		I_{1j}		I_{1W}	$\sigma_{1\Sigma}=\sqrt[2]{\frac{1}{W-1}\sum_{j=1}^{W}\left(I_{1j}-\bar{I}_1\right)^2}$
	...						
	i	I_{i1}		I_{ij}		I_{iW}	$\sigma_{i\Sigma}=\sqrt[2]{\frac{1}{W-1}\sum_{j=1}^{W}\left(I_{ij}-\bar{I}_i\right)^2}$
	...						
	M	I_{M1}		I_{Mj}		I_{MW}	$\sigma_{M\Sigma}=\sqrt[2]{\frac{1}{W-1}\sum_{j=1}^{W}\left(I_{Mj}-\bar{I}_M\right)^2}$
Степень редукции класса		$\sigma_{\Sigma 1}$		$\sigma_{\Sigma j}$		$\sigma_{\Sigma W}$	$H=\sqrt[2]{\frac{1}{(W\cdot M-1)}\sum_{j=1}^{W}\sum_{i=1}^{M}\left(I_{ij}-\bar{I}\right)^2}$

Отметим, что как значимость значения фактора, степень детерминированности класса и ценность или качество модели в АСК-анализе рассматривается <u>вариабельность</u> значений частных критериев этого значения фактора, класса или модели в целом (таблица 4).

Численно эта вариабельность может измеряться разными способами, например средним отклонением модулей частных критериев от среднего, дисперсией или среднеквадратичным отклонением или его квадратом. В системе «Эйдос» принят последний вариант, т.к. эта величина совпадает с ***мощностью*** *сигнала, в частности мощностью информации, а в АСК-анализе все модели рассматриваются в как источник информации об объекте моделирования.*

Поэтому есть все основания уточнить традиционную терминологию АСК-анализа (таблица 5):

Таблица 6 – Уточнение терминологии АСК-анализа

№	Традиционные термины (синонимы)	Новый термин	Формула
1	1. Значимость значения фактора (признака). 2. Дифференцирующая мощность значения фактора (признака). 3. Ценность значения фактора (признака) для решения задачи идентификации и других задач	Корень из информационной мощности значения фактора	$\sigma_{i\Sigma} = \sqrt[2]{\frac{1}{W-1}\sum_{j=1}^{W}\left(I_{ij} - \bar{I}_i\right)^2}$
2	1. Степень детерминированности класса. 2. Степень обусловленности класса.	Корень из информационной мощности класса	$\sigma_{\Sigma j} = \sqrt[2]{\frac{1}{M-1}\sum_{i=1}^{M}\left(I_{ij} - \bar{I}_j\right)^2}$
3	1. Качество модели. 2. Ценность модели. 3. Степень сформированности модели. 4. Количественная мера степени выраженности закономерностей в моделируемой предметной области	Корень из информационной мощности модели	$H = \sqrt[2]{\frac{1}{(W \cdot M - 1)}\sum_{j=1}^{W}\sum_{i=1}^{M}\left(I_{ij} - \bar{I}\right)^2}$

12.3.4. Верификация моделей

Оценка достоверности моделей в системе «Эйдос» осуществляется путем решения задачи классификации объектов обучающей выборки по обобщенным образам классов и подсчета количества истинных и ложных положительных и отрицательных решений по F-мере Ван Ризбергена, а также по критериям L1- L2-мерам проф. Е.В.Луценко, которые предложены для того, чтобы смягчить или полностью преодолеть некоторые недостатки F-меры [8].

Достоверность моделей можно оценивать и путем решения других задач, например задач прогнозирования, выработки управляющих решений, исследования объекта моделирования путем исследования его модели. Но это более трудоемко и даже всегда возможно, особенно на экономических и политических моделях.

В режиме 3.4 системы «Эйдос» и ряде других изучается достоверность каждой частной модели в соответствии с этими мерами достоверности.

12.3.5. Выбор наиболее достоверной модели

Все последующие задачи решаются в наиболее достоверной модели.

Причины этого просты. Если модель достоверна, то:

– *идентификация* объекта с классом достоверна, т.е. модель относит объекты к классам, к которым они действительно принадлежат;

– *прогнозирование* достоверно, т.е. действительно наступают те события, которые прогнозируются;

– *принятие решений* адекватно (достоверно), т.е. после реализации принятых управляющих решений объект управления действительно переходит в целевые будущие состояния;

– *исследование* достоверно, т.е. полученные в результате исследования модели объекта моделирования выводы могут быть с полным основанием отнесены к объекту моделирования.

Технически сам выбор наиболее достоверной модели осуществляется в режиме 5.6 системы «Эйдос» и проходит быстро. Это необходимо лишь для решения задачи идентификации и прогнозирования (в режиме 4.1.2), которая требует наибольшие вычислительные ресурсы и поэтому решается только для модели, заданной текущей. Все остальные расчеты проводятся в системе «Эйдос» сразу во всех моделях.

12.3.6. Решение задачи идентификации и прогнозирования

При решении *задачи идентификации* каждый объект распознаваемой выборки сравнивается по всем своим признакам с каждым из обобщенных образов классов. Смысл решения задачи идентификации заключается в том, что при определении принадлежности конкретного объекта к обобщенному образу классу об этом конкретном объекте ***по аналогии*** *становится известно все, что известно об объектах этого класса, по крайней мере, самое существенное о них, т.е. чем они отличаются от объектов других классов.*

Задачи идентификации и прогнозирования взаимосвязаны и мало чем отличаются друг от друга. Главное различие между ними в том, что при идентификации значения свойств и принадлежность объекта к классу относятся к одному моменту времени, а при прогнозировании значения факторов относятся к прошлому, а переход объекта под действием этих факторов в состояние, соответствующее классу относится к будущему.

Задача решается в модели, заданной в качестве текущей, т.к. является весьма трудоемкой в вычислительном отношении. Правда с использованием графического процессора (GPU) для расчетов эта проблема практически снялась.

Сравнение осуществляется путем применения *неметрических интегральных критериев*, которых в настоящее время в системе «Эйдос» используется два. Эти интегральные критерии интересны тем, что корректны[18] в неортонормированных пространствах, которые всегда и встречаются на практике, и являются фильтрами подавления шума.

12.3.6.1. Интегральный критерий «Сумма знаний»

Интегральный критерий «Сумма знаний» представляет собой суммарное количество знаний, содержащееся в системе факторов различной природы, характеризующих сам объект управления,

[18] В отличие от Евклидова расстояния, которое используется для подобных целей наиболее часто

управляющие факторы и окружающую среду, о переходе объекта в будущие целевые или нежелательные состояния.

Интегральный критерий представляет собой аддитивную функцию от частных критериев знаний, представленных в help режима 5.5:

$$I_j = (\vec{I}_{ij}, \vec{L}_i).$$

В выражении круглыми скобками обозначено скалярное произведение. В координатной форме это выражение имеет вид:

$$I_j = \sum_{i=1}^{M} I_{ij} L_i,$$

где: M – количество градаций описательных шкал (признаков);

$\vec{I}_{ij} = \{I_{ij}\}$ – вектор состояния j–го класса;

$\vec{L}_i = \{L_i\}$ – вектор состояния распознаваемого объекта, включающий все виды факторов, характеризующих сам объект, управляющие воздействия и окружающую среду (массив–локатор), т.е.:

$$\vec{L}_i = \begin{cases} 1, \textit{если } i-\textit{й фактор действует}; \\ n, \textit{где}: n > 0, \textit{если } i-\textit{й фактор действует с истинностью } n; \\ 0, \textit{если } i-\textit{й фактор не действует}. \end{cases}$$

В текущей версии системы «Эйдос-X++» значения координат вектора состояния распознаваемого объекта принимались равными либо 0, если признака нет, или n, если он присутствует у объекта с интенсивностью n, т.е. представлен n раз (например, буква «о» в слове «молоко» представлена 3 раза, а буква «м» - один раз).

12.3.6.2. Интегральный критерий «Семантический резонанс знаний»

Интегральный критерий «Семантический резонанс знаний» представляет собой *нормированное* суммарное количество знаний, содержащееся в системе факторов различной природы, характеризующих сам объект управления, управляющие факторы и окружающую среду, о переходе объекта в будущие целевые или нежелательные состояния.

Интегральный критерий представляет собой аддитивную функцию от частных критериев знаний, представленных в help режима 3.3 и имеет вид:

$$I_j = \frac{1}{\sigma_j \sigma_l M} \sum_{i=1}^{M} \left(I_{ij} - \bar{I}_j\right)\left(L_i - \bar{L}\right),$$

где:

M – количество градаций описательных шкал (признаков); $\bar{I}_j$ – средняя информативность по вектору класса; $\overline{L}$ – среднее по вектору объекта;

σ_j – среднеквадратичное отклонение частных критериев знаний вектора класса; σ_l – среднеквадратичное отклонение по вектору распознаваемого объекта.

$\vec{I}_{ij} = \{I_{ij}\}$ – вектор состояния j–го класса; $\vec{L}_i = \{L_i\}$ – вектор состояния распознаваемого объекта (состояния или явления), включающий все виды факторов, характеризующих сам объект, управляющие воздействия и окружающую среду (массив–локатор), т.е.:

$$\vec{L}_i = \begin{cases} 1, \textit{если } i-\textit{й фактор действует}; \\ n, \textit{где}: n>0, \textit{если } i-\textit{й фактор действует с истинностью } n; \\ 0, \textit{если } i-\textit{й фактор не действует}. \end{cases}$$

В текущей версии системы «Эйдос-X++» значения координат вектора состояния распознаваемого объекта принимались равными либо 0, если признака нет, или n, если он присутствует у объекта с интенсивностью n, т.е. представлен n раз (например, буква «о» в слове «молоко» представлена 3 раза, а буква «м» - один раз).

Приведенное выражение для интегрального критерия «Семантический резонанс знаний» получается непосредственно из выражения для критерия «Сумма знаний» после замены координат перемножаемых векторов их стандартизированными значениями: $I_{ij} \to \frac{I_{ij} - \bar{I}_j}{\sigma_j}$, $L_i \to \frac{L_i - \bar{L}}{\sigma_l}$.Поэтому по своей сути он также является скалярным произведением двух стандартизированных (единичных) векторов класса и объекта. Существуют и много других способов нормирования, например, путем применяя сплайнов, в частности линейной интерполяции: $I_{ij} \to \frac{I_{ij} - I_j^{\min}}{I_j^{\max} - I_j^{\min}}$, $L_i \to \frac{L_i - L^{\min}}{L^{\max} - L^{\min}}$, Это позволяет предложить другие виды интегральных критериев. Но они в настоящее время не реализованы в системе «Эйдос».

12.3.6.3. Важные математические свойства интегральных критериев

Данные интегральные критерии обладают очень интересными ***математическими свойствами***, которые обеспечивают ему важные достоинства:

Во-первых, интегральный критерий имеет **неметрическую** природу, т.е. он являются мерой сходства векторов класса и объекта, но не расстоянием между ними, а косинусом угла между ними, т.е. это

межвекторное или информационное расстояние. Поэтому его применение является корректным в **неортонормированных** пространствах, которые, как правило, и встречаются на практике и в которых применение Евклидова расстояния (теоремы Пифагора) является некорректным.

Во-вторых, данный интегральный критерий являются **фильтром**, подавляющим белый **шум**, который всегда присутствует в эмпирических исходных данных и в моделях, созданных на их основе. Это свойство подавлять белый шум проявляется у данного критерия тем ярче, чем больше в модели градаций описательных шкал.

В-третьих, интегральный критерий сходства представляет собой количественную меру сходства/различия конкретного объекта с обобщенным образом класса и имеет тот же смысл, что и **функция принадлежности** элемента множеству в нечеткой логике Лотфи Заде. **Однако** в нечеткой логике эта функция задается исследователем априорно путем выбора из нескольких возможных вариантов, а в АСК-анализе и его программном инструментарии – интеллектуальной системе «Эйдос» она рассчитывается в соответствии с хорошо обоснованной математической моделью непосредственно на основе эмпирических данных.

В-четвертых, кроме того значение интегрального критерия сходства представляет собой адекватную самооценку **степени уверенности** системы в положительном или отрицательном решении о принадлежности/непринадлежности объекта к классу или **риска ошибки** при таком решении.

В-пятых, по сути, при распознавании происходит расчет коэффициентов I_j разложения функции объекта L_i в ряд по функциям классов I_{ij}, т.е. определяется **вес** каждого обобщенного образа класса в образе объекта, что подробнее описано в монографии: Луценко Е. В. Сценарный и спектральный автоматизированный системно-когнитивный анализ: научная монография / Е. В. Луценко. – Краснодар: КубГАУ, 2021. – 288 с., ISBN 978-5-907474-67-3, DOI: 10.13140/RG.2.2.22981.37608, https://www.researchgate.net/publication/353555996

12.3.7. Решение задачи принятия решений

12.3.7.1. Упрощенный вариант принятия решений как обратная задача прогнозирования, позитивный и негативный информационные портреты классов, SWOT-анализ

Задачи прогнозирования и принятия решений относятся друг к другу как прямая и ***обратная*** задачи:

– при прогнозировании по значениям факторов, действующих на объект моделирования, определяется в какое будущее состояние он перейдет под их действием;

– при принятии решений, наоборот, по будущему целевому состоянию объекта моделирования определяются значения факторов, которые обуславливают его переход в это будущее целевое состояние.

Таким образом, задача принятия решений является обратной по отношению к задаче прогнозирования. Но это так только в простейшем случае: в случае использования SWOT-анализа (режим 4.4.8 системы «Эйдос») [9].

12.3.7.2. Развитый алгоритм принятия решений в АСК-анализе

Однако SWOT-анализ (режим 4.4.8 системы «Эйдос») имеет свои ограничения: может быть задано только одно будущее целевое состояние, некоторые рекомендуемые факторы может не быть технологической и финансовой возможности использовать.

Поэтому в АСК-анализе и системе «Эйдос» реализован развитый алгоритм принятия решений (режим 6.3) в котором кроме SWOT-анализа используются также результаты решения задачи прогнозирования и результаты кластерно-конструктивного анализа классов и значений факторов, т.е. некоторые результаты решения задачи исследования предметной области. Этот алгоритм описан в работе [10] и ряде последующих работ.

12.3.8. Решение задачи исследования объекта моделирования путем исследования его модели

12.3.8.1. Инвертированные SWOT-диаграммы значений описательных шкал (семантические потенциалы)

Инвертированные SWOT-диаграмм (предложены автором в работе [9]), отражают силу и направление влияния конкретной градации описательной шкалы на переход объекта моделирования в состояния, соответствующие градациям классификационных шкал (классы). Это и есть ***смысл*** (семантический потенциал) этой градации описательной шкалы. Инвертированные SWOT-диаграммы выводятся в режиме 4.4.9 системы «Эйдос».

12.3.8.2. Кластерно-конструктивный анализ классов

В системе «Эйдос» (в режиме 4.2.2.1) рассчитывается матрица сходства классов по системе их детерминации и на основе этой матрицы рассчитывается и выводится три основных формы:

– круговая 2d-когнитивная диаграмма классов (режим 4.2.2.2);

– агломеративных дендрограмм, полученных в результате ***когнитивной (истинной) кластеризации классов*** (предложена автором в 2011 году в работе [11]) (режим 4.2.2.3);

– график изменения межкластерных расстояний (режим 4.2.2.3);

– 3d-когнитивная диаграмма классов и признаков (режим 4.4.12).

12.3.8.3. Кластерно-конструктивный анализ значений описательных шкал

В системе «Эйдос» (в режиме 4.3.2.1) рассчитывается матрица сходства признаков по системе их смыслу и на основе этой матрицы рассчитывается и выводится три основных формы:

– круговая 2d-когнитивная диаграмма признаков (режим 4.3.2.2);

– агломеративных дендрограмм, полученных в результате ***когнитивной (истинной) кластеризации признаков*** (предложена автором в 2011 году в работе [11]) (режим 4.3.2.3);

– график изменения межкластерных расстояний (режим 4.3.2.3);

– 3d-когнитивная диаграмма классов и признаков (режим 4.4.12).

12.3.8.4. Модель знаний системы «Эйдос» и нелокальные нейроны

Модель знаний системы «Эйдос» относится к ***нечетким декларативным*** гибридным моделям и объединяет в себе некоторые положительные особенности нейросетевой и фреймовой моделей представления знаний.

Классы в этой модели соответствуют нейронам и фреймам, а признаки рецепторам и шпациям (описательные шкалы – слотам).

От фреймовой модели представления знаний модель системы «Эйдос» отличается своей эффективной и простой программной реализацией, полученной за счет того, что разные фреймы отличаются друг от друга не набором слотов и шпаций, а лишь информацией в них. *Поэтому в системе «Эйдос» при увеличении числа фреймов само количество баз данных не увеличивается, а увеличивается лишь их размерность*. Это является очень важным свойством моделей системы «Эйдос», существенно облегчающим и упрощающим программную реализации.

От нейросетевой модели представления знаний модель системы «Эйдос» отличается тем, что [12]:

1) весовые коэффициенты на рецепторах не подбираются итерационным методом обратного распространения ошибки, а рассчитываются методом прямого счета на основе хорошо теоретически обоснованной модели, основанной на ***теории информации*** (это напоминает байесовские сети);

2) весовые коэффициенты имеют хорошо теоретически обоснованную ***содержательную интерпретацию***, основанную на теории информации;

3) нейросеть является ***нелокальной***, как сейчас говорят «полносвязной».

В системе «Эйдос» нелокальные нейроны визуализируются (режим 4.4.10 системы «Эйдос») в виде специальных графических форм, на

которых сила и направление влияния рецепторов нейрона на степень его активации/торможения отображается в форме цвета и толщины дендрита.

12.3.8.5. Нелокальная нейронная сеть

В системе «Эйдос» есть возможность построения моделей, соответствующих многослойным нейронным сетям [12].

Есть также возможность визуализации любого одного слоя нелокальной нейронной сети (режим 4.4.11 системы «Эйдос»).

Такой слой в наглядной форме отражает силу и направление влияния рецепторов ряда нейрона на степень их активации/торможения в форме цвета и толщины дендритов.

Нейроны на изображении слоя нейронной сети расположены слева направо в порядке убывания модуля суммарной силы их детерминации рецепторами, т.е. слева находятся результаты, наиболее жестко обусловленные действующими на них значениями факторов, а справа – менее жестко обусловленные.

12.3.8.6. 3D-интегральные когнитивные карты

3d-интегральная когнитивная карта является отображением на одном рисунке когнитивных диаграмм классов и значений факторов вверху и внизу соответственно и одного слоя нейронной сети (режим 4.4.12 системы «Эйдос»).

12.3.8.7. 2D-интегральные когнитивные карты содержательного сравнения классов (опосредованные нечеткие правдоподобные рассуждения)

В 2d-когнитивных диаграммах сравнения классов по системе их детерминации видно, насколько сходны или насколько отличаются друг от друга классы по значениям обуславливающих их факторов.

Однако мы не видим из этой диаграммы, чем именно конкретно сходны и чем именно отличаются эти классы по значениям обуславливающих их факторов.

Это мы можем увидеть из когнитивной диаграммы содержательного сравнения классов, которая отображается в режиме 4.2.3 системы «Эйдос».

2D-интегральные когнитивные карты содержательного сравнения классов являются примерами опосредованных нечетких правдоподобных логических заключений, о которых может быть первым писал Дьердь Пойа [13]. Впервые об автоматизированной реализации рассуждений подобного типа в интеллектуальной системе «Эйдос» написано в 2002 году в работе [1] на странице 521[19]. Позже об этом писалось в работе [7][20] и ряде других

[19] https://www.elibrary.ru/download/elibrary_18632909_64818704.pdf, Таблица 7. 17, стр. 521

[20] http://ej.kubagro.ru/2013/07/pdf/15.pdf, стр.44.

работ автора, поэтому здесь подробнее рассматривать этот вопрос нецелесообразно.

Например, нам известно, что один человек имеет голубые глаза, а другой черные волосы. Спрашивается, эти признаки вносят вклад в сходство или в различие этих двух людей? В АСК-анализе и системе «Эйдос» этот вопрос решается так. В модели на основе кластерно-конструктивного анализа классов и значений факторов (признаков) известно, насколько те или иные признаки сходны или отличаются по их влиянию на объект моделирования. Поэтому понятно, что человек с голубыми глазами вероятнее всего блондин, а брюнет, скорее всего, имеет темные глаза. Так что понятно, что эти признаки вносят вклад в различие этих двух людей.

12.3.8.8. 2D-интегральные когнитивные карты содержательного сравнения значений факторов (опосредованные нечеткие правдоподобные рассуждения)

Из 2d-когнитивных диаграммах сравнения значений факторов по их влиянию на объект моделирования, т.е. на его переходы в состояния, соответствующие классам вполне понятно, насколько сходны или отличаются любые два значения факторов по их смыслу.

Напомним, что смысл, согласно концепции смысла Шенка-Абельсона, используемой в АСК-анализе, состоит в знании причин и последствий [14].

Однако из этой диаграммы не видно, чем именно ***конкретно*** сходны или отличаются значения факторов по их смыслу.

Это видно из когнитивных диаграмм, которые можно получить в режиме 4.3.3 системы «Эйдос».

12.3.8.9. Когнитивные функции

Когнитивные функции являются обобщением классического математического понятия функции на основе системной теории информации и предложены Е.В.Луценко в 2005 году [7, 15-22].

Когнитивные функции отображают, какое количество информации содержится в градациях описательной шкалы о переходе объекта моделирования в состояния, соответствующие градациям классификационной шкалы. При этом в статистических и системно-когнитивных моделях в каждой градации описательной шкалы содержится информация обо всех градациях классификационной шкалы, т.е. ***каждому значению аргумента соответствуют <u>все</u> значения функции, но соответствуют <u>в разной степени</u>, причем как положительной, так и отрицательной, которая отображается цветом***.

В системе «Эйдос» когнитивные функции отображаются в режиме 4.5.

12.3.8.10. Значимость описательных шкал и их градаций

В АСК-анализе все факторы рассматриваются с одной единственной точки зрения: сколько информации содержится в их значениях о переходе объекта моделирования и управления, на который они действуют, в определенное будущее состояние, описываемое классом (градация классификационной шкалы), и при этом сила и направление влияния всех значений факторов на объект измеряется в одних общих для всех факторов единицах измерения: единицах количества информации [6].

Значимость (селективная сила) градаций описательных шкал в АСК-анализе – это вариабельность частных критериев в статистических и системно-когнитивных моделях, например в модели Inf1, это вариабельность информативностей (режим 3.7.5 системы «Эйдос»)..

Значимость всей описательной шкалы является средним от степени значимости ее градаций (режим 3.7.4 системы «Эйдос»).

Если рассортировать все градации факторов (признаки) в порядке убывания селективной силы и получить сумму селективной силы системы значений факторов нарастающим итогом, то получим Парето-кривую.

12.3.8.11. Степень детерминированности классов и классификационных шкал

Степень детерминированности (обусловленности) класса в системе «Эйдос» количественно оценивается ***степенью вариабельности значений факторов*** (градаций описательных шкал) в колонке матрицы модели, соответствующей данному классу.

Чем выше степень детерминированности класса, тем более достоверно он прогнозируется по значениям факторов.

Степень детерминированности (обусловленности) всей классификационной шкалы является средним от степени детерминированности ее градаций, т.е. классов (режим 3.7.2 системы «Эйдос»).

ГЛАВА 13. СЦЕНАРНЫЙ АВТОМАТИЗИРОВАННЫЙ СИСТЕМНО-КОГНИТИВНЫЙ АНАЛИЗ

13.1. Объект, предмет, проблема, цель, метод и задачи исследования

Объектом исследования в данной работе является фундаментальная теорема А.Н.Колмогорова (1957) [1]:

«Т е о р е м а. При любом целом $n \geq 2$ *существуют такие определенные на единичном отрезке* $E^1 = [0;1]$ *непрерывные действительные функции* $\psi^{pq}(x)$*, что каждая определенная на n-мерном единичном кубе* E^n *непрерывная действительная функция* $f(x_1,...,x_n)$ *представима в виде:*

$$f(x_1,...,x_n) = \sum_{q=1}^{q=2n+1}\left(\chi_q\left[\sum_{p=1}^{n}\psi^{pq}(x_p)\right]\right), \qquad (1)$$

где функции $\chi_q(y)$ *действительны и непрерывны.»* [1].

Эта замечательная фундаментальная теорема означает, что для реализации функций многих переменных достаточно операций взвешенного суммирования (суперпозиции) функций одной переменной.

Последствия этого очень важны и многочисленны и относятся не только к математике, где теорема А.Н.Колмогорова связана, например, с решением 13-й проблемы Гильберта, но и ко многим другим направлениям науки, например, к интеллектуальным технологиям [2].

Но в данной работе для нас важнее, что теорема А.Н.Колмогорова [1], по мнению автора, фактически является теоретическим фундаментом всей математической теории разложения функций в ряды, т.е. так называемой теории рядов [3].

Чтобы убедиться в этом предлагается рассмотреть *частный* случай теоремы А.Н.Колмогорова, который получается из (1) путем замены функции $\chi_q(y)$ на *частный* случай этой функции, когда она равна собственному аргументу, умноженному на некоторую константу g_q (2).

$$\chi_q\left[\sum_{p=1}^{n}\psi^{pq}(x_p)\right] \Rightarrow g_q\sum_{p=1}^{n}\psi^{pq}(x_p) \qquad (2)$$

Отметим, что подобный подход не раз применялся для исследования различных вариантов теоремы А.Н.Колмогорова *для конкретных видов функций* [4, 5, 6] и в нем ничего необычного.

С учетом (2) выражение (1) примет вид:

$$f(x_1,...,x_n)=\sum_{q=1}^{\infty}\left(g_q\sum_{p=1}^{n}\psi^{pq}\left(x_p\right)\right) \quad (3)$$

Кроме того, в выражении (3) как принято в теории рядов верхний предел суммирования в первой сумме заменен на бесконечность.

Выражение (3) является частным случаем выражения (1), которое строго математически доказано, поэтому выражение (3) тоже можно считать строго математически доказанным.

В терминологии теории рядов константу g_q в выражении (3) естественно интерпретировать как весовые коэффициенты ряда, а функции $\psi^{pq}\left(x_p\right)$ – как базисные функции, по которым производится разложение в ряд функции $f(x_1,...,x_n)$.

Однако, определение вида базисных функций $\psi^{pq}\left(x_p\right)$ и весовых коэффициентов g_q для данной конкретной функции $f(x_1,...,x_n)$ представляет собой *математическую* **проблему**, *для которой пока не найдено общего математически строго решения*. При этом для частных случаев, т.е. конкретных видов базисных функций и весовых коэффициентов, таких решений найдено довольно много.

Предметом исследования является математическая модель автоматизированного системно-когнитивного анализа (АСК-анализа), которая рассматривается как один из возможных вариантов ***общего и универсального практического решения проблемы*** разработки базисных функций и весовых коэффициентов для разложения в ряд по ним произвольной функции. В этом контексте функция $f(x_1,...,x_n)$ интерпретируется как конкретный образ состояния идентифицируемого объекта или ситуации, функции $\psi^{pq}\left(x_p\right)$ – как обобщенные образы классов, а функция g_q – как меры сходства конкретного образа объекта или ситуации с обобщенным образом *q-го* класса.

Предлагаемый путь решения проблемы. На взгляд автора *источником или причиной существования* поставленной **проблемы** является то, что в математической теории рядов считается, что для разложения функций в ряд должна использоваться полная ортогональная система базисных функций.

Справка: «ПОЛНАЯ СИСТЕМА ФУНКЦИЙ в некотором линейном пространстве функций L — система функций $\{\varphi(x)\}$ такая, что в L не существует ненулевой функции, ортогональной всем функциям семейства (см. Ортогональные функции) в смысле определенного в L скалярного произведения. Если в L существует ***полная ортонормированная система***

функций, то любую функцию из *L* можно разложить в ряд по функциям этой системы.»[21]

О смысловой связи теоремы А.Н.Колмогорова и АСК-анализа.

Прежде всего необходимо отметить, что теорема А.Н.Колмогорова является фундаментальной математической теоремой, безупречно строго доказанной математически для *действительных и непрерывных функций.* Эта теорема имеет очень высокий статус в математике и играет большую роль в перспективных исследованиях, «*поскольку она, как путеводная звезда, указывает путь*» (профессор А.Н.Орлов[22]).

Математическая модель АСК-анализа (Е.В.Луценко, 1979) является дискретной (численной) моделью. Говоря строго математически она ниоткуда строго не выведена, но имеет эвристический правдоподобный характер [8]. В тоже время обоснованию и описанию применений этой модели посвящено много работ автора с соавторами [9-38].

Но, по мнению автора, между математической моделью АСК-анализа и теоремой А.Н.Колмогорова (по крайней мере с ее частным случаем (3)) существует определенная смысловая связь, хотя и недоказанная строго математически. И данная работа посвящена не доказательству этой смысловой взаимосвязи, а ее использованию для развития сценарного АСК-анализа и решения с его применением новых задач, интересных для науки и для практики.

По этому поводу профессор А.И.Орлов в частной переписке по поводу данной работы пишет: «Реальные расчеты в АСК-анализе проводятся по формулам, которые на сегодня не выведены из теоремы А.Н.Колмогорова, поскольку соответствующие предельные теоремы пока не получены (и получить их, возможно, трудно). Связь между этими результатами идейная, но не математическая. Тем не менее полезно отметить эту связь. Я думаю, что Ваши подходы и алгоритмы переходят в подходы А.Н.Колмогорова при соответствующем предельном переходе (при переходе от дискретности к непрерывности при уменьшении разностей между ближайшими значениями переменных). Математически строгих **формулировок** и тем более доказательств этого на сегодня нет, и получить их, видимо, весьма сложно. Таким образом выявились новые математические проблемы – есть чем заняться будущим поколениям исследователей». Сказано исчерпывающе.

Система обобщенных образов классов в АСК-анализе в общем случае не является полной ортогональной системой функций. Тем ни менее предлагается использовать эту систему функций для разложения в ряд функции, описывающей состояние объекта или ситуации. Это обеспечивает ***практическое*** решение поставленной проблемы, а также

[21] См., например: http://dict.scask.ru/index.php?id=1259

[22] http://orlovs.pp.ru, http://ej.kubagro.ru/a/viewaut.asp?id=2744

решение на этой основе ряда задач, представляющих большой научный и практический интерес.

В частности, предлагается интерпретировать операцию разложения функции, описывающей состояние объекта или ситуации в ряд по функциям обобщенных образов классов как решение задачи идентификации или прогнозирования. При прогнозировании текущая ситуация сравнивается с этими обобщенными образами и разлагается в ряд по ним (прямое преобразование, объектный анализ). Средневзвешенный прогноз формируется путем обратного преобразования образов классов с их весами, т.е. как их взвешенная суперпозиция. При этом в качестве базисных функций используются обобщенные образы прогнозируемых сценариев того что будет и того что не будет с их весами, в качестве которых используется достоверность прогноза.

Кроме того, созданную модель можно использовать для решения и других задач, таких как принятие решений (обратная задача прогнозирования) и исследование моделируемой предметной области путем исследования ее модели.

АСК-анализ предоставляет математический метод формирования системы базисных функций обобщенных образов классов и весовых коэффициентов для разложения в ряд функции состояния объекта или ситуации на основе непосредственно эмпирических данных.

Более того, АСК-анализ имеет свой программный инструментарий, в качестве которого в настоящее время выступает интеллектуальная система «Эйдос» (открытое программное обеспечение), которая реализует этот математический метод.

Целью данной работы является решение поставленной проблемы путем обобщения теории рядов с применением теории информации и разработки реализующего этот подход программного инструментария.

Рассмотрим теоретическое решение поставленной проблемы на уровне математической модели, а затем подробный численный пример практического решения проблемы с применением специально разработанного для этой цели программного инструментария.

В качестве ***примеров*** применения предлагаемых подходов рассматриваются технический, фундаментальный и техно-фундаментальный сценарный АСК-анализ. В этих примерах на основе анализа ретроспективных исходных данных выявляются фактически наблюдавшиеся прошлые и будущие сценарии развития событий. Путем их обобщения формируются образы будущих сценариев развития событий, которые рассматриваются как базисные функции классов. Будущие сценарии обуславливаются прошлыми сценариями развития событий (значениями факторов).

13.2. Теоретическое решение проблемы исследования

13.2.1. Суть математической модели классического АСК-анализа

13.2.1.1. Способ формализации предметной области в АСК-анализе, классификационные и описательные шкалы и градации и обучающая выборка

Формализация предметной области – это такое ее описание, которое пригодно для обработки на компьютере. Этот процесс состоит в том, что создаются классификационные и описательные шкалы и градации, а затем с их помощью кодируются исходные данные и таким образом формируется обучающая (тренировочная) выборка. По сути формализация предметной области повышает степень формализации ее описания путем нормализации исходных данных, до уровня, достаточного для обработки на компьютере.

В АСК-анализе и системе «Эйдос» для формализации предметной области используются различные автоматизированные программные интерфейсы (API), которых довольно много. Это различные интерфейсы с текстовыми, табличными и графическими данными. В результате формируются классификационные и описательные шкалы и градации, которые могут быть различных типов [11]: числовыми и текстовыми, причем текстовые могут быть номинальными и порядковыми.

В АСК-анализе используется 3 способа ***интерпретации смысла классификационных и описательных шкал и градаций***:

– *1-й статическая интерпретация,* когда градации классификационных шкал, т.е. классы, соответствуют обобщенным категориям объектов, а градации описательных шкал рассматриваются как признаки объектов, т.е. наличие или степень выраженности у них определенных физических, социальных и других свойств;

– *2-динамическая интерпретация,* когда градации классификационных шкал, т.е. классы, соответствуют будущим состояниям объекта моделирования в которые он переходит под действием различных факторов, а градации описательных шкал рассматриваются как значения факторов, влияющих на поведение объекта моделирования;

– *3-универсальная интерпретация,* когда не уточняется статическая или динамическая интерпретация используется, а используются термины: «Классификационная шкала», «Градация классификационной шкалы (класс)» «Описательная шкала», «Градация описательной шкалы» .

В нашем случае больше подходит динамическая интерпретация, поэтому и будем пользоваться преимущественно соответствующей терминологией, иногда для уточнения смысла используя термины из других интерпретаций.

Рассмотрим ***принцип формирования описательных шкал и градаций*** (факторы и их значения). Каждому фактору (описательной шкале) соответствует свой диапазон изменения значений аргумента. Каждому значению аргумента соответствует градация шкалы. У разных шкал может быть различное количество градаций.

Описательные шкалы (факторы)	**Градации описательных шкал (значения факторов)**
1-й фактор	$X_{1min}=1 <= X_1<=X_{1max}$
2-й фактор	$X_{2min}= X_{1max} <= X_2<=X_{2max}$
...	...
i-й фактор	$Xi_{min} <= Xi<=Xi_{max}$
...	...
n-й фактор	$Xn_{min} <= Xn<=Xn_{max}=M$

Если шкала числовая, то ее градации представляют собой числовые диапазоны. У каждого числового диапазона есть границы (наименьшее и наибольшее значения) и среднее значение.

Если шкала текстовая, то ее градациями являются уникальные текстовые значения, соответствующие этой шкале.

Если текстовая шкала порядковая, то при сортировке по алфавиту ее градации располагаются в правильном смысловом порядке от минимального значения до максимального, например:

1/5-минимальное значение;

2/5 малое значение;

3/5-среднее значение;

4/5-большое значение;

5/5-максимальное значение.

Если такого осмысленного прядка градаций текстовой шкалы при их сортировке по алфавиту не получается, то значит это текстовая шкала номинального типа.

Совершенно аналогично строятся и классификационные шкалы, и градации. Поэтому это нет особого смысла подробно описывать. Но если градация описательной шкалы является значением фактора, то градация классификационной шкалы представляет собой класс. Обычно классы соответствуют либо обобщенным категориям объектов, либо результатам действия факторов, т.е. описывают результирующие состояния системы, в которые она переходит под действием факторов.

Если исходные данные представлены в табличном виде, то каждой шкале обычно соответствует колонка или строка этой таблицы (чаще колонка).

Отметим, что каждый фактор (описательную шкалу) можно рассматривать как ось в некотором многомерном пространстве. Понятно, что в общем случае это пространство неортонормированное, т.е. факторы зависят друг от друга. Остается также открытым вопрос о метрике и

топологии этого пространства, т.е. о том, имеет ли оно кривизну, какую меру расстояния на нем корректно использовать, к какому классу топологических структур относится топология этого пространства. Все эти вопросы требуют дополнительных исследований. Автор на всякий случай использует в системно-когнитивных моделях информационную меру расстояния между двумя векторами (межсекторное расстояние), корректное для неортонормированных пространств.

13.2.1.2. Синтез системно-когнитивных моделей как разработка обобщенных базисных функций классов путем многопараметрической типизации функций состояний конкретных объектов или ситуаций моделирования

Математическая модель АСК-анализа и системы «Эйдос» основана на системной нечеткой интервальной математике [9, 10, 11] и обеспечивает сопоставимую обработку больших объемов фрагментированных и зашумленных взаимозависимых данных, представленных в различных типах шкал (номинальных, порядковых и числовых) и различных единицах измерения [11].

Суть математической модели АСК-анализа состоит в следующем. Непосредственно на основе эмпирических данных, после их формализации, как описано в предыдущем разделе, рассчитывается матрица абсолютных частот (матрица сопряженности) (таблица 1).

А этой таблице строки соответствуют градациям описательных шкал (значениям факторов), а колонки соответствуют классам, т.е. градациям классификационных шкал. На их пересечении находится число случаев ***наблюдения*** определенного значения признака у объектов определенного класса. Наблюдение определенного признака у объекта определенного класса является ***фактом***. Также фактом является наблюдении перехода объекта моделирования в определенное будущее состояние, если на него действовало определенное значение некоторого фактора. Это означает, что для установления факта необходимо получить информацию о признаках объекта, создать на ее основе конкретный образ объекта и идентифицировать этот конкретный образ, т.е. сравнить его с обобщенными образами и определить степень их сходства, т.е. выполнить довольно много достаточно сложных, даже интеллектуальных операций.

Таким образом, понятие факта не является таким уж простым и элементарным, скорее наоборот. Подробнее о сложности установления фактов можно почитать в работе [34].

Таблица 7– Матрица абсолютных частот (статистическая модель ABS)

	Описательные шкалы (факторы)	Градации описательных шкал (значения факторов)	Классы: 1	...	j	...	W	Сумма
Описательные шкалы и градации (факторы и их значения)	1-й фактор	X_{1min}=1	N_{11}		N_{1j}		N_{1W}	
		…						
		X_{1max}						
	2-й фактор	X_{2min}						
		…						
		X_{2max}						
	…	…						
	i-й фактор	X_{imin}						
		…						
		X_i	N_{i1}		N_{ij}		N_{iW}	$N_{i\Sigma}=\sum_{j=1}^{W} N_{ij}$
		…						
		X_{imax}						
	…	…						
	n-й фактор	X_{nmin}						
		...						
		X_{nmax}=M	N_{M1}		N_{Mj}		N_{MW}	
Суммарное количество признаков по классу					$N_{\Sigma j}=\sum_{i=1}^{M} N_{ij}$			$N_{\Sigma\Sigma}=\sum_{i=1}^{W}\sum_{j=1}^{M} N_{ij}$
Суммарное количество объектов обучающей выборки по классу					$N_{\Sigma j}$			$N_{\Sigma\Sigma}=\sum_{j=1}^{W} N_{\Sigma j}$

На основе таблицы 1 рассчитываются матрицы условных и безусловных процентных распределений (таблица 2).

Здесь необходимо дать пояснение по поводу того, чем являются значения в таблице 2: относительными частотами, процентами или вероятностями. Вообще-то они являются относительными частотами, но выраженными в процентах. Причем за 100% принимается либо «Суммарное количество признаков по классу», либо «Суммарное количество объектов обучающей выборки по классу» из таблицы 1. В результате и получается две модели: PRC1 и PRC2, которые отличаются только этим.

Проценты используются исключительно для удобства восприятия результатов и более эффективного использования разрядной сетке при отображении результатов. Понятно, что вероятность есть ***предел***, к которому ***асимптотически***, т.е. никогда его не достигая, стремится относительная частота при ***бесконечном*** (неограниченном) увеличении объема выборки.

Таблица 8 – Матрица условных и безусловных процентных распределений (статистические модели PRC1 и PRC2)

	Описательные шкалы (факторы)	Градации описательных шкал (значения факторов)	Классы: 1	...	j	...	W	Безусловная вероятность признака
Описательные шкалы и градации (факторы и их значения)	1-й фактор	X_{1min}=1	P_{11}		P_{1j}		P_{1W}	
		…						
		X_{1max}						
	2-й фактор	X_{2min}						
		…						
		X_{2max}						
	…	…						
	i-й фактор	X_{imin}						
		…						
		X_i	P_{i1}		$P_{ij} = \frac{N_{ij}}{N_{\Sigma j}}$		P_{iW}	$P_{i\Sigma} = \frac{N_{i\Sigma}}{N_{\Sigma\Sigma}}$
		…						
		X_{imax}						
	…	…						
	n-й фактор	X_{nmin}						
		...						
		X_{nmax} =M	P_{M1}		P_{Mj}		P_{MW}	
Безусловная вероятность класса					$P_{\Sigma j}$			

Поэтому, конечно, строго говоря, в таблице 2 приведены не вероятности. Но при увеличении объема выборки относительные частоты, в приведенные в таблице 2, все меньше и меньше отличаются от вероятностей. Таким образом называя их вероятностями мы допускаем некоторую неточность или погрешность в наших высказываниях. Но автор считает, что для практических целей это допустимо, учитывая, что при больших выборках эта погрешность и неточность очень мала. Тем более, что мы довольно редко изрекаем абсолютные истины и чаще всего в наших высказываниях есть неточности и погрешности. Допуская эту небольшую вольность мы поступаем точно так же, т.е. следуя той же ***традиции***, что и ученые, которые используют на практике другие математические абстракции, типа математической и материальной точки, бесконечно малых, линий, окружностей и треугольников и т.д. и т.п. [10]. А *так поступают абсолютно все ученые и не ученые*, хотя мнение последних для нас сейчас и не так важно. Например, когда ученый говорит, что у автомобиля колесо круглое, то он конечно не имеет в виду, что оно абсолютно точно соответствует математическому понятию: «Круг» или «Окружность». Совершено ясно, что колесо соответствует

этим строгим математическим понятиям весьма приблизительно, а часто и вообще не очень соответствует, поэтому его и отдают на балансировку.

Затем на основе таблицы 2 или непосредственно таблицы 1 с использованием частных критериев, знаний приведенных таблице 3, рассчитываются матрицы системно-когнитивных моделей (таблица 4).

Таблица 9 – Различные аналитические формы частных критериев знаний

Наименование модели знаний и частный критерий	Выражение для частного критерия	
	через относительные частоты	через абсолютные частоты
ABS, матрица абсолютных частот	---	N_{ij}
PRC1, матрица условных и безусловных процентных распределений, в качестве $N_{\Sigma j}$ используется суммарное количество признаков по классу	---	$P_{ij} = \frac{N_{ij}}{N_{\Sigma j}}$
INF1, частный критерий: количество знаний по А.Харкевичу, 1-й вариант расчета вероятностей: *Nj* – суммарное количество признаков по *j*-му классу. Вероятность того, что если у объекта j-го класса обнаружен признак, то это i-й признак	$I_{ij} = \Psi \times Log_2 \frac{P_{ij}}{P_i}$	$I_{ij} = \Psi \times Log_2 \frac{N_{ij} N}{N_i N_j}$
INF3, частный критерий: **Хи-квадрат:** разности между фактическими и теоретически ожидаемыми абсолютными частотами	---	$I_{ij} = N_{ij} - \frac{N_i N_j}{N}$
INF4, частный критерий: ROI - Return On Investment, 1-й вариант расчета вероятностей: *Nj* – суммарное количество признаков по *j*-му классу	$I_{ij} = \frac{P_{ij}}{P_i} - 1 = \frac{P_{ij} - P_i}{P_i}$	$I_{ij} = \frac{N_{ij} N}{N_i N_j} - 1$
INF6, частный критерий: разность условной и безусловной вероятностей, 1-й вариант расчета вероятностей: *Nj* – суммарное количество признаков по *j*-му классу	$I_{ij} = P_{ij} - P_i$	$I_{ij} = \frac{N_{ij}}{N_j} - \frac{N_i}{N}$

Обозначения к таблице 4:

i – значение прошлого параметра;

j - значение будущего параметра;

N_{ij} – количество встреч j-го значения будущего параметра при i-м значении прошлого параметра;

M – суммарное число значений всех прошлых параметров;

W - суммарное число значений всех будущих параметров.

N_i – количество встреч i-м значения прошлого параметра по всей выборке;

N_j – количество встреч j-го значения будущего параметра по всей выборке;

N – количество встреч j-го значения будущего параметра при i-м значении прошлого параметра по всей выборке.

I_{ij} – частный критерий знаний: количество знаний в факте наблюдения i-го значения прошлого параметра о том, что объект перейдет в состояние, соответствующее j-му значению будущего параметра;

Ψ – нормировочный коэффициент (Е.В.Луценко, 2002), преобразующий количество информации в формуле А.Харкевича в биты и обеспечивающий для нее соблюдение принципа соответствия с формулой Р.Хартли;

P_i – безусловная относительная частота встречи i-го значения прошлого параметра в обучающей выборке;

P_{ij} – условная относительная частота встречи i-го значения прошлого параметра при j-м значении будущего параметра .

Таблица 10 – Матрица системно-когнитивной модели (СК-модель)

Описательные шкалы и градации (факторы и их значения)	Описательные шкалы (факторы)	Градации описательных шкал (значения факторов)	Классы: 1	...	j	...	W	Значимость значений факторов
	1-й фактор	X_{1min}=1	I_{11}		I_{1j}		I_{1W}	$\sigma_{1\Sigma}=\sqrt[2]{\frac{1}{W-1}\sum_{j=1}^{W}\left(I_{1j}-\bar{I}_1\right)^2}$
		…						
		X_{1max}						
	2-й фактор	X_{2min}						
		…						
		X_{2max}						
	…	…						
	i-й фактор	X_{imin}						
		…						
		X_i	I_{i1}		I_{ij}		I_{iW}	$\sigma_{i\Sigma}=\sqrt[2]{\frac{1}{W-1}\sum_{j=1}^{W}\left(I_{ij}-\bar{I}_i\right)^2}$
		…						
		X_{imax}						
	…	…						
	n-й фактор	X_{nmin}						
		...						
		X_{nmax}=M	I_{M1}		I_{Mj}		I_{MW}	$\sigma_{M\Sigma}=\sqrt[2]{\frac{1}{W-1}\sum_{j=1}^{W}\left(I_{Mj}-\bar{I}_M\right)^2}$
Степень редукции класса			$\sigma_{\Sigma 1}$		$\sigma_{\Sigma j}$		$\sigma_{\Sigma W}$	$H=\sqrt[2]{\frac{1}{(W\cdot M-1)}\sum_{j=1}^{W}\sum_{i=1}^{M}\left(I_{ij}-\bar{I}\right)^2}$

Отметим, что в АСК-анализе и его программном инструментарии интеллектуальной системе «Эйдос» используется два способа расчета матриц условных и безусловных процентных распределений (таблица 2):

1-й способ: в качестве $N_{\Sigma j}$ используется суммарное количество признаков по классу;

2-й способ: в качестве $N_{\Sigma j}$ используется суммарное количество объектов обучающей выборки по классу.

Поэтому в АСК-анализе и системе «Эйдос» есть модели, аналогичные PRC1, INF1, INF5 и INF6, в которых относительные частоты рассчитываются по тем же формулам, как в этих моделях, но не 1-м, а 2-м способом. Это модели: PRC2, INF2, INF4 и INF7 соответственно.

Суть этих методов в том, что вычисляется количество информации в значении фактора о том, что объект моделирования перейдет под его действием в определенное состояние, соответствующее классу. Это позволяет сопоставимо и корректно обрабатывать разнородную информацию о наблюдениях объекта моделирования, представленную в

различных типах измерительных шкал и различных единицах измерения [11].

На основе системно-когнитивных моделей, представленных в таблице 4 (отличаются частыми критериями, приведенными в таблице 3), решаются задачи идентификации (классификации, распознавания, диагностики, прогнозирования), поддержки принятия решений (обратная задача прогнозирования), а также задача исследования моделируемой предметной области путем исследования ее системно-когнитивной модели. В качестве развернутого *методически детально проработанного* примера полного исследования с применением АСК-анализа и системы «Эйдос» можно рассматривать главу 4 в работе [21]. По этому методическому образцу оформлена и 3-я часть данной работы.

Таким образом в ***классическом АСК-анализе:***

1. В качестве прошлых значений факторов, влияющих на поведение объекта моделирования, рассматриваются сценарии изменения значений этих факторов. В качестве результата влияния факторов рассматривается сценарии поведения объекта моделирования под влиянием этих факторов.

2. На основе анализа исходных данных выявляются ранее наблюдавшиеся сценарии изменения значений факторов, влияющих на объект моделирования, и сценарии поведения объекта моделирования под влиянием этих значений факторов.

3. Путем обобщения (многопараметрической типизации) конкретных сценариев поведения объекта моделирования формируются обобщенные образы сценариев развития событий (классы) под влиянием сценариев изменения значений факторов.

Математической моделью класса является вектор частных критериев, соответствующий колонке из таблицы 4. Сами частные критерии, используемые в текущей версии системы «Эйдос», приведены в таблице 3.

13.2.1.3. Прогнозирование и системная идентификация как разложение функции ситуации (объекта) в ряд по функциям классов (объектный анализ)

Как влияет на поведение объекта моделирования одно значение фактора, отражено в системно-когнитивных моделях. Как влияет система значений факторов, определяется с помощью интегральных критериев. В интегральном критерии используется система частных критериев и их значения сводятся к одному значению интегрального критерия. Поэтому вычисление значений интегрального критерия сходства объекта распознаваемой (ее еще называют тестовой) выборки с обобщенными образами всех классов называется **системной идентификацией.**

В настоящее время в системе «Эйдос» используется два ***аддитивных*** интегральных критерия:

– сумма знаний;

– резонанс знаний.

1-й интегральный критерий «Сумма знаний» представляет собой суммарное количество знаний, содержащееся в системе значений факторов различной природы, характеризующих сам объект управления, управляющие факторы и окружающую среду, о переходе объекта в будущие целевые или нежелательные состояния.

Интегральный критерий представляет собой аддитивную функцию от частных критериев знаний:

$$I_j = (\vec{I}_{ij}, \vec{L}_i).$$

В выражении круглыми скобками обозначено скалярное произведение. В координатной форме это выражение имеет вид:

$$I_j = \sum_{i=1}^{M} I_{ij} L_i,$$

где: M – количество градаций описательных шкал (признаков);

$\vec{I}_{ij} = \{I_{ij}\}$ – вектор состояния j–го класса;

$\vec{L}_i = \{L_i\}$ – функция состояния (вектор) распознаваемого объекта, включающий все виды факторов, характеризующих сам объект, управляющие воздействия и окружающую среду (массив–локатор), т.е.:

$$\vec{L}_i = \begin{cases} 1, \textit{если } i-\textit{й фактор действует}; \\ n, \textit{где}: n > 0, \textit{если } i-\textit{й фактор действует с истинностью } n; \\ 0, \textit{если } i-\textit{й фактор не действует}. \end{cases}$$

В текущей версии системы «Эйдос» значения координат вектора состояния распознаваемого объекта принимались равными либо 0, если признака нет, или n, если он присутствует у объекта с интенсивностью n, т.е. представлен n раз (например, буква «о» в слове «молоко» представлена 3 раза, а буква «м» – один раз).

Если представить информацию распознаваемой выборки в виде матрицы, в которой каждая строка будет описывать один объект распознаваемой выборки, то *операцию распознавания этой выборки с помощью 1-го интегрального критерия можно представить себе как операцию умножения матрицы распознаваемой выборки на матрицу статистической или системно-когнитивной модели*. Результатом является матрица произведения, в которой каждый элемент является суммой произведений элементов соответствующих строки распознаваемой матрицы и столбца модели.

2-й интегральный критерий «Семантический резонанс знаний» представляет собой нормированное суммарное количество знаний, содержащееся в системе факторов различной природы, характеризующих

сам объект управления, управляющие факторы и окружающую среду, о переходе объекта в будущие целевые или нежелательные состояния.

Интегральный критерий представляет собой аддитивную функцию от частных критериев знаний и имеет вид:

$$I_j = \frac{1}{\sigma_j \sigma_l M} \sum_{i=1}^{M} \left(I_{ij} - \bar{I}_j\right)\left(L_i - \bar{L}\right),$$

где:

σ_j – среднеквадратичное отклонение частных критериев знаний вектора класса;

σ_l – среднеквадратичное отклонение по вектору распознаваемого объекта.

Свое наименование интегральный критерий сходства «Семантический резонанс знаний» получил потому, что по своей математической форме является корреляцией двух векторов: состояния j–го класса и состояния распознаваемого объекта.

По своему смыслу интегральные критерии количественно отражают степень сходства идентифицируемого состояния объекта моделирования с обобщенными образами классов, т.е. по сути степень «присутствия» обобщенного образа класса в этом идентифицируемом состоянии объекта.

Все это позволяет обоснованно рассматривать функцию описания идентифицируемых объектов как взвешенную суперпозицию обобщенных образов классов различного типа с различными амплитудами. По сути это позволяет рассматривать процесс идентификации или прогнозирования состояния объекта как разложение его конкретного образа (функции, описывающей его состояние) в ряд по обобщенным образам классов [12, 13].

Таким образом, в предложенной семантической информационной модели при идентификации и прогнозировании, по сути, осуществляется разложение векторов идентифицируемых объектов по векторам классов распознавания, т.е. осуществляется ***"объектный анализ"*** *(по аналогии с спектральным, гармоническим или* ***Фурье-анализом****),* ***что позволяет рассматривать идентифицируемые объекты как взвешенную суперпозицию обобщенных образов классов различного типа с различными амплитудами [12]. При этом вектора обобщенных образов классов, с математической точки зрения, представляют собой произвольные функции и не обязательно образуют полную (необходимую и достаточную) и не избыточную (ортонормированную) систему функций.***

Впервые эта мысль была высказана автором в 1999 году[23] работе [12][24] в разделе 5.7. Распознавание как объектный анализ (разложение в ряд по профилям образов), а затем развита в ряде работ, в частности в [13].

Таким образом, в данной работе предлагается рассматривать предлагаемую математическую модель АСК-анализа как вариант общего и универсального практического решения проблемы разработки базисных функций и весовых коэффициентов для разложения в ряд по ним функции состояния идентифицируемого объекта.

В этом контексте функция $f(x_1,...,x_n)$ ***интерпретируется как конкретный образ состояния идентифицируемого объекта, функция*** $\psi^{pq}(x_p)$ ***– обобщенный образ q-го класса, а функция*** g_q ***– мера сходства конкретного образа объекта с обобщенным образом класса.***

Отметим также, что между мультипликативными и аддитивными интегральными критериями сходства нет принципиального различия, т.к. логарифм от мультипликативного интегрального критерия представляет собой аддитивный интегральный критерий, в котором логарифмы сомножителей мультипликативного интегрального критерия представляют собой слагаемые аддитивного интегрального критерия.

13.2.1.4. Математические определения основных понятий АСК-анализа, связанных с теоремой А.Н.Колмогорова

Дадим более строгие математические определения базовым понятиям АСК-анализа, которые были использованы выше на интуитивном уровне понимания. Это следующие понятия: ***конкретный образ состояния идентифицируемого объекта, функция*** $\psi^{pq}(x_p)$ ***– обобщенный образ q-го класса, а функция*** g_q ***– мера сходства конкретного образа объекта с обобщенным образом класса.***

Конкретный образ состояния идентифицируемого объекта или ситуации – в АСК-анализе это массив (вектор, функция) $\vec{L}_i = \{L_i\}$ – функция состояния (вектор) распознаваемого объекта, включающий все виды факторов, характеризующих сам объект, управляющие воздействия и окружающую среду (массив–локатор), т.е.:

$$\vec{L}_i = \begin{cases} 1, \textit{если } i-\textit{й фактор действует}; \\ n, \textit{где}: n>0, \textit{если } i-\textit{й фактор действует с истинностью } n; \\ 0, \textit{если } i-\textit{й фактор не действует}. \end{cases}$$

[23] Фактически реализована в математической модели эта мысль была еще в 1979 году, а в системе «Эйдос» изначально, например

[24] См., например: http://lc.kubagro.ru/aidos/aidos99/index.htm

В текущей версии системы «Эйдос» значения координат вектора состояния распознаваемого объекта принимались равными либо 0, если признака нет, или n, если он присутствует у объекта с интенсивностью n, т.е. представлен n раз (например, буква «о» в слове «молоко» представлена 3 раза, а буква «м» – один раз).

В теореме А.Н.Колмогорова (3) этому соответствует функция:

$$f(x_1,...,x_n).$$

Обобщенный образ j-го класса – это $\vec{I}_{ij}=\{I_{ij}\}$ – вектор состояния j–го класса; представляет собой колонку таблицы 4, соответствующую j-му классу. В теореме А.Н.Колмогорова (3) этому соответствует функция: $\psi^{pq}(x_p)$.

Функция g_q – мера сходства конкретного образа объекта с обобщенным образом q-го класса – это один из ***аддитивных*** интегральных критериев сходства, используемых в настоящее время в АСК-анализе и системе «Эйдос» (приведены в предыдущем разделе):

– сумма знаний:

$$I_j=\sum_{i=1}^{M} I_{ij}L_i,$$

– резонанс знаний:

$$I_j=\frac{1}{\sigma_j\sigma_l M}\sum_{i=1}^{M}\left(I_{ij}-\bar{I}_j\right)\left(L_i-\bar{L}\right),$$

где:

M – количество градаций описательных шкал (признаков);

σ_j – среднеквадратичное отклонение частных критериев знаний вектора класса;

σ_l – среднеквадратичное отклонение по вектору распознаваемого объекта.

В теореме А.Н.Колмогорова (3) интегральным критериям АСК-анализа соответствует весовой коэффициент (функция): g_q.

В итоге получаем следующую таблицу соответствий основных понятий АСК-анализа и теоремы А.Н.Колмогорова:

№	Наименование	Теорема А.Н.Колмогорова	АСК-анализ
1	Конкретный образ состояния идентифицируемого объекта или ситуации	$f(x_1,...,x_n)$	$\vec{L}_i=\{L_i\}$
2	Обобщенный образ *j-го* или *q-го* класса	$\psi^{pq}(x_p)$	$\vec{I}_{ij}=\{I_{ij}\}$
3	Функция g_q – мера сходства конкретного образа объекта с обобщенным образом q-го класса	g_q	$I_j=\sum_{i=1}^{M} I_{ij}L_i,$

			$I_j = \frac{1}{\sigma_j \sigma_l M} \sum_{i=1}^{M} \left(I_{ij} - \bar{I}_j\right)\left(L_i - \bar{L}\right),$
4	Строка матрицы системно-когнитивной модели - значение фактора (таблица 4)	p	i
5	Колонка матрицы системно-когнитивной модели – класс (таблица 4)	q	j

13.2.1.5. Математическая формулировка теоремы А.Н.Колмогорова для классического АСК-анализа

Учитывая, что M – это число строк, соответствующих значениям факторов в матрице модели (таблица 4), а W – число колонок, соответствующих классам, в этой матрице, теорема А.Н.Колмогорова, в интерпретации, принятой в данной работе (3) примет вид (4):

$$f(x_1,...,x_n) = \sum_{q=1}^{W}\left(g_q \sum_{p=1}^{M} \psi^{pq}\left(x_p\right)\right) \quad (4)$$

В терминологии теории рядов константу g_q в выражении (3) естественно интерпретировать как весовые коэффициенты ряда, а функции $\psi^{pq}\left(x_p\right)$ – как базисные функции, по которым производится разложение в ряд функции $f(x_1,...,x_n)$. Эта теорема означает, что для реализации функций многих переменных достаточно операций взвешенного суммирования (суперпозиции) функций одной переменной.

Удивительно, что в этом представлении лишь функции весовых коэффициентов g_q зависят от представляемой функции $f(x_1,...,x_n)$, а функции $\psi^{pq}\left(x_p\right)$ универсальны.

Однако, определение вида базисных функций $\psi^{pq}\left(x_p\right)$ и весовых коэффициентов g_q для данной конкретной функции $f(x_1,...,x_n)$ представляет собой *математическую* **проблему**, *для которой пока не найдено общего математически строго решения*.

При этом для частных случаев, т.е. конкретных видов базисных функций и весовых коэффициентов, таких решений найдено довольно много. В математике разработано довольно много различных *конкретных* вариантов разложений функций в ряды, обычно, но не всегда, названных в честь разработавших их математиков: это бином Ньютона, ряд Тейлора (разложение в ряд по степенным функциям), ряд Маклорена, ряд Фурье, ряд Лагранжа и Бюрмана-Лагранжа, полиномы Чебышева, ряд Лорана, разложение в ряд по экспонентам, разложение по специальным функциям[25], таким как полиномы Лежандра, полиномы Лагерра, полиномы Эрмита, функции Бесселя и т.д. [7]. Благодаря наличию рекуррентных соотношений для большинства рядов их численный расчет не является проблемой.

[25] См., например: http://eqworld.ipmnet.ru/ru/library/mathematics/special.htm

В данной работе предлагается рассматривать предлагаемую математическую модель АСК-анализа как вариант общего и универсального практического решения проблемы разработки базисных функций и весовых коэффициентов для разложения в ряд функции состояния идентифицируемого объекта или ситуации. В этом контексте функция $f(x_1,...,x_n)$ ***интерпретируется как конкретный образ состояния идентифицируемого объекта или ситуации, функция*** $\psi^{pq}(x_p)$ ***– как обобщенный образ q-го класса, а функция*** g_q ***– мера сходства конкретного образа объекта или ситуации с обобщенным образом класса.***

Что же конкретно имеется в виду? В разделе 2.1.1 мы привели следующую таблицу:

Описательные шкалы (факторы)	**Градации описательных шкал (значения факторов)**
1-й фактор	$X_{1min}=1 <= X_1 <= X_{1max}$
2-й фактор	$X_{2min}= X_{1max} <= X_2 <= X_{2max}$
…	…
i-й фактор	$Xi_{min} <= Xi <= Xi_{max}$
…	…
n-й фактор	$Xn_{min} <= Xn <= Xn_{max}=M$

Из выражения (3) мы видим, что базисная функция $\psi^{pq}(x_p)$ для разложения функции $f(x_1,...,x_n)$ в ряд, представляет собой функцию от аргумента x_i, диапазон изменения которого соответствует одному фактору или одной описательной шкале. Иначе говоря, это часть колонки таблицы 4, соответствующая *q-му* классу и *p-му* фактору.

Возникает естественный вопрос о том, что же в АСК-анализе соответствует сумме этих функций: $\sum_{p=1}^{n}\psi^{pq}(x_p)$? Поскольку индекс p – это индекс по всем строкам матрицы системно-когнитивной модели, всем диапазонам изменения аргумента x_i, то на взгляд автора ответ вполне очевиден: ***это вся колонка таблицы 4, соответствующая q-му классу***, т.е. это ***<u>обобщенный образ q-го класса (3) по всем факторам:</u>***

$$\psi^q(x)=\sum_{p=1}^{M}\psi^{pq}(x_p), \qquad (4)$$

где: $X_{1min}=1 <= X_p <= Xn_{max}=M$

Таким образом выражение для теоремы А.Н.Колмогорова (3) с учетом (4) примет вид (5):

$$f(x)=\sum_{q=1}^{W}\left(g_q\psi^q(x)\right) \qquad (5)$$

Выражение (5) – это классическое выражение для разложения функции $f(x)$ в ряд по базисным функциям $\psi^q(x)$, *т.е. это взвешенная суперпозиция функций* $\psi^q(x)$ *с весами:* g_q.

Функции $\psi^q(x)$ в АСК-анализе формируются в процессе синтеза моделей и представляют собой обобщенные образы классов, функция $f(x)$ описывает идентифицируемый объект или прогнозируемую ситуацию, а весовые коэффициенты разложения в ряд g_q представляют собой интегральные критерии сходства функции состояния объекта или ситуации с обобщенными образами классов и вычисляются при распознавании, идентификации или прогнозировании.

Графики базисных функций $\psi^q(x)$ построить не сложно: для этого в MS Excel надо отобразить в виде графика *q-ю* колонку матрицы соответствующей статистической или системно-когнитивной модели (СК-модель).

Отметим ***принципиальную важность выражения (5) для проектирования структуры баз знаний в АСК-анализе и системе «Эйдос»***. Для этого сравним модели представления знаний системы «Эйдос» (Луценко Е.В., 1979) и фреймовую модель Марвина Мински (1975). В модели Мински каждому обобщенному образу класса (фрейма-прототипа) соответствует много слотов (описательных шкал) со своими шпациями (градациями) и в каждом фрейме они в общем случае разные. Поэтому при увеличении количества фреймов-прототипов (классов) в модели Мински количество таблиц и отношений межу ними расчет как снежный ком. Напрашивается идея как-то упростить фреймовую модель представления знаний. В 1979 году Е.В.Луценко (в то время старший инженер-программист вычислительного центра Краснодарского медицинского института) предложил следующее решение: описывать все фреймы-прототипы в одной общей системе слотов и шпаций (описательных шкал и градаций), т.е. по сути в одной таблице вида таблиц 1-4. Справочники классификационных и описательных шкал и градаций составляли еще 6 таблиц, обучающей выборки – еще 3, тестовой выборки – еще 3. Это решение приводило к независимости количества таблиц и отношений между ними в базах знаний системы «Эйдос»[26] от числа классификационных и описательных шкал и градаций (т.е. от числа фреймов-прототипов, слотов и шпаций). ***Корректность*** этого решения обосновывалось именно теоремой А.Н.Колмогорова, а именно выражением (5), т.е. тем, что весовые коэффициенты, соответствующие разным слотам и шпациям (т.е. разным описательным шкалам и

[26] Именно в 1979 году была разработана математическая модель системы «Эйдос», точнее суть этой модели. Тогда же она положительно прошла экспертизу на уровне докторов физ.-мат. Наук, профессоров, занимающихся интеллектуальными технологиями.

градациям), соответствующих разным фреймам-прототипам (классам), можно просто складывать.

13.2.1.6. Объекты математической модели АСК-анализа как алгебраические структуры в рамках высшей алгебры

Важно отметить, что в АСК-анализе и классификационные, и описательные шкалы могут быть как ***числовыми****, так и* ***текстовыми****, а текстовые могут быть либо* ***номинальными****, либо* ***порядковыми****.*

Таким образом классификационные и описательные шкалы в АСК-анализе можно рассматривать как ***алгебраические структуры*** (группы, кольца и поля), на которых определены те или иные операции над их градациями:

– текстовые шкалы: номинальные: операция эквивалентности;

– текстовые шкалы: порядковые: операции эквивалентности и больше/меньше;

– числовые шкалы: операции эквивалентности, больше/меньше, сложения, вычитания, умножения и деления.

Поэтому и все остальные объекты математической модели АСК-анализа, такие как описания объектов обучающей и распознаваемой выборки, матрица абсолютных частот, матрицы условных и безусловных процентных распределений, матрица информативностей и других системно-когнитивных моделей (см. таблицу 3), а также матрицы сходства, базы агломеративной древовидной классификации, SWOT-анализа и другие, также *можно рассматривать как алгебраические структуры в рамках высшей алгебры*. В частности, матрица информативностей по своей математической структуре является тензором, описывающим метрику многомерного неевклидового неортонормированного когнитивного пространства, отражающего предметную область в системно-когнитивной модели. Однако более подробное рассмотрение этих вопросов не входит в задачи данной статьи, тем более что этому вопросу посвящено довольно много работ автора [9-38][27].

13.2.1.7. Значимость значения фактора, степень детерминированности класса и ценность модели

Отметим, что как значимость значения фактора, степень детерминированности класса и ценность или качество модели в АСК-анализе рассматривается <u>вариабельность</u> значений частных

[27] Более полный список этих работ можно посмотреть, например здесь: http://lc.kubagro.ru/aidos/Work_on_emergence.htm

критериев этого значения фактора, класса или модели в целом (таблица 4).

Численно эта вариабельность может измеряться разными способами, например средним отклонением модулей частных критериев от среднего, дисперсией или среднеквадратичным отклонением или его квадратом. В системе «Эйдос» принят последний вариант, т.к. эта величина совпадает с ***мощностью сигнала****, в частности мощностью информации, а* ***в АСК-анализе все модели рассматриваются в как источник информации об объекте моделирования.***

Поэтому есть все основания уточнить традиционную терминологию АСК-анализа (таблица 5).

Термины каждой строки по сути являются синонимами. Исследование погрешности (дисперсии) для этих выражений – это предмет дальнейшего исследования. Отметим, что впервые количественное выражение для корня информационной мощности модели предложено проф. Е.В.Луценко в работе [9] еще в 2002 году.[28] Для синтеза 3 статистических и 7 системно-когнитивных моделей используется режим 3.5 системы «Эйдос», описанный ниже.

Таблица 11 – Уточнение терминологии АСК-анализа

№	Традиционные термины (синонимы)	Новый термин	Формула
1	1. Значимость значения фактора (признака). 2. Дифференцирующая мощность значения фактора (признака). 3. Ценность значения фактора (признака) для решения задачи идентификации и других задач	Корень из информационной мощности значения фактора	$\sigma_{i\Sigma}=\sqrt[2]{\frac{1}{W-1}\sum_{j=1}^{W}\left(I_{ij}-\bar{I}_i\right)^2}$
2	1. Степень детерминированности класса. 2. Степень обусловленности класса.	Корень из информационной мощности класса	$\sigma_{\Sigma j}=\sqrt[2]{\frac{1}{M-1}\sum_{i=1}^{M}\left(I_{ij}-\bar{I}_j\right)^2}$
3	1. Качество модели. 2. Ценность модели. 3. Степень сформированности модели. 4. Количественная мера степени выраженности закономерностей в моделируемой предметной области	Корень из информационной мощности модели	$H=\sqrt[2]{\frac{1}{(W\cdot M-1)}\sum_{j=1}^{W}\sum_{i=1}^{M}\left(I_{ij}-\bar{I}\right)^2}$

13.2.1.8. Абсолютная и относительная сходимость прогнозного ряда. Ортонормирование системы функций классов: в какой степени оно действительно необходимо?

При дальнейшем развитии аналогии между распознаванием и разложением функции ситуации по обобщенным функциям классов естественно возникают вопросы: о полноте, избыточности и ортонормированности системы векторов классов как функций, по которым

[28] http://elibrary.ru/item.asp?id=18632909 формула (3.81) на стр.290

проводится разложение вектора объекта; о сходимости, т.е. вообще возможности и корректности такого разложения.

В общем случае вектор объекта совершенно не обязательно должен разлагаться в ряд по векторам классов таким образом, что сумма ряда во всех точках точно совпадала со значениями исходной функции. Это означает, что система векторов классов может быть ***неполна*** по отношению к профилю распознаваемого объекта, и, тем более, всех возможных объектов.

Предлагается считать не разлагаемые в ряд, т.е. плохо распознаваемые объекты, суперпозицией хорошо распознаваемых объектов ("похожих" на те, которые использовались для формирования обобщенных образов классов), и объектов, которые и не должны распознаваться, так как объекты этого типа не встречались в обучающей выборке и не использовались для формирования обобщенных образов классов, а также не относятся к представляемой обучающей выборкой генеральной совокупности.

Нераспознаваемую компоненту можно рассматривать либо как шум, либо считать ее полезным сигналом, несущим ценную информацию о неисследованных объектах интересующей нас предметной области (в зависимости от целей и тезауруса исследователей).

Первый вариант не приводит к осложнениям, так как примененный в математической модели алгоритм сравнения векторов объектов и классов, основанный на вычислении нормированной корреляции Пирсона (сумма произведений), является *весьма устойчивым к наличию белого шума* в идентифицируемом сигнале.

Во втором варианте необходимо дообучить систему распознаванию объектов, несущих такую компоненту (в этой возможности и заключается адаптивность модели). Технически этот вопрос решается просто копированием описаний плохо распознавшихся объектов из распознаваемой выборки в обучающую, их идентификацией экспертами и дообучением системы.

Кроме того, может быть целесообразным дополнить справочник классификационных шкал и градаций новыми классами, соответствующими этим объектам, а справочник описательных шкал и градаций – новыми признаками, необходимыми для описания этих объектов.

Однако на практике гораздо чаще наблюдается противоположная ситуация (можно даже сказать, что она типична), когда система векторов ***избыточна***, т.е. в системе классов распознавания есть очень похожие классы (между которыми имеет место высокая корреляция, наблюдаемая в режиме: "кластерно-конструктивный анализ"). Практически это означает, что в системе сформировано несколько практически одинаковых образов с разными наименованиями. Для исследователя это само по себе является

очень ценной информацией. Однако если исходить только из потребности разложения распознаваемого объекта в ряд по векторам классов (чтобы определить суперпозицией каких образов он является, т.е. "разложить его на компоненты"), то наличие сильно коррелирующих друг с другом векторов представляется неоправданным, так как просто увеличивает размерности данных, внося в них мало нового по существу. Поэтому возникает задача *исключения избыточности системы классов распознавания*, т.е. выбора из всей системы классов распознавания такого минимального их набора, в котором профили классов минимально коррелируют друг с другом, *т.е. ортогональны в фазовом пространстве признаков*. Это условие в теории рядов называется "ортонормируемостью" системы базовых функций, а в факторном анализе связано с идеей выделения "главных компонент".

В предлагаемой математической модели реализованы два варианта выхода из данной ситуации:

1) исключение неформирующихся, расплывчатых классов;

2) объединение почти идентичных по содержанию (дублирующих друг друга) классов.

Однако выбрать нужный вариант и реализовать его, используя соответствующие режимы, пользователь технологии АСК-анализа должен сам. Возможно в будущем эти процессы будут автоматизированы. Вся необходимая и достаточная информация для принятия соответствующих решений предоставляется пользователю инструментария АСК-анализа, в качестве которого в настоящее время выступает система «Эйдос».

Если считать, что функции образов составляют формально-логическую систему, к которой применима теорема Геделя, то можно сформулировать эту **теорему** для данного случая следующим образом:

Для любой системы базисных функций $\{\varphi(x)\}$ в некотором линейном пространстве функций L всегда существует по крайней мере одна такая ***ненулевая*** функция, что она ***не может*** быть разложена в ряд по данной системе базисных функций, т.е. ***функция, которая является ортогональной ко всей системе базисных функций в целом". Этим утверждается, что ЛЮБАЯ система базисных функций принципиально неполна.*** Добавление этой новой функцию в систему базисных функций $\{\varphi(x)\}$ приводит к ***повышению размерности*** линейного пространства функций L. Принципиально размерность этого пространства ничем не ограничена.

Строгое математическое доказательство этой теоремы не входит в задачи данной статьи и является делом будущего. Сейчас же уместно отметить лишь, что на взгляд автора математическое представление об обязательной ортогональности и полноте базисных функций для

разложения в ряд является скорее абстрактным математическим требованием, имеющим мало относящимся к реальности, примерно как реально невыполнимые требования факторного анализа об абсолютной точности исходных данных, полной независимости друг от друга факторов и аддитивности их действия на объект моделирования (что эквивалентно требованию его абсолютной линейности).

Очевидно, не взаимосвязанными друг с другом могут быть только четко оформленные, детерминистские образы, т.е. образы с высокой степенью редукции ("степень сформированности конструкта"). Поэтому в процессе выявления взаимно-ортогональных базисных образов, в первую очередь, будут выброшены аморфные "расплывчатые" образы, которые связаны практически со всеми остальными образами.

В некоторых случаях результат такого процесса представляет интерес, и это делает оправданным его реализацию. Однако можно предположить, что наличие расплывчатых образов в системе является оправданным, так как в этом случае система образов не будет формальной и подчиняющейся теореме Геделя. Следовательно, система распознавания будет более полна в том смысле, что увеличится вероятность идентификации *любого объекта*, предъявленного ей на распознавание. Конечно, уровень сходства с аморфным образом не может быть столь высоким, как с четко оформленным. Поэтому в этом случае более уместно применить термин "ассоциация" или нечеткая, расплывчатая идентификация, чем "однозначная идентификация".

Итак, можно сделать следующий вывод: допустимость в математической модели АСК-анализа не только четко оформленных (детерминистских) образов, но и образов аморфных, нечетких, расплывчатых не только не является недостатком, но наоборот, является важным достоинством данной модели. Это обусловлено тем, что данная модель распознавания обеспечивает корректные результаты анализа, идентификации и прогнозирования даже в тех случаях, когда модели идентификации и информационно-поисковые системы детерминистского типа традиционных АСУ практически неработоспособны. В этих условиях данная модель АСК-анализа работает как система *ассоциативной (нечеткой) идентификации.*

Таким образом можно обоснованно сделать общий вывод о том, что если в чисто математической теории разложения функций в ряды требование ортонормированности базисных функций является вполне обоснованным, то для практических приложений это не играет принципиальной роли, более того, в практических приложениях использование для разложения в ряд неортонормированной системы базисных функций представляет большой интерес, т.к. открывает новые широкие перспективы исследований взаимосвязей между факторами, а также между

факторами и поведением объекта моделирования. Кроме того это может эффективно использоваться при принятии решений (см. раздел:2.4).

Совершенно аналогичная ситуация наблюдается с другими строго математическими понятиями и является обычной устоявшейся практикой. Например строго математические понятия материальной и математической точки, бесконечно малых и т.п. на практике, например в физике и в численных расчетах на компьютерах, т.е. в численных методах и дискретной математике, заменяются на элементы малых, но конечных размеров, например на конечные разности. При этом интегралы заменяются на суммы.

Более того, использование неортонормированной системы базисных функций не только вполне корректно для практических приложений, но и представляет особый большой интерес, т.к. при этом появляется возможность ***изучения сходства/различия базисных функций по их смыслу***, т.е. по влиянию на вид функций состояний объектов и ситуаций, разлагаемые в ряд по ним. Для этого могут использоваться, например, когнитивные диаграммы и дендрограммы агломеративной кластеризации. С одной стороны, это позволяет исследовать ***нелинейные*** системы, для которых не выполняется большая предельная теорема о действии большого количества ***независимых*** друг от друга факторов, а значит не выполняется нормальное распределение и ***неприменимы методы параметрической статистики***.

13.2.2. Суть математической модели сценарного АСК-анализа

13.2.2.1. Идея и концепция сценарного АСК-анализа

Идея сценарного АСК-анализа очень проста: *к базовым шкалам, созданным точно как в классическом АСК-анализе, добавить шкалы сценариев, отражающие динамику изменения показателей, отраженных базовыми шкалами.*

Концепция сценарного АСК-анализа:

1. К каждой классификационной шкале модели, отражающей точечные значения будущих показателей объекта моделирования, добавить классификационную шкалу, градации которой (новые классы), будут отражать динамику изменений этих точечных показателей в будущем, т.е. будущие сценарии изменения показателя, отраженного базовой шкалой.

2. К каждой описательной шкале модели, отражающей точечные значения прошлых показателей объекта моделирования, добавить описательную шкалу, градации которой (новые значения факторов), будут отражать динамику изменений этих точечных показателей в прошлом, т.е. прошлые сценарии изменения показателя, отраженного базовой шкалой.

На рисунке 6 ниже показано, как на основе одной базовой описательной шкалы и одной базовой классификационной шкалы с их градациями (выделены желтым фоном) образованы соответствующие им классификационные и описательные сценарные шкалы, и градации (выделены зеленым фоном).

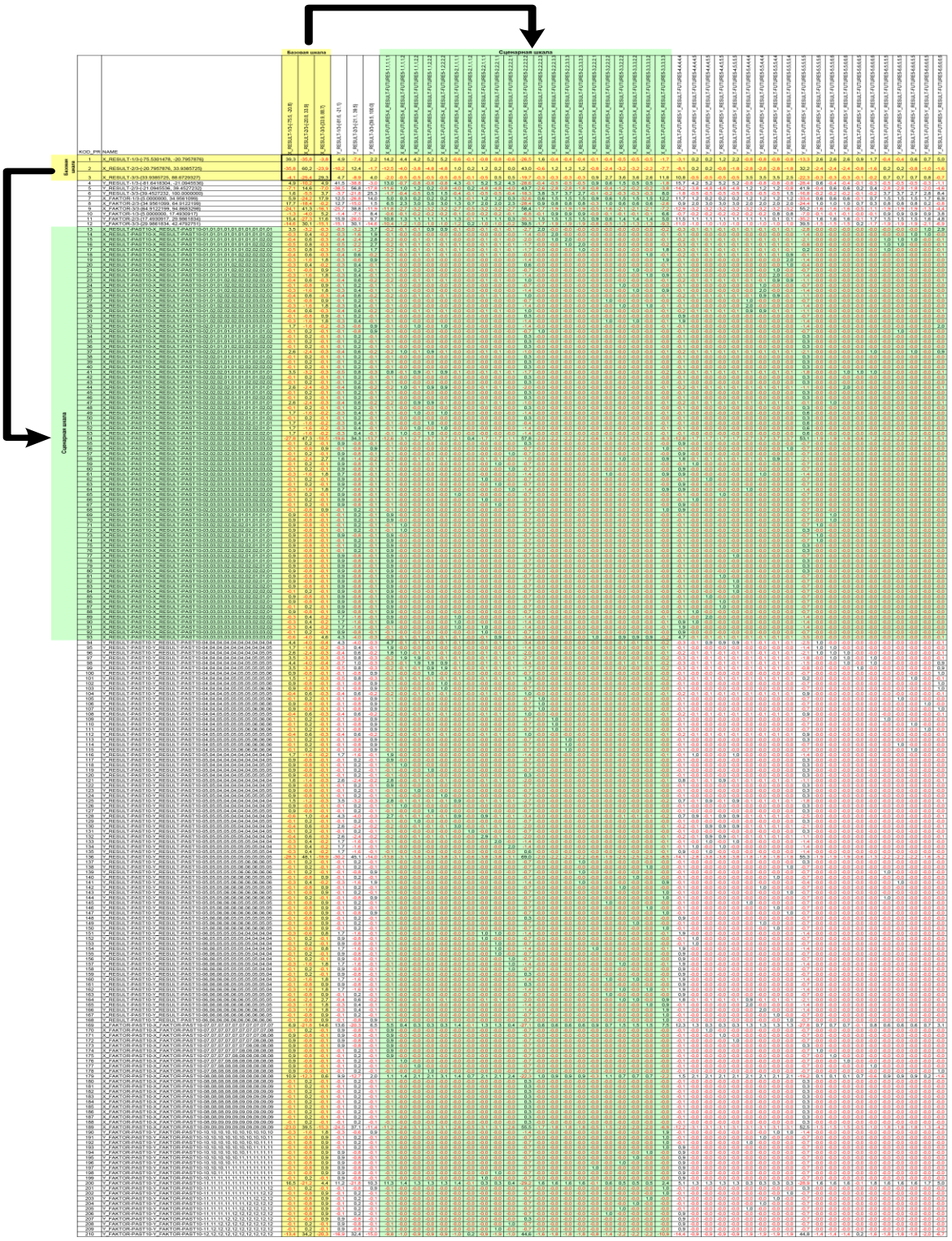


Рисунок 6.

При этом глубина предыстории составляет 10 точечных значений показателя базовой шкалы, а горизонт прогнозирования – 5 точек.

Эти возможности реализованы в автоматизированном программном интерфейсе импорта данных из внешних источников данных (API) системы «Эйдос».

В результате в модели кроме тех шкал и градаций, которые были и в классическом АСК-анализе, добавляются новые классификационные и описательные шкалы и градации, отражающие прошлые и будущие сценарии изменения показателей соответствующих базовых шкал.

Эти новые шкалы и градации обрабатываются в сценарном АСК-анализе абсолютно также, как в классическом АСК-анализе, но кроме этого дополнительно только в сценарном АСК-анализе реализуются интересные новые возможности, подробнее описанные ниже в данном разделе.

13.2.2.2. Математическая формулировка теоремы А.Н.Колмогорова для сценарного АСК-анализа

В ***сценарном АСК-анализе:***

1. В качестве прошлых значений факторов, влияющих на поведение объекта моделирования, рассматриваются сценарии изменения значений этих факторов. В качестве результата влияния факторов рассматривается сценарии поведения объекта моделирования под влиянием этих факторов.

2. На основе анализа исходных данных выявляются ранее наблюдавшиеся сценарии изменения значений факторов, влияющих на объект моделирования, и сценарии поведения объекта моделирования под влиянием этих значений факторов.

3. Путем обобщения (многопараметрической типизации) конкретных сценариев поведения объекта моделирования формируются обобщенные образы сценариев развития событий (классы) под влиянием сценариев изменения значений факторов.

Так же как в классическом АСК-анализе, в сценарном АСК-анализе математической моделью класса является вектор частных критериев, соответствующий колонке из таблицы 4. Сами частные критерии, используемые в текущей версии системы «Эйдос», приведены в таблице 3.

Поэтому все выводы, полученные ранее по теореме А.Н.Колмогорова для классического АСК-анализа сохраняют силу и для сценарного АСК-анализа, в частности выражение для теоремы А.Н.Колмогорова (5):

$$f(x)=\sum_{q=1}^{W}\left(g_q\psi^q(x)\right) \qquad (5)$$

Кроме того, в сценарном АСК-анализе сами классы в сценарных классификационных шкалах являются сценариями, т.е. ***функциями***

будущих прогнозных сценариев: s(t), отражающими динамику точечных показателей соответствующей базовой шкалы.

Поэтому выражение для теоремы А.Н.Колмогорова для сценарного АСК-анализа может быть записано в виде (5):

$$\boxed{f(t)=\sum_{j=1}^{W}\left(g_j s_j(t)\right)}, \qquad (5)$$

где:

t – время;

f(t) – средневзвешенный прогнозируемый будущий сценарий;

s(t) – обобщенный образ сценарного класса (функция будущего сценария);

g_j – уровень сходства функции прогнозируемой ситуации $\vec{L}_i=\{L_i\}$ с обобщенным образом сценарного класса (функцией будущего сценария *s(t)*).

При решении задачи идентификации и прогнозирования эти сценарные классы, т.е. будущие сценарии, также прогнозируются с различной достоверностью (с различными уровнями сходства). При значениях интегрального критерия сходства больше нуля это прогнозы того, что будет (положительные прогнозы), а при значениях меньше нуля – того, чего не будет (отрицательные прогнозы).

13.2.2.3. Постановка задачи прогнозирования сценариев будущих событий (классов) на основе сценариев прошлых событий (значений факторов)

В сценарном методе АСК-анализа сценарии развития событий в прошлом рассматриваются как значения факторов, обуславливающие сценарии развития событий в будущем.

На основе анализа исходных данных выявляются ранее наблюдавшиеся сценарии и на основе их обобщения формируются обобщенные образы сценариев развития событий, т.е. классов.

При синтезе системно-когнитивных моделей вычисляется количество информации, которое содержится в конкретных прошлых сценариях о наступлении или не наступлении конкретных будущих сценариев.

Например, фондовый рынок описывается временными рядами курсов ценных бумаг и валют, а также временными рядами, описывающих различные внутренние и внешние факторы, влияющие на фондовый рынок. Среди ***внутренних факторов*** фондового рынка можно отметить саму динамику взаимных курсов различных ценных бумаг и валют, динамику числа банков, участвующих в торгах, динамику спрос и предложение на различные ценные бумаги и валюты. Среди ***внешних факторов*** можно выделить общую политическую ситуацию, уровень

информационного и вооруженного противостояния в горячих точках и на основных (стратегических) транспортных и энергетических магистралях, уровень мировой экономической активности, наличие различных глобальных заболеваний, типа пандемии Covid-19, а также выступления и заявления ведущих политических лидеров мира, лидеров наиболее мощных экономик мира, террористические акты, особенно такие масштабные как 11 сентября в США, а также такие казалось бы курьезные случаи, как падение президента США Дж.Буша с трапа военного вертолета при прибытии его в Японию. Наблюдения за ситуацией на фондовом рынке образуют базу данных, в которой строки соответствуют различным наблюдениям, привязанным ко времени, а столбцы отражают факторы и результаты их влияния.

В программном инструментарии АСК-анализа системе «Эйдос» есть развитые программные интерфейсы, позволяющие ввести подобные данные в систему «Эйдос», создать на их основе модели и применить эти модели для решения задач прогнозирования, принятия решений и исследования моделируемой предметной области путем исследования ее модели.

Для этого выполняются следующие этапы АСК-анализа:

1. Когнитивно-целевая структуризация предметной области.
2. Формализация предметной области (автоматическая разработка классификационных и описательных шкал и градаций, кодирование исходных с их помощью и генерация обучающей выборки).
3. Синтез и верификация моделей.
4. Решение задач идентификации и прогнозирования.
5. Решение задач поддержки принятия решений.
6. Исследование моделируемой предметной области путем исследования ее модели.

Ниже мы подробнее рассмотрим содержание и выполнение этих этапов на численном примере.

13.2.2.4. Алгоритм выявления сценариев изменения значений факторов и сценариев поведения объекта моделирования

Рассмотрим **алгоритм выявления сценариев изменения значений факторов и сценариев поведения объекта моделирования** (п.2) в случае, когда исходные данные представляют собой *временные ряды*.

Шаг 1-й. Базовые шкалы и значения шкал формируются как обычно при формализации предметной области. При этом в качестве значений градаций числовых шкал рассматриваются числовые диапазоны, а в качестве значений текстовых шкал (номинальных и порядковых) рассматриваются уникальные текстовые значения. Числовые диапазоны

могут быть либо равными с разным числом наблюдений, либо разными (адаптивными) с примерно одинаковым числом наблюдений.

Шаг 2-й: Организуется цикл по текущей записи базы исходных данных от 1-й записи до последней.

Шаг 3-й. Организуется цикл по всем измерительным шкалам, как классификационным, так и описательным. Классификационные шкалы и градации используются для формального описания и кодирования будущих состояний объекта моделирования, а описательные, – как для формального описания и кодирования прошлых состояний самого объекта моделирования (его предыстории), так и для описания различных факторов, действующих на объект моделирования. Эти факторы могут быть классифицированы как зависящие от нашей воли (факторы управления, применение различных технологий), так и не зависящие от нее – это факторы окружающей среды. Факторы окружающей среды могут быть классифицированы в соответствии с иерархическими уровнями организации внешней среды: природной, технологической, организационной, экономической и политической и т.д.

Шаг 4-й. Относительно текущей записи базы исходных данных *по каждой шкале* определяются коды градаций базовых классификационных и описательных шкал на заданную глубину предыстории в прошлое и на заданный горизонт прогнозирования в будущее. На основе этой информации формируются и добавляются в справочники шкалы прошлых и будущих сценариев. Будущие сценарии образуются на основе базовых классификационных шкал, а прошлые – на основе базовых описательных шкал. Название шкалы-сценария образуется из названия базовой шкалы, но основе которой она образована, слова "Будущее" или "Прошлое" (FUTURE or PAST) и КОДОВ градаций базовой шкалы сценария.

Шаг 5-й. Конец цикла по шкалам.

Шаг 6-й. Конец цикла по записям базы исходных данных.

В сценарном АСК-анализе вектора классов (таблица 4) рассматриваются как базисные функции для разложения в ряд сценария идентифицируемой ситуации. При этом в качестве весовых коэффициентов разложения в ряд используются значения интегральных критериев сходства идентифицируемой ситуации с соответствующими классами [8, 9].

13.2.2.5. Разработка частных положительных и отрицательных прогнозов и оценка их достоверности как разложение функции ситуации в ряд по функциям классов

При прогнозировании текущая ситуация, описанная прошлыми сценариями, сравнивается с обобщенными образами классов, т.е. с будущими сценариями, и разлагается в спектр по ним аналогично прямому преобразованию Фурье.

По сути в обозначениях теоремы А.Н.Колмогорова (1957) для сценарного АСК-анализа:

$$\boxed{f(t)=\sum_{j=1}^{W}\left(g_j s_j(t)\right)}, \qquad (5)$$

распознавание (разложение функции ситуации в ряд по базисным функциям классов) сводится к нахождению весовых коэффициентов g_j, при этом в качестве базисных функций $s(t)$ используются обобщенные образы классов, т.е. будущих сценариев.

Весовые коэффициенты g_j представляют собой интегральные критерии сходства идентифицируемой ситуации с обобщенными образами классов, используемые в настоящее время в АСК-анализе: сумма знаний и резонанс знаний, рассмотренных выше.

При этом оказывается, что текущая ситуация имеет положительное сходство разной степени с одними конкретными будущими сценариями, и отрицательное сходство с другими будущими сценариями.

Если уровень сходства текущей ситуации с будущим сценарием больше нуля, то такой прогноз называется положительным. Положительный прогноз описывает прогноз того, «что будет».

Если уровень сходства текущей ситуации с будущим сценарием меньше нуля, т.е. по сути это уровень различия, то такой прогноз называется отрицательным. Отрицательный прогноз описывает прогноз того, «чего не будет».

Модуль уровня сходства/различия описания текущей ситуации с прогнозами отражает оценку системой «Эйдос» уровня достоверности этих прогнозов.

Таким образом при прогнозировании описание текущей ситуации по сути разлагается в ряд по обобщенным образам классов, соответствующих будущим сценариям развития событий. Коэффициентами этого ярда являются урони сходства/различия описания текущей ситуации с обобщенными образами классов.

13.2.2.6. Формирование средневзвешенных положительных (что будет) и отрицательных (чего не будет) прогнозов как преобразование, обратное разложению функции ситуации в ряд по функциям классов

Средневзвешенный прогноз формируется путем обратного преобразования, аналогичного обратному преобразованию Фурье, в котором в качестве базисных функций используются обобщенные образы классов прогнозируемых сценариев того что будет и того что не будет с их весами.

По сути средневзвешенный прогноз является взвешенной суперпозицией обобщенных образов классов с весами, равными сходству/различию описания текущей ситуации с этими обобщенными образами классов. Отметим, что каждый обобщенный образ класса соответствует определенному сценарию развития событий, который реально наблюдался в эмпирических данных.

13.2.2.7. Технический и фундаментальный подходы и их синтез в сценарном АСК-анализе

Технический анализ предполагает прогнозирование хода временных рядов на основе данных из тех же временных рядов за прошлый период. В терминологии АСК-анализа это просто означает, что одни и те же временные ряды используются и в качестве классификационных шкал, и в качестве описательных шкал. Классификационные шкалы позволяют формально описать будущие события, которые необходимо прогнозировать. Описательные шкалы позволяют формально описать прошлые события, которые рассматриваются в качестве факторов (причин), обуславливающих будущие события.

Фундаментальный анализ предполагает прогнозирование хода временных рядов на основе данных из других временных рядов за прошлый период, отражающих динамику различных внутренних и внешних факторов. В терминологии АСК-анализа это означает, что одни временные ряды используются и в качестве классификационных шкал, описывающих будущие события, а другие используются в качестве описательных шкал, описывающих факторы (причины), обуславливающие эти будущие события. В сценарном АСК-анализе нет никаких проблем использовать для прогнозирования хода временных рядов ***одновременно*** и данные из тех же временных рядов за прошлый период (как в техническом анализе), так и данные из других временных рядов за прошлый период, отражающих динамику различных внутренних внешних факторов, действующих на ситуацию (как в фундаментальном анализе).

Таким образом сценарный АСК-анализ позволяет легко объединить в одном приложении и технический, и фундаментальный анализ, что и

*отражено в названии этого синтетического подхода: «**техно-фундаментальный сценарный АСК-анализ**».*

13.2.3. Развитый алгоритм принятия решений АСК-анализа

Традиционно, управляющие решения принимаются путем многократного решения задачи прогнозирования при различных значениях управляющих факторов и выбора такого их сочетания, которое обеспечивает перевод объекта управления в целевое состояние. Однако на реальные объекты управления действуют сотни и тысячи управляющих факторов, каждый из которых может иметь десятки значений. Полный перебор всех возможных сочетаний значений управляющих факторов приводит к необходимости решения задачи прогнозирования десятки и сотни тысяч и даже миллионы раз для принятия одного решения, и это является совершенно неприемлемым на практике. Поэтому необходим метод принятия решений не требующий значительных вычислительных ресурсов. Таким образом, налицо противоречие между фактическими и желаемым, в чем и состоит проблема, решаемая в работе. В данной работе предлагается развитый алгоритм принятия решений путем однократного решения обратной задачи прогнозирования (автоматизированный SWOT-анализ), использующий результаты кластерно-конструктивного анализа целевых состояний объекта управления и значений факторов и однократного решения задачи прогнозирования. Этим и обуславливается актуальность темы работы. Цель работы состоит в решении поставленной проблемы. Путем декомпозиции цели сформулированы следующие задачи, являющиеся этапами достижения цели. Когнитивно-целевая структуризация предметной области; формализация предметной области (разработка классификационных и описательных шкал и градаций и формирование обучающей выборки); синтез, верификация и повышение достоверности модели объекта управления; прогнозирование, принятие решений и исследование объекта управления путем исследования его модели. В качестве метода решения поставленных задач применяется автоматизированный системно-когнитивный анализ и его программный инструментарий – интеллектуальная система «Эйдос». В результате работы предложен развитый алгоритм приятия решений, применимый в интеллектуальных системах управления. Основной вывод по результатам работы состоит в том, что предлагаемый подход позволил успешно решить поставленную проблему [18].

Предлагается следующий развитый алгоритм принятия решений в интеллектуальных системах управления на основе АСК-анализа и системы «Эйдос» (рисунок 7).

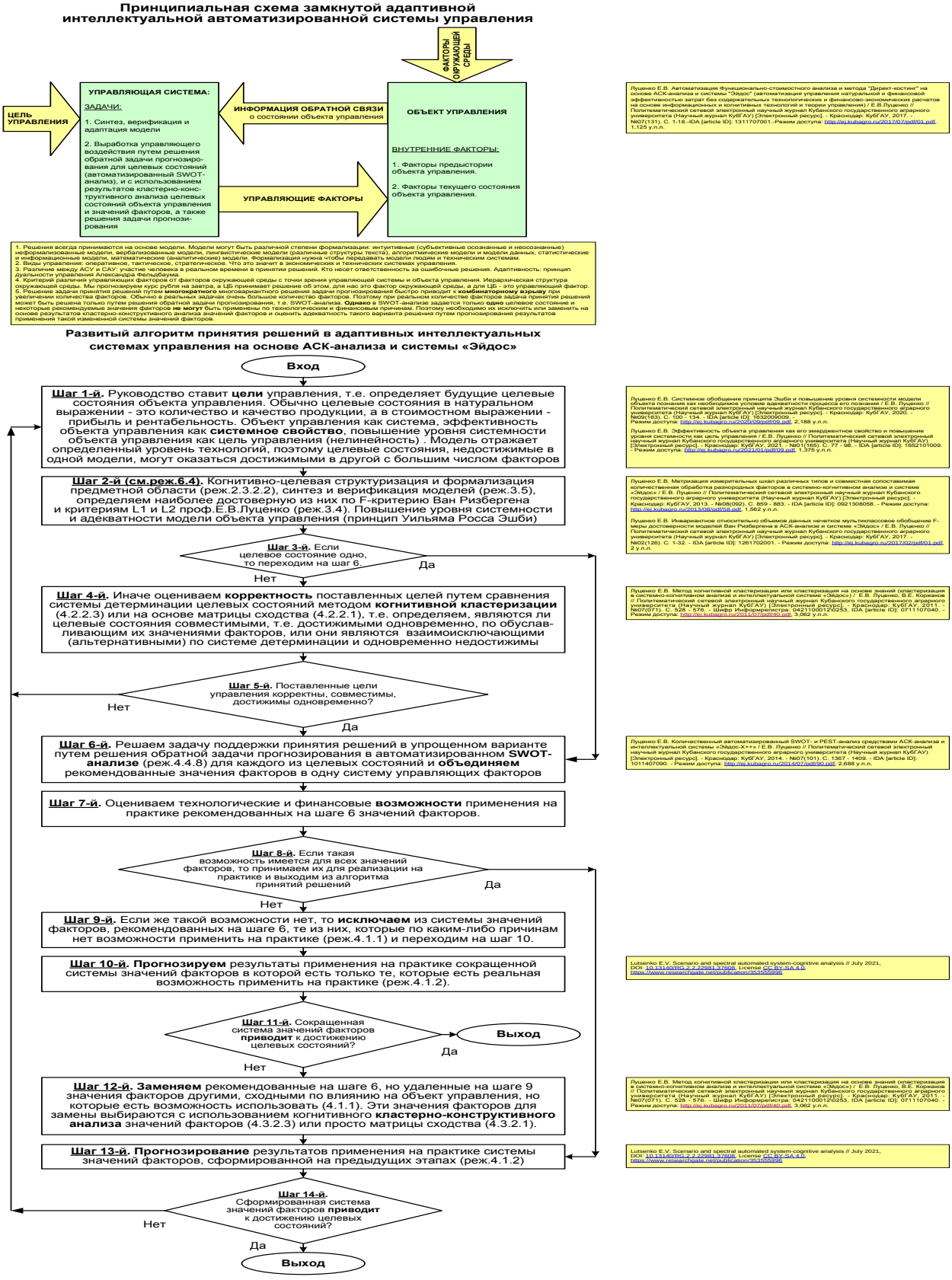


Рисунок выполнен автором

Рисунок 7. Развитый алгоритм принятия решений в интеллектуальных системах управления на основе АСК-анализа и системы «Эйдос»

Развитый алгоритм принятия решений АСК-анализа при его применении в интеллектуальных системах управления

Шаг 1-й. Ставим цели управления, т.е. определяем одно или несколько целевых состояний объекта управления. В натуральном выражении целевые состояния - это обычно количество и качество продукции, а в стоимостном выражении - прибыль и рентабельность ее производства и продажи.

Шаг 2-й. Проводим когнитивно-целевую структуризацию и формализацию предметной области, синтез и верификация статистических и системно-когнитивных моделей (СК-модели), определяем наиболее достоверную из них по F-критерию Ван Ризбергена и критериям L1 и L2 проф.Е.В.Луценко.

Шаг 3-й. Если целевое состояние одно, то переходим на шаг 6.

Шаг 4-й. Иначе оцениваем корректность поставленных целей путем сравнения системы детерминации целевых состояний методом когнитивной кластеризации или просто на основе матрицы сходства, т.е. определяем, являются ли целевые состояния совместимыми, т.е. достижимыми одновременно, по обуславливающим их значениями факторов, или они являются взаимоисключающими (альтернатив-ными) по системе детерминации и одновременно достигнуты быть не могут.

Шаг 5-й. Поставленные цели управления корректны, совместимы, достижимы одновременно?

Шаг 6-й. Решаем задачу поддержки принятия решений в упрощенном варианте путем автоматизированного когнитивного SWOT-анализа целевых состояний.

Шаг 7-й. Оцениваем технологические и финансовые возможности применения на практике рекомендованных на шаге 6 значений факторов.

Шаг 8-й. Если такая возможность имеется для всех значений факторов, то принимаем их для реализации на практике и выходим из алгоритма принятий решений

Шаг 9-й. Если же такой возможности нет, то исключаем из системы значений факторов, рекомендованных на шаге 6, те из них, которые по каким-либо причинам нет возможности применить на практике и переходим на следующий шаг.

Шаг 10-й. Прогнозируем результаты применения на практике сокращенной системы значений факторов в которой есть только те, которые есть реальная возможность применить на практике.

Шаг 11-й. Сокращенная система значений факторов приводит к достижению целевых состояний?

Шаг 12-й. Заменяем рекомендованные на шаге 6, но удаленные на шаге 9 значения факторов другими, сходными по влиянию на объект управления, но такими, которые есть возможность использовать. Эти значения факторов для замены выбираются с использованием результатов

когнитивного кластерно-конструктивного анализа значений факторов или просто матрицы сходства.

Шаг 13-й. Прогнозирование результатов применения на практике системы значений факторов, сформированной на шаге 12.

Шаг 14-й. Сформированная система значений факторов приводит к достижению целевых состояний? Если прогнозируемый результат применения на практике системы значений факторов, сформированной на шаге 12, по результатам прогнозирования приводит к переходу объекта управления в целевые состояния, то принимаем данную систему значений факторов для реализации на практике и выходим из алгоритма принятия решений. Если же прогноз показывает, что целевое состояние при использовании этой системы значений факторов не достигается, то задача управления не имеет решения в данной модели и осуществляется переход на шаг 2 для качественного изменения модели с новыми исходными данными и расширенной системной значений факторов.

После выхода из алгоритма и реализации управляющих решений цикл управления, представленный на рисунке 2, повторяется. При этом результаты управления в любом случае, т.е. как при успешном достижении целевых состояний, так и в противном случае, учитываются в исходных данных для создания модели и осуществляется пересинтез модели. Поэтому непосредственно в процессе управления происходит постоянное улучшение качества интеллектуальной модели приятия решений путем ее самообучения с учетом фактических результатов управления. Это обеспечивается тем, что интеллектуальная система «Эйдос» является одновременно инструментом для синтеза и верификации моделей объекта управления, инструментом применения этих моделей для решения задач идентификации, прогнозирования, принятия решений и исследования моделируемой предметной области путём исследования ее модели. Достоверность созданных моделей оценивается с помощью F-меры Ван Ризбергена и ее мультиклассовых, нечетких обобщений, инвариантных относительно объема выборки (Луценко 2017). Система «Эйдос» не только обеспечивает решение этих задач, но и на данный момент, по-видимому, является единственной в мире системой, обеспечивающей решение всех этих задач на единой математической и технологической основе. При этом решение некоторых из этих задач по отдельности на данный момент автоматизировано только в системе «Эйдос», например автоматизированный когнитивный SWOT-анализ, когнитивный кластерно-конструктивный анализ, построение когнитивных диаграмм и когнитивных функций (Луценко 2017). Таким образом, развитый алгоритм принятия решений в интеллектуальных системах управления на основе АСК-анализа и реализуемый в системе «Эйдос», соответствует известному принципу дуального управления, предложенному в 50-х годах XX века в

теории самонастраивающихся и самообучающихся систем замечательным советским ученым Александром Ароновичем Фельдбаумом.

Необходимо отметить, что система «Эйдос» обеспечивает решение всех задач, решение которых необходимо для реализации предлагаемого алгоритма: обратной задачи прогнозирования (автоматизированный SWOT-анализ) [19]; кластерно-конструктивный анализ целевых состояний объекта управления и значений факторов [20]; задачи прогнозирования [9-38].

13.3. Практическое решение проблемы исследования в системе «Эйдос» на примере прогнозирования курсов акций компании Google и сценариев их изменения

13.3.1. Введение. Постановка цели и задач исследования

Задача, решаемая в данной работе, поставлена на портале Kaggle молодым исследователем из Индии Шриниди Хиппараги (https://www.kaggle.com/shreenidhihipparagi). Им же предоставлены и исходные данные для решения этой задачи: https://www.kaggle.com/shreenidhihipparagi/google-stock-prediction.

Шриниди Хиппараги пишет на портале Kaggle: «Все практики, изучающие DL, обязательно встретят RNN и LSTM. Поэтому я подумал, позвольте мне добавить набор данных, который можно использовать в качестве ступени к прогнозам акций.

Этот набор данных содержит 14 столбцов и 1257 строк. Каждый столбец назначается атрибуту, а строки содержат значения этого атрибута.

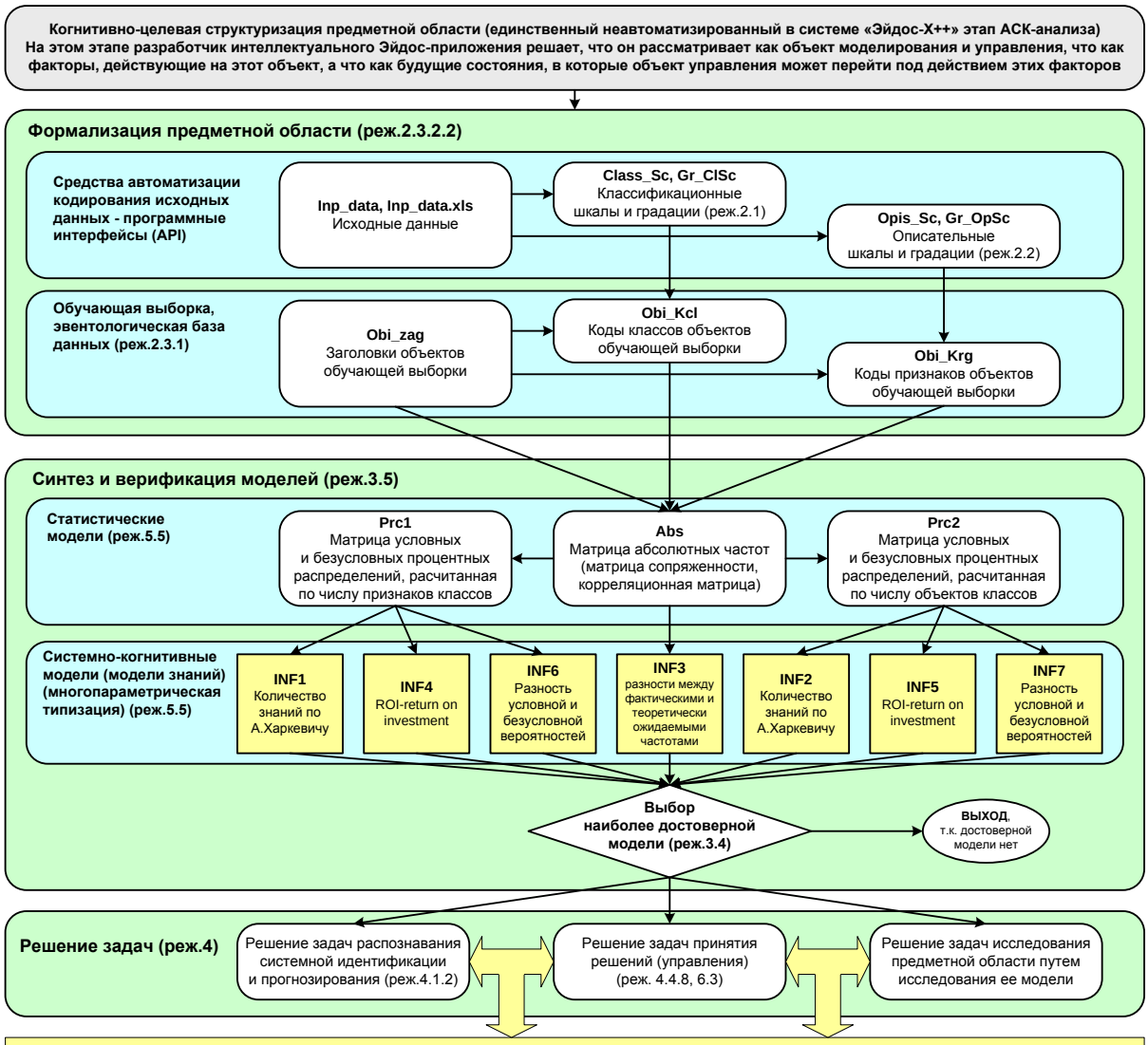
….. Я хотел бы поблагодарить Tiingo за предоставление такой замечательной платформы, которая поддерживает финансовые и биржевые данные и обновляет их изо дня в день.

Предскажите значения закрытия и открытия на следующие 30 дней. Вы можете это сделать?»

Таким образом, ставится **цель** прогнозирования значений закрытия и открытия акций на определенный период вперед в будущее.

В соответствии с последовательностью обработки данных, информации и знаний в системе «Эйдос» (рисунок 8) путем декомпозиции поставленной цели получена следующая последовательность ***задач***, решение которых является ***этапами*** достижения этой цели:

Последовательность обработки данных, информации и знаний в системе «Эйдос», повышение уровня системности данных, информации и знаний, повышение уровня системности моделей



Если модель адекватна, **достоверна**, т.е. соответствует действительности, то и результаты решения задач в этой модели также соответствуют действительности. Это значит, что если достоверность модели низка или неизвестна, то применять ее для решения реальных задач нельзя. Если же это делается, то является авантюризмом и профанацией науки.

О соотношении задач.

- распознавание, классификация, идентификация и диагностика (это одно и тоже, т.е. синонимы). При решении этих задач определяется степень сходства/различия образа конкретного объекта с обобщенными образами классов.
- идентификация и прогнозирование (при идентификации значения свойств и принадлежность объекта к классу относятся к одному моменту времени, а при прогнозировании значения факторов относятся к прошлому, а переход объекта под действием этих факторов в состояние, соответствующее классу относится к будущему, по сути, **прогнозирование - это идентификация будущих состояний, т.е. это тоже идентификация, но не в пространстве (настоящем), а в пространстве-времени**;
- прогнозирование и принятие решений (при прогнозировании по значениям факторов, действующих на объект моделирования, определяется в какое будущее состояние он перейдет под их действием. При принятии решений, наоборот, по будущему целевому состоянию объекта моделирования определяются значения факторов, которые обуславливают его переход в это будущее целевое состояние. Таким образом **задача принятия решений является обратной по отношению к задаче прогнозирования**);
- принятие решений путем многократного многовариантного прогнозирования при различных сочетаниях значений факторов невозможно из-за комбинаторного взрыва. Прогнозирование может быть **элементом** принятия решения, т.е. применено для оценки **адекватности** рассматриваемого уже ранее сформированного другим методом варианта решения, но оно в реальных случаях, т.е. когда много факторов, не может быть применено для выработки самого варианта решения;
- принятие решений и исследование моделируемой предметной области (задача принятия решений является обратной по отношению к задаче прогнозирования только в простейшем случае: в случае использования SWOT-анализа. Однако SWOT-анализ имеет свои ограничения: может быть задано только **одно** будущее целевое состояние, некоторые рекомендуемые факторы может **не быть** технологической и финансовой **возможности использовать**. Поэтому в АСК-анализе и системе «Эйдос» реализован развитый алгоритм принятия решений п.6.3 в котором кроме SWOT-анализа используются также результаты решения задачи прогнозирования и результаты кластерно-конструктивного анализа классов и значений факторов, т.е. некоторые результаты решения задачи исследования предметной области.)

Рисунок 8. Последовательность обработки данных, информации и знаний в системе «Эйдос»

Задача 1: когнитивная структуризация предметной области.

Задача 2: подготовка исходных данных и формализация предметной области (разработка классификационных и описательных шкал и градаций

и кодирование исходных с их помощью, т.е. получение обучающей выборки).

Задача 3: синтез и верификация статистических и системно-когнитивных моделей и выбор наиболее достоверной модели.

Задача 4: решение различных задач в наиболее достоверной модели:

– подзадача 4.1. Прогнозирование (диагностика, классификация, распознавание, идентификация);

– подзадача 4.2. Поддержка принятия решений (в упрощенном и развитом вариантах);

– подзадача 4.3. Исследование моделируемой предметной области путем исследования ее модели: когнитивные диаграммы классов и значений факторов, агломеративная когнитивная кластеризация классов и значений факторов, нелокальные нейроны и нейронные сети, 3d-интегральные когнитивные карты, когнитивные функции), исследование силы и направления влияния факторов и степени детерминированности классов, обуславливающими их значениями факторов.

В данной работе рассмотрим подробный численный пример в интеллектуальной системе «Эйдос». Эта система будет использована, т.к. в настоящее время именно она представляет собой программный инструментарий Автоматизированного системно-когнитивного анализа (АСК-анализ). Полная информация об АСК-анализе и системе «Эйдос» приведена в работах [1, 2].а также на сайте автора: http://lc.kubagro.ru/ и на портале РесчеГейт: https://www.researchgate.net/profile/Eugene-Lutsenko. Саму систему «Эйдос» также можно скачать на сайте автора: http://lc.kubagro.ru/aidos/_Aidos-X.htm.

В ходе рассмотрения численного примера решим поставленные выше задачи. При этом будем придерживаться (в упрощенном варианте) методики изложения, описанной в работе [3].

13.3.2. Задача 1: когнитивная структуризация предметной области

На этапе когнитивно-целевой структуризации предметной области мы неформализуемым путем решаем на качественном уровне, что будем рассматривать в качестве факторов, действующих на моделируемый объект (причин), а что в качестве результатов действия этих факторов (последствий). По сути это постановка решаемой проблемы.

Описательные шкалы служат для формального описания факторов, а классификационные – результатов их действия на объект моделирования. Шкалы могут быть числовые и текстовые [4].

При этом необходимо отметить, что статистические и системно-когнитивные модели (СК-модели) отражают лишь сам *факт* наличия зависимостей между значениями факторов и результатами их действия. Но они не отражают *причин и механизмов* такого влияния.

Более того, иногда *при моделировании встречается ситуация, когда на результаты влияют не сами рассматриваемые в модели факторы, а некие причины, влияющие на эти рассматриваемые в модели факторы.* Причем эти причины в модели вообще не упоминаются и рассматриваются.

Если исследовать зависимость поведения людей от положения стрелок часов, то получится довольно тесная взаимосвязь. Но это не означает, что существуют некие физические силы, типа сил гравитации, с помощью которых стрелки часов влияют на поведение людей. Все выглядит так, что человек посмотрел на часы, и стал что-то делать, что нужно в это время. На самом деле на поведение людей влияет положение Солнца над горизонтом, а не положение стрелок часов, а *часы просто адекватно отражают это положение Солнца, сообщают* ***информацию*** *об этом*. Аналогичная ситуация с геномом, который влияет и на почерк, и на успеваемость, поэтому почерк и успеваемость выглядят взаимосвязанными или влияющими друг на друга, хотя на самом деле они связаны не друг с другом, а с геномом.

Важно не перепутать местами причины и следствия: ветер дует не потому, что у деревьев шатаются ветки и дождь идет не потому, что это показывает приложение Gismeteo на телефоне или ласточки летают низко[29]. Английские ученые в результате исследования очень большой выборки респондентов из разных стран установили, что чем больше человек отпраздновал дней рождения, тем больше у него продолжительность жизни. На основе этого исследования они настоятельно рекомендовали как можно чаще праздновать дни рождения.

Система «Эйдос» выявляет ***эмпирические закономерности*** в моделируемой предметной области и отображает их в различных формах: табличной, графической и аналитической. Это соответствует эмпирическому этапу развития. Этим самым она вплотную подводит исследователя к теоретическому уровню познания [5][30]

Это значит:

– *во-первых,* что содержательная интерпретация СК-моделей – это компетенция специалистов-экспертов хорошо разбирающихся в данной предметной области. Иногда встречается ситуация, когда и то, что на первый взгляд является причинами, и то, что, казалось бы, является их последствиями, на самом деле является последствиями неких глубинных причин, которых мы не видим и никоим образом непосредственно не отражаем в модели;

[29] Хотя…

[30] См., также: http://lc.kubagro.ru/aidos/Works_on_identification_presentation_and_use_of_knowledge.htm

– *во-вторых,* даже если содержательной интерпретации обнаруженных эмпирических закономерностей не разработано, то в принципе это совершенно не исключает возможности эффективно пользоваться их знанием на *практике* для достижения заданных результатов и поставленных целей, т.е. для управления.

Данная работа основана на исходных данных, размещенных на портале Kaggle: https://www.kaggle.com/shreenidhihipparagi/google-stock-prediction:

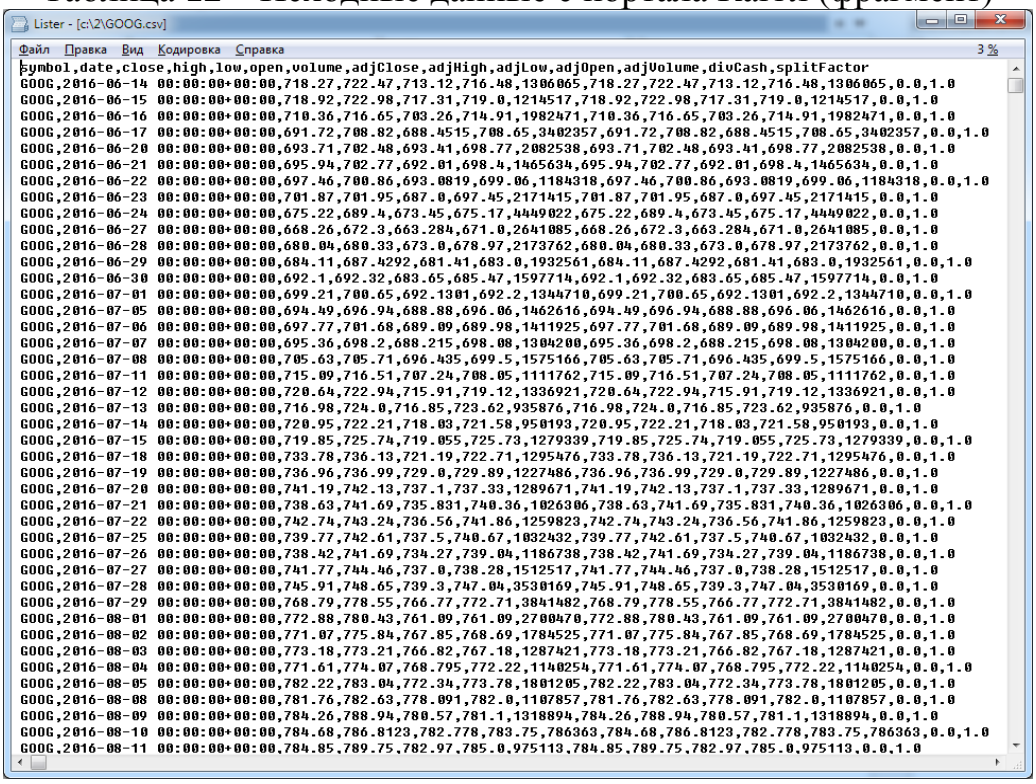
По описанию задачи, приведенному на портале Kaggle можно сделать вывод о том, что ее смысл состоит в том, чтобы по динамике значений различных показателей акций Гугл на финансовом рынке спрогнозировать курсы их открытия и закрытия на конец заданного периода.

Научное значение разработки методики подобных прогнозов состоит в том, что это довольно сложная задача, для которой пока не найдено качественного общего решения. И это не смотря на огромные усилия, в этом направлении, осуществляемые большим количеством специалистов очень высокой квалификации.

Практическое значение подобных прогнозов состоит в том, что на их основе можно принимать обоснованные решения о приобретении или продаже данных акций. Чем выше достоверность прогнозов, тем выше адекватность решений, тем выше прибыль от этой деятельности.

Исходные данные содержат следующие параметры (таблица 1):

Таблица 12 – Исходные данные с портала Каггл (фрагмент)



В таблице 1 классификационные шкалы поставлены начале таблицы, как принято в системе «Эйдос».

В данной работе в качестве классификационных шкал выберем начальную и конечную стоимость акций на день (выделены желтым фоном) (таблица 2), а в качестве факторов, влияющих на этот результаты – все остальные показатели (таблица 3):

Таблица 13

Классификационные шкалы

Код	Наименование
1	OPEN
2	CLOSE
3	OPEN-FUTURE3
4	CLOSE-FUTURE3
5	OPEN-FUTURE3-Point1
6	OPEN-FUTURE3-Point2
7	OPEN-FUTURE3-Point3
8	CLOSE-FUTURE3-Point1
9	CLOSE-FUTURE3-Point2
10	CLOSE-FUTURE3-Point3

Таблица 14

Описательные шкалы

Код	Наименование
1	HIGH
2	LOW
3	VOLUME
4	ADJCLOSE
5	ADJHIGH
6	ADJLOW
7	ADJOPEN
8	ADJVOLUME
9	HIGH-PAST3
10	LOW-PAST3
11	VOLUME-PAST3
12	ADJCLOSE-PAST3
13	ADJHIGH-PAST3
14	ADJLOW-PAST3
15	ADJOPEN-PAST3
16	ADJVOLUME-PAST3
17	HIGH-PAST3-Point1
18	HIGH-PAST3-Point2
19	HIGH-PAST3-Point3
20	LOW-PAST3-Point1
21	LOW-PAST3-Point2
22	LOW-PAST3-Point3
23	VOLUME-PAST3-Point1
24	VOLUME-PAST3-Point2
25	VOLUME-PAST3-Point3
26	ADJCLOSE-PAST3-Point1
27	ADJCLOSE-PAST3-Point2
28	ADJCLOSE-PAST3-Point3
29	ADJHIGH-PAST3-Point1
30	ADJHIGH-PAST3-Point2
31	ADJHIGH-PAST3-Point3
32	ADJLOW-PAST3-Point1
33	ADJLOW-PAST3-Point2
34	ADJLOW-PAST3-Point3
35	ADJOPEN-PAST3-Point1
36	ADJOPEN-PAST3-Point2
37	ADJOPEN-PAST3-Point3
38	ADJVOLUME-PAST3-Point1
39	ADJVOLUME-PAST3-Point2
40	ADJVOLUME-PAST3-Point3

В соответствии с методологией сценарного АСК-анализа [6, 7, 8, 9] кроме базовых классификационных и описательных шкал, непосредственно отражающих значения из таблицы 1, в модели используются еще и *автоматически* созданные на основе базовых шкал:

– сценарные шкалы, отражающие *динамику* изменения значений базовых показателей;

– шкалы, отражающие значения в заданных *точках* этих сценариев.

Смысл этих шкал, приведенных в таблицах 2 и 3, понятен из их названий.

13.3.3. Задача 2: подготовка исходных данных и формализация предметной области

13.3.3.1. Автоматизированный программный интерфейс (API) ввода числовых и текстовых данных и таблиц

Технически мы можем решить задачу прогнозирования не только на период 30, как просят на портале Kaggle, но и на значительно больший период. Но не будем этого делать и выберем на порядок меньший период прогнозирования всего в 3 дня. Мы это сделаем для уменьшения размерности задачи и удобства ее описания в полном виде в данной статье.

Исходные данные для данной работы (таблица 1) получены непосредственно с портала Kaggle по прямой ссылке: https://www.kaggle.com/shreenidhihipparagi/google-stock-prediction/download.

Эти данные представлены в виде CSV-файла. После скачивания этого файла для ввода в систему «Эйдос» с ним было выполнено несколько простых преобразований:

1. CSV-файл был переименован с «GOOG.csv» на «Inp_data.csv» и размещен в папке: ..\Aidos-X\AID_DATA\Inp_data\ системы «Эйдос» для исходных данных табличного типа.

2. CSV-файл был преобразован в XLS-файл для удобства дальнейшей корректировки и ввода в систему «Эйдос».

Само CSV-XLS преобразование (конвертирование) может быть осуществлено онлайн с помощью одного из онлайн-конвертеров. Рекомендуется использовать следующие CSV-XLS-онлайн конвертеры, которые очень хорошо работают со стандартными CSV-файлами:

https://convertio.co/ru/csv-xls/,
https://onlineconvertfree.com/ru/convert-format/csv-to-xls/;
https://document.online-convert.com/ru/convert/csv-to-excel.

В простейшем случае CSV-файл это текст, состоящий из строк, в каждой из которых содержится ***одинаковое*** количество элементов, разделенных каким-либо разделителем, чаще всего запятой. Таким образом, строки CSV-файла можно поставить в соответствие строкам таблицы, а элементы строк – колонкам таблицы.

Но следует иметь в виду, что сам CSV-стандарт (форматированный текст) еще не совсем устоялся. Но в CSV-файлах в качестве разделителя могут быть использованы и другие символы, например, точка с запятой или табуляция. Иногда, когда необходимо, чтобы внутри элементов использовалась запятая, эти элементы выделяют кавычками. Поэтому иногда (достаточно редко) встречаются CSV-файлы с необычными форматами, которые не всякий конвертер сможет корректно преобразовать. В этом случае рекомендуется попробовать подобрать другой конвертер, которых очень много в открытом доступе. Потратив на это некоторое время, обычно удается получить желаемый результат.

3. После преобразования CSV-файла в XLS-файл в нем средствами MS-Excel были произведены следующие корректировки:

– колонки: «open» и «close», советующие классификационным шкалам, были перемещены в начало таблицы и выделены желтым фоном;

– удалена колонка «symbol» с названием фирмы, т.к. в ней не было других фирм, кроме Гугл;

– XLS-файл (стандарт MS Excel-2003) записан в стандарте более новых версий MS Excel как XLSX. Это сделано потому, что в новом стандарте файл имеет размер примерно в два раза меньше, чем в старом.

В результате всех этих операций получилась таблица исходных данных (таблица 4):

Таблица 15– Исходные данные для ввода в систему «Эйдос» (фрагмент)

date	open	close	high	low	volume	adjClose	adjHigh	adjLow	adjOpen	adjVolume
2016-02-16 00:00:00+00:00	692,98	691	698	685,05	2520021	691	698	685,05	692,98	2520021
2016-02-17 00:00:00+00:00	699	708,4	709,75	691,38	2492634	708,4	709,75	691,38	699	2492634
2016-02-18 00:00:00+00:00	710	697,35	712,35	696,03	1883248	697,35	712,35	696,03	710	1883248
2016-02-19 00:00:00+00:00	695,03	700,91	703,0805	694,05	1589281	700,91	703,0805	694,05	695,03	1589281
2016-02-22 00:00:00+00:00	707,45	706,46	713,24	702,51	1949816	706,46	713,24	702,51	707,45	1949816
2016-02-23 00:00:00+00:00	701,45	695,85	708,4	693,58	2009280	695,85	708,4	693,58	701,45	2009280
2016-02-24 00:00:00+00:00	688,92	699,56	700	680,78	1963573	699,56	700	680,78	688,92	1963573
2016-02-25 00:00:00+00:00	700,01	705,75	705,98	690,585	1642166	705,75	705,98	690,585	700,01	1642166
2016-02-26 00:00:00+00:00	708,58	705,07	713,43	700,86	2243522	705,07	713,43	700,86	708,58	2243522
2016-02-29 00:00:00+00:00	700,32	697,77	710,89	697,68	2481145	697,77	710,89	697,68	700,32	2481145
2016-03-01 00:00:00+00:00	703,62	718,81	718,81	699,77	2151419	718,81	718,81	699,77	703,62	2151419
2016-03-02 00:00:00+00:00	719	718,85	720	712	1629003	718,85	720	712	719	1629003
2016-03-03 00:00:00+00:00	718,68	712,42	719,45	706,02	1957974	712,42	719,45	706,02	718,68	1957974
2016-03-04 00:00:00+00:00	714,99	710,89	716,49	706,02	1972077	710,89	716,49	706,02	714,99	1972077
2016-03-07 00:00:00+00:00	706,9	695,16	708,0912	686,9	2988026	695,16	708,0912	686,9	706,9	2988026
2016-03-08 00:00:00+00:00	688,59	693,97	703,79	685,34	2058471	693,97	703,79	685,34	688,59	2058471
2016-03-09 00:00:00+00:00	698,47	705,24	705,68	694	1421515	705,24	705,68	694	698,47	1421515
2016-03-10 00:00:00+00:00	708,12	712,82	716,44	703,36	2833525	712,82	716,44	703,36	708,12	2833525
2016-03-11 00:00:00+00:00	720	726,82	726,92	717,125	1970815	726,82	726,92	717,125	720	1970815
2016-03-14 00:00:00+00:00	726,81	730,49	735,5	725,15	1718252	730,49	735,5	725,15	726,81	1718252
2016-03-15 00:00:00+00:00	726,92	728,33	732,29	724,77	1720965	728,33	732,29	724,77	726,92	1720965
2016-03-16 00:00:00+00:00	726,37	736,09	737,47	724,51	1624370	736,09	737,47	724,51	726,37	1624370
2016-03-17 00:00:00+00:00	736,45	737,78	743,07	736	1860834	737,78	743,07	736	736,45	1860834
2016-03-18 00:00:00+00:00	741,86	737,6	742	731,83	2980709	737,6	742	731,83	741,86	2980709
2016-03-21 00:00:00+00:00	736,5	742,09	742,5	733,5157	1836503	742,09	742,5	733,5157	736,5	1836503
2016-03-22 00:00:00+00:00	737,46	740,75	745	737,46	1269749	740,75	745	737,46	737,46	1269749
2016-03-23 00:00:00+00:00	742,36	738,06	745,7199	736,15	1432099	738,06	745,7199	736,15	742,36	1432099
2016-03-24 00:00:00+00:00	732,01	735,3	737,747	731	1594891	735,3	737,747	731	732,01	1594891
2016-03-28 00:00:00+00:00	736,79	733,53	738,99	732,5	1301327	733,53	738,99	732,5	736,79	1301327
2016-03-29 00:00:00+00:00	734,59	744,77	747,25	728,76	1903758	744,77	747,25	728,76	734,59	1903758
2016-03-30 00:00:00+00:00	750,1	750,53	757,88	748,74	1782427	750,53	757,88	748,74	750,1	1782427
2016-03-31 00:00:00+00:00	749,25	744,95	750,85	740,94	1718798	744,95	750,85	740,94	749,25	1718798
2016-04-01 00:00:00+00:00	738,6	749,91	750,34	737	1576745	749,91	750,34	737	738,6	1576745

При разработке *реальных* научных интеллектуальных приложений убедительно рекомендуется в числовых колонах в обязательном порядке указывать единицы измерения, в нашем случае это доллары США, а также делать одинаковое число знаков после запятой в колонке. В данном случае мы этого не делали, чтобы сохранить полное совпадение названий базовых шкал с оригиналом на портале Kaggle.

Отметим, что в таблице 4 приведен лишь небольшой фрагмент исходных данных, т.к. в этой таблице 1259 строк. Полностью файл исходных данных можно скачать из Эйдос-облака по прямой ссылке: http://aidos.byethost5.com/Source_data_applications/Applications-000295/Inp_data.xlsx.

После подготовки таблицы исходных данных Inp_data.xlsx и размещения ее в папке для исходных данных: c:\Aidos-

X\AID_DATA\Inp_data запустим режим 2.3.2.2 системы «Эйдос» (рисунок 9), представляющий собой автоматизированный программный интерфейс (API) с внешними числовыми и текстовыми данными табличного типа. При этом используем параметры, приведенные на рисунке 9:

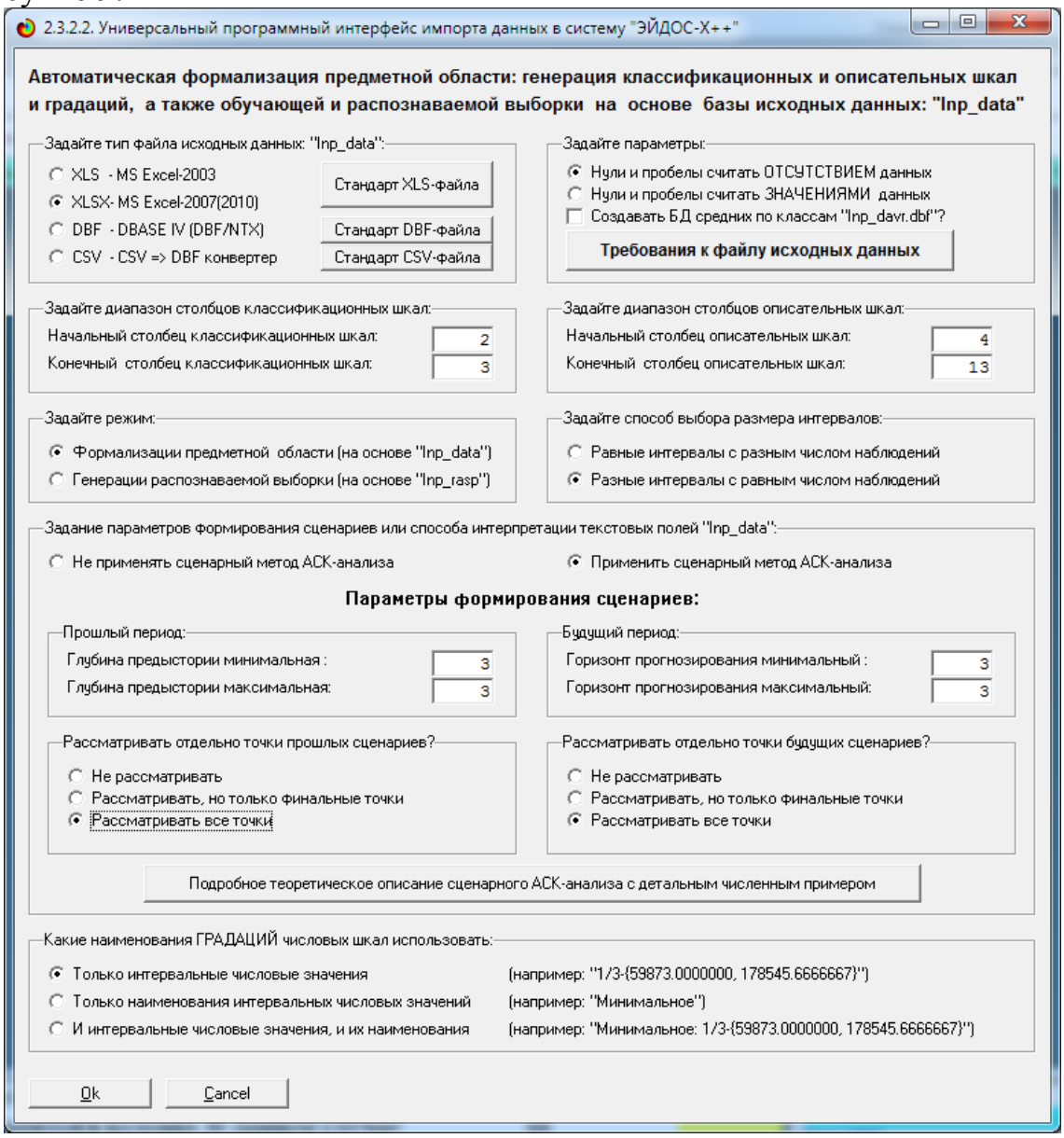
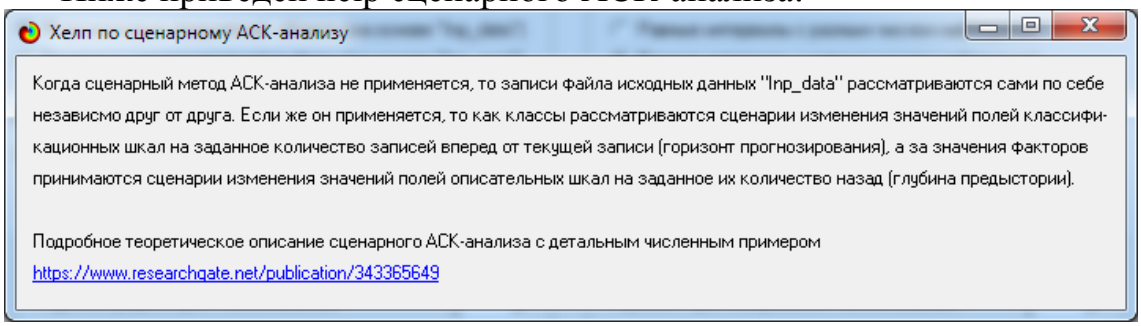


Рисунок 9. Экранная форма управления режимом 2.3.2.2 системы «Эйдос»

Ниже приведен help сценарного АСК-анализа:



Приведенная на этой экранной гиперссылка: https://www.researchgate.net/publication/343365649 является активной (действующей). По ней находится наиболее фундаментальная на данный момент опубликованная работа автора по сценарному АСК-анализу [6].

На рисунках 10 приведены экранные формы API- 2.3.2.2, отражающие последующие этапы выполнения этого режима:

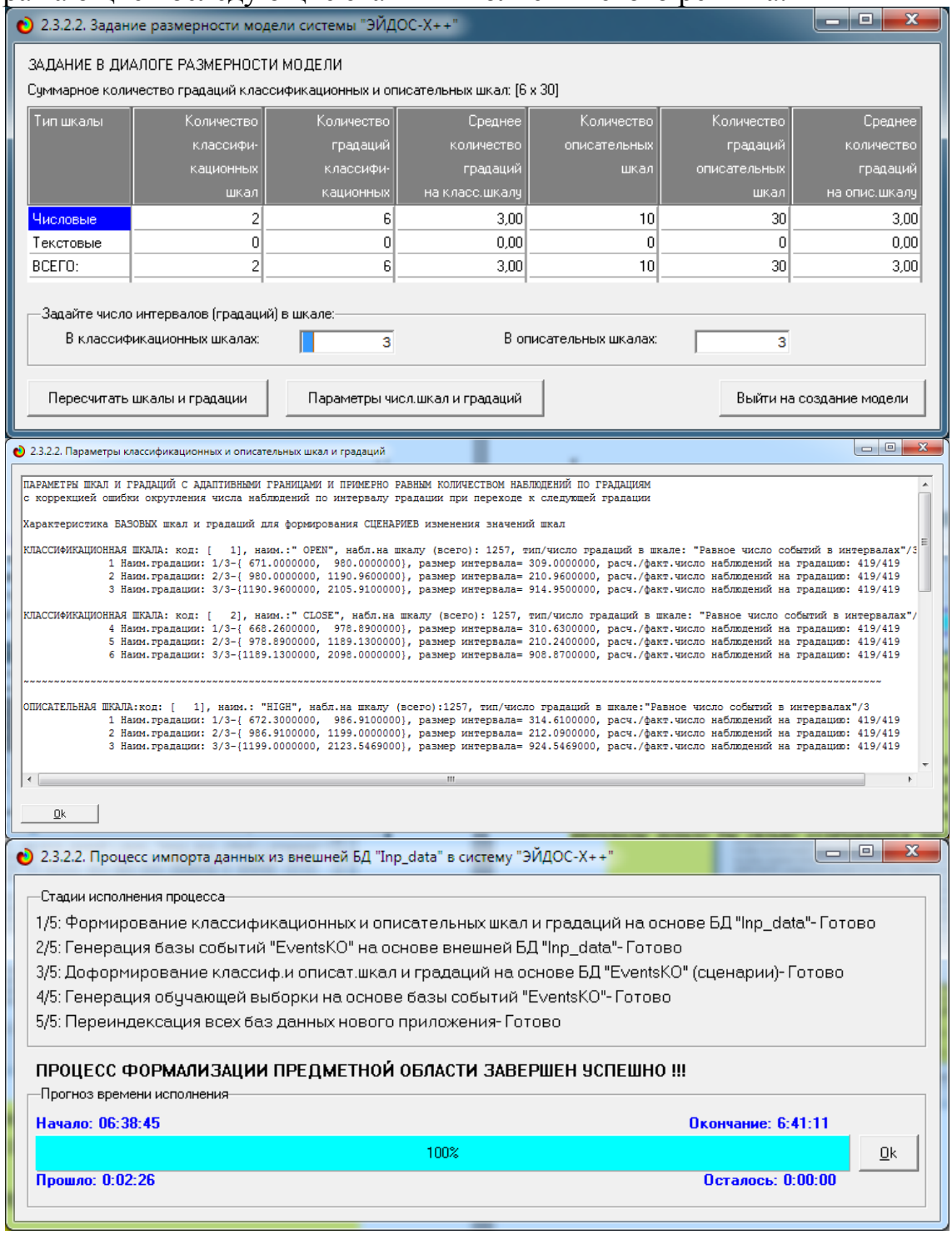


Рисунок 10. Экранные форма программного интерфейса (API) 2.3.2.2 системы «Эйдос» с внешними данными табличного типа

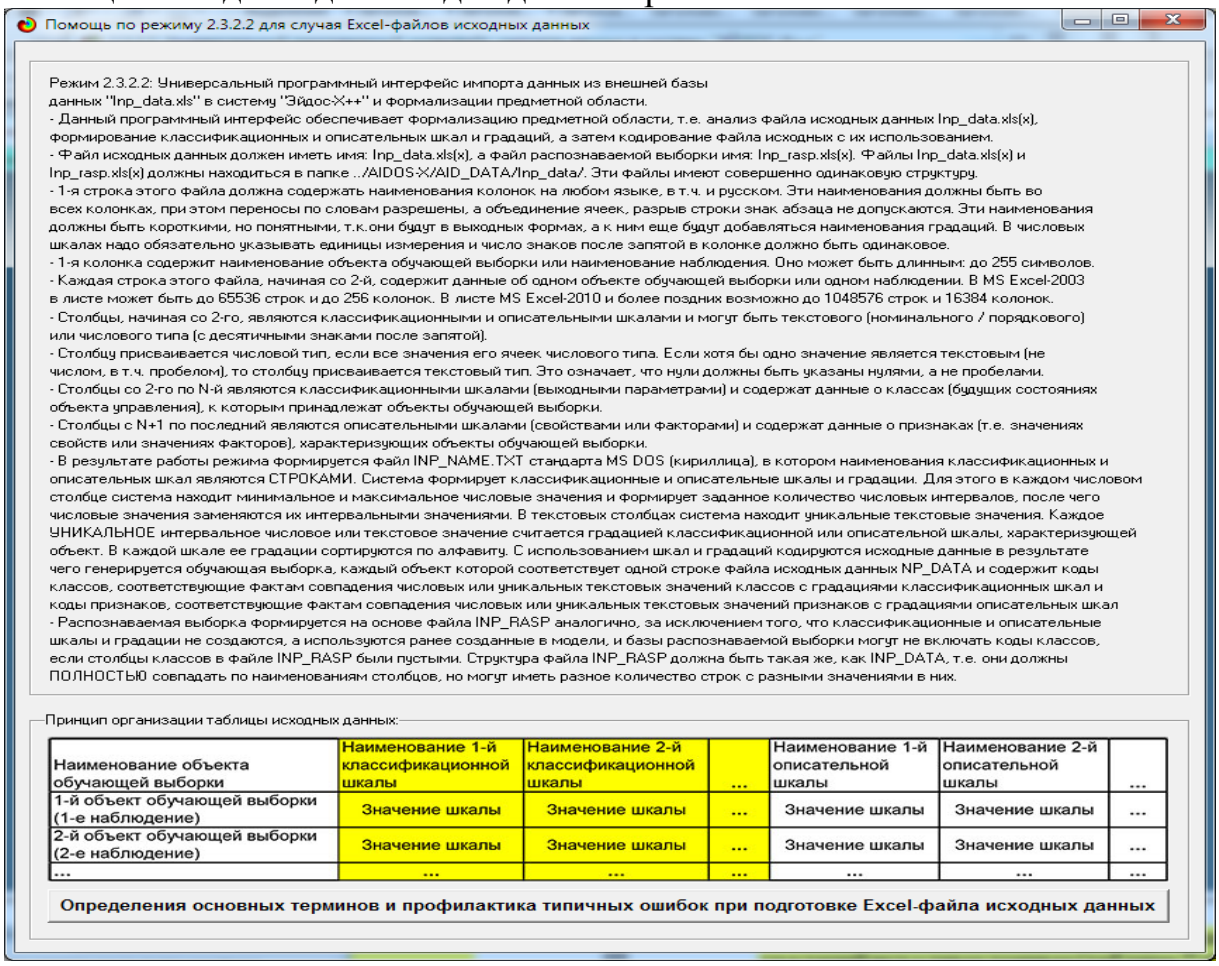
Как видно из рисунка 10 весь процесс вода исходных данных в систему «Эйдос» занял 2 минуты 26 секунд.

Обратим внимание на то, что заданы *адаптивные* интервалы, учитывающее неравномерность распределения данных по диапазону значений, что важно при относительно небольшом числе наблюдений. Если бы интервалы были заданы равными по величине, то в различные интервалы попало бы сильно отличающееся число наблюдений, а в некоторых интервалах их бы могло не оказаться вовсе.

Здесь же обратим внимание на то, что в таблице исходных данных (таблица 2) колонки содержать как числовые, так и текстовые значения. В шкалах текстового число числовых интервалов (диапазонов), естественно, не задается. В нашем случае в исходных данных текстовых колонок нет.

В классификационных и описательных шкалах задано 3 адаптивных числовых интервала. Как видно из рисунка 10 на каждое интервальное числовое значение приходится около 419 наблюдений, что более чем достаточно для того, чтобы можно было обосновано говорить о наличии достаточной статистики для обоснованных выводов.

На рисунке 11 приведены две экранные формы с исчерпывающим хелпом по API-2.3.2.2. В этом help объясняется принцип организации таблицы исходных данных для данного режима.



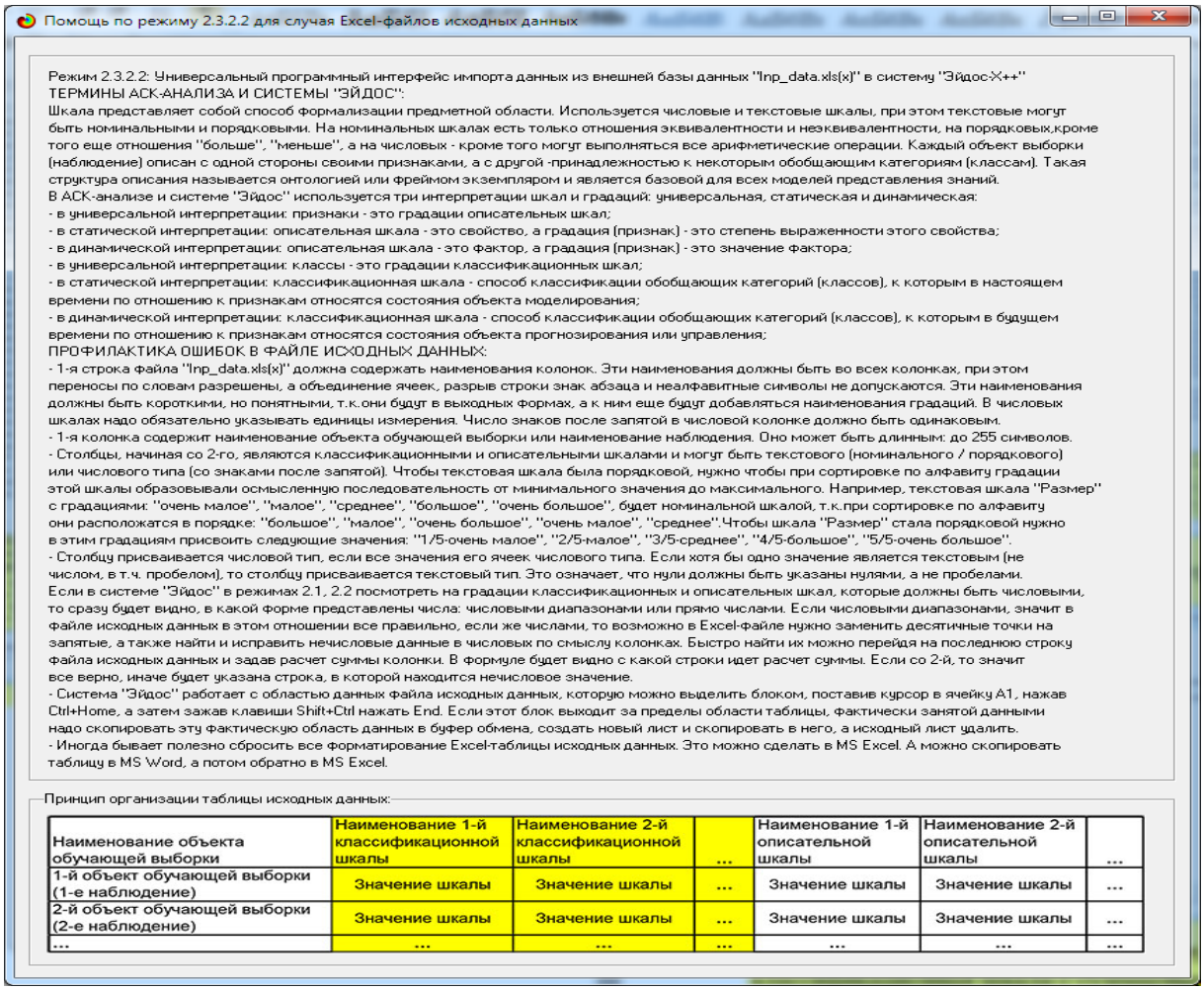


Рисунок 11. Экранные формы HELP универсального программного интерфейса API-2.3.2.2

После окончания работы API-2.3.2.2 выводится экранная форма, приведенная на рисунке 5. В этой экранной форме содержится информация об обнаружении в таблице исходных данных (таблица 1) колонок без вариабельности значений. В таблице 2 эти колонки не показаны.

13.3.3.2. Классификационные и описательные шкал и градации и обучающая выборка

В результате работы API-2.3.2.2 сформировано 10 классификационных шкал с суммарным количеством градаций (классов) 54 (рисунок 13, таблица 16) и 40 описательных шкал с суммарным числом градаций 238 (рисунок 14, таблица 17).

С использованием классификационных и описательных шкал и градаций исходные данные (таблица 15) были закодированы и в результате получена обучающая выборка (рисунок 15, таблица 18):

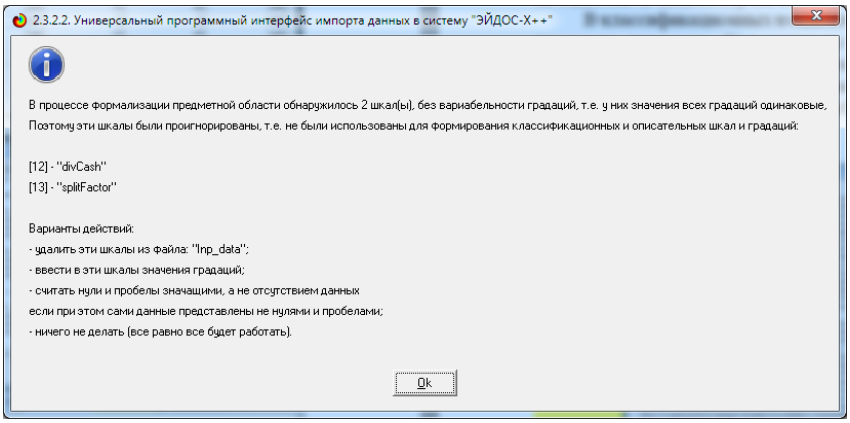


Рисунок 12. Экранная форма API-2.3.2.2 с информацией об обнаружении в таблице исходных данных колонок без вариабельности значений

Таблица 16 – Классификационные шкалы и градации

Код	Наименование
1	OPEN-1/3-{671.0, 980.0}
2	OPEN-2/3-{980.0, 1191.0}
3	OPEN-3/3-{1191.0, 2105.9}
4	CLOSE-1/3-{668.3, 978.9}
5	CLOSE-2/3-{978.9, 1189.1}
6	CLOSE-3/3-{1189.1, 2098.0}
7	OPEN-FUTURE3-OPEN-FUTURE3-1,1,1
8	OPEN-FUTURE3-OPEN-FUTURE3-1,1,2
9	OPEN-FUTURE3-OPEN-FUTURE3-1,2,1
10	OPEN-FUTURE3-OPEN-FUTURE3-1,2,2
11	OPEN-FUTURE3-OPEN-FUTURE3-2,1,1
12	OPEN-FUTURE3-OPEN-FUTURE3-2,1,2
13	OPEN-FUTURE3-OPEN-FUTURE3-2,2,1
14	OPEN-FUTURE3-OPEN-FUTURE3-2,2,2
15	OPEN-FUTURE3-OPEN-FUTURE3-2,2,3
16	OPEN-FUTURE3-OPEN-FUTURE3-2,3,2
17	OPEN-FUTURE3-OPEN-FUTURE3-2,3,3
18	OPEN-FUTURE3-OPEN-FUTURE3-3,2,2
19	OPEN-FUTURE3-OPEN-FUTURE3-3,2,3
20	OPEN-FUTURE3-OPEN-FUTURE3-3,3,2
21	OPEN-FUTURE3-OPEN-FUTURE3-3,3,3
22	CLOSE-FUTURE3-CLOSE-FUTURE3-4,4,4
23	CLOSE-FUTURE3-CLOSE-FUTURE3-4,4,5
24	CLOSE-FUTURE3-CLOSE-FUTURE3-4,5,4
25	CLOSE-FUTURE3-CLOSE-FUTURE3-4,5,5
26	CLOSE-FUTURE3-CLOSE-FUTURE3-5,4,4
27	CLOSE-FUTURE3-CLOSE-FUTURE3-5,4,5
28	CLOSE-FUTURE3-CLOSE-FUTURE3-5,5,4
29	CLOSE-FUTURE3-CLOSE-FUTURE3-5,5,5
30	CLOSE-FUTURE3-CLOSE-FUTURE3-5,5,6
31	CLOSE-FUTURE3-CLOSE-FUTURE3-5,6,5
32	CLOSE-FUTURE3-CLOSE-FUTURE3-5,6,6
33	CLOSE-FUTURE3-CLOSE-FUTURE3-6,5,5
34	CLOSE-FUTURE3-CLOSE-FUTURE3-6,5,6
35	CLOSE-FUTURE3-CLOSE-FUTURE3-6,6,5
36	CLOSE-FUTURE3-CLOSE-FUTURE3-6,6,6
37	OPEN-FUTURE3-POINT1-OPEN-FUTURE3-Point1-1/3-{671.0, 980.0}
38	OPEN-FUTURE3-POINT1-OPEN-FUTURE3-Point1-2/3-{980.0, 1191.0}
39	OPEN-FUTURE3-POINT1-OPEN-FUTURE3-Point1-3/3-{1191.0, 2105.9}
40	OPEN-FUTURE3-POINT2-OPEN-FUTURE3-Point2-1/3-{671.0, 980.0}
41	OPEN-FUTURE3-POINT2-OPEN-FUTURE3-Point2-2/3-{980.0, 1191.0}
42	OPEN-FUTURE3-POINT2-OPEN-FUTURE3-Point2-3/3-{1191.0, 2105.9}
43	**OPEN-FUTURE3-POINT3-OPEN-FUTURE3-Point3-1/3-{671.0, 980.0}**
44	**OPEN-FUTURE3-POINT3-OPEN-FUTURE3-Point3-2/3-{980.0, 1191.0}**
45	**OPEN-FUTURE3-POINT3-OPEN-FUTURE3-Point3-3/3-{1191.0, 2105.9}**
46	CLOSE-FUTURE3-POINT1-CLOSE-FUTURE3-Point1-1/3-{668.3, 978.9}
47	CLOSE-FUTURE3-POINT1-CLOSE-FUTURE3-Point1-2/3-{978.9, 1189.1}
48	CLOSE-FUTURE3-POINT1-CLOSE-FUTURE3-Point1-3/3-{1189.1, 2098.0}
49	CLOSE-FUTURE3-POINT2-CLOSE-FUTURE3-Point2-1/3-{668.3, 978.9}
50	CLOSE-FUTURE3-POINT2-CLOSE-FUTURE3-Point2-2/3-{978.9, 1189.1}
51	CLOSE-FUTURE3-POINT2-CLOSE-FUTURE3-Point2-3/3-{1189.1, 2098.0}
52	**CLOSE-FUTURE3-POINT3-CLOSE-FUTURE3-Point3-1/3-{668.3, 978.9}**
53	**CLOSE-FUTURE3-POINT3-CLOSE-FUTURE3-Point3-2/3-{978.9, 1189.1}**
54	**CLOSE-FUTURE3-POINT3-CLOSE-FUTURE3-Point3-3/3-{1189.1, 2098.0}**

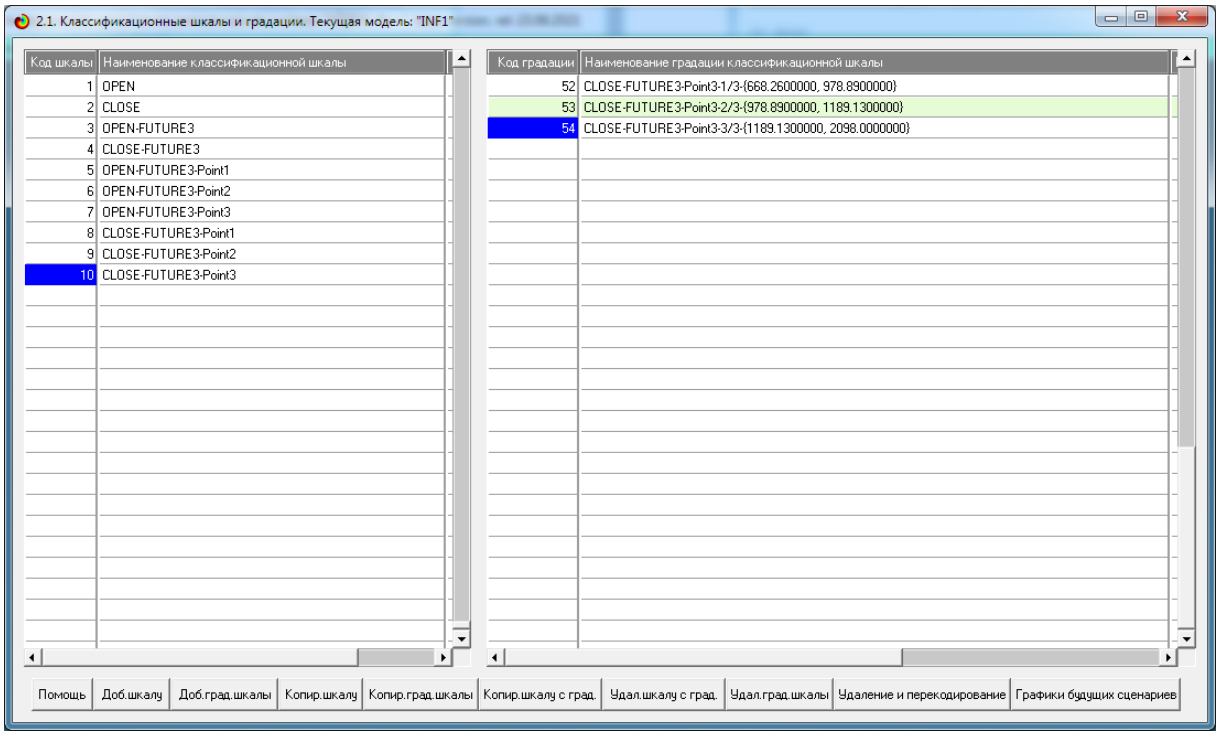


Рисунок 13. Экранная форма режима 2.1 системы «Эйдос»: классификационные шкалы и градации

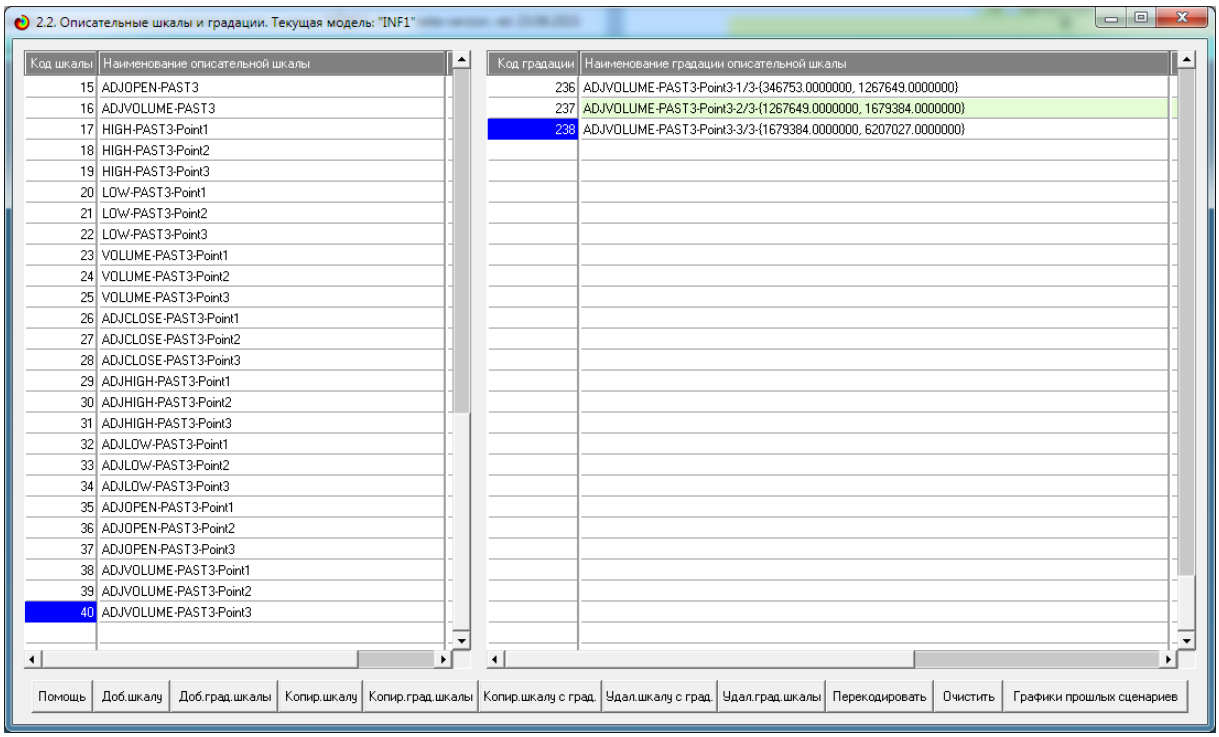


Рисунок 14. Экранная форма режима 2.2 системы «Эйдос»: описательные шкалы и градации

Таблица 17 – Описательные шкалы и градации

Код	Наименование
1	HIGH-1/3-{672.3000000, 986.9100000}
2	HIGH-2/3-{986.9100000, 1199.0000000}
3	HIGH-3/3-{1199.0000000, 2123.5469000}
4	LOW-1/3-{663.2840000, 972.2500000}
5	LOW-2/3-{972.2500000, 1181.1200000}
6	LOW-3/3-{1181.1200000, 2078.5400000}
7	VOLUME-1/3-{346753.0000000, 1267649.0000000}
8	VOLUME-2/3-{1267649.0000000, 1679384.0000000}
9	VOLUME-3/3-{1679384.0000000, 6207027.0000000}
10	ADJCLOSE-1/3-{668.2600000, 978.8900000}
11	ADJCLOSE-2/3-{978.8900000, 1189.1300000}
12	ADJCLOSE-3/3-{1189.1300000, 2098.0000000}
13	ADJHIGH-1/3-{672.3000000, 986.9100000}
14	ADJHIGH-2/3-{986.9100000, 1199.0000000}
15	ADJHIGH-3/3-{1199.0000000, 2123.5469000}
16	ADJLOW-1/3-{663.2840000, 972.2500000}
17	ADJLOW-2/3-{972.2500000, 1181.1200000}
18	ADJLOW-3/3-{1181.1200000, 2078.5400000}
19	ADJOPEN-1/3-{671.0000000, 980.0000000}
20	ADJOPEN-2/3-{980.0000000, 1190.9600000}
21	ADJOPEN-3/3-{1190.9600000, 2105.9100000}
22	ADJVOLUME-1/3-{346753.0000000, 1267649.0000000}
23	ADJVOLUME-2/3-{1267649.0000000, 1679384.0000000}
24	ADJVOLUME-3/3-{1679384.0000000, 6207027.0000000}
25	HIGH-PAST3-HIGH-PAST3-01,01,01
26	HIGH-PAST3-HIGH-PAST3-01,01,02
27	HIGH-PAST3-HIGH-PAST3-01,02,01
28	HIGH-PAST3-HIGH-PAST3-01,02,02
29	HIGH-PAST3-HIGH-PAST3-02,01,01
30	HIGH-PAST3-HIGH-PAST3-02,02,01
31	HIGH-PAST3-HIGH-PAST3-02,02,02
32	HIGH-PAST3-HIGH-PAST3-02,02,03
33	HIGH-PAST3-HIGH-PAST3-02,03,02
34	HIGH-PAST3-HIGH-PAST3-02,03,03
35	HIGH-PAST3-HIGH-PAST3-03,02,02
36	HIGH-PAST3-HIGH-PAST3-03,02,03
37	HIGH-PAST3-HIGH-PAST3-03,03,02
38	HIGH-PAST3-HIGH-PAST3-03,03,03
39	LOW-PAST3-LOW-PAST3-04,04,04
40	LOW-PAST3-LOW-PAST3-04,04,05
41	LOW-PAST3-LOW-PAST3-04,05,04
42	LOW-PAST3-LOW-PAST3-04,05,05
43	LOW-PAST3-LOW-PAST3-05,04,04
44	LOW-PAST3-LOW-PAST3-05,04,05
45	LOW-PAST3-LOW-PAST3-05,05,04
46	LOW-PAST3-LOW-PAST3-05,05,05
47	LOW-PAST3-LOW-PAST3-05,05,06
48	LOW-PAST3-LOW-PAST3-05,06,05
49	LOW-PAST3-LOW-PAST3-05,06,06
50	LOW-PAST3-LOW-PAST3-06,05,05
51	LOW-PAST3-LOW-PAST3-06,05,06
52	LOW-PAST3-LOW-PAST3-06,06,05
53	LOW-PAST3-LOW-PAST3-06,06,06
54	VOLUME-PAST3-VOLUME-PAST3-07,07,07
55	VOLUME-PAST3-VOLUME-PAST3-07,07,08
56	VOLUME-PAST3-VOLUME-PAST3-07,07,09
57	VOLUME-PAST3-VOLUME-PAST3-07,08,07
58	VOLUME-PAST3-VOLUME-PAST3-07,08,08
59	VOLUME-PAST3-VOLUME-PAST3-07,08,09
60	VOLUME-PAST3-VOLUME-PAST3-07,09,07
61	VOLUME-PAST3-VOLUME-PAST3-07,09,08
62	VOLUME-PAST3-VOLUME-PAST3-07,09,09
63	VOLUME-PAST3-VOLUME-PAST3-08,07,07
64	VOLUME-PAST3-VOLUME-PAST3-08,07,08
65	VOLUME-PAST3-VOLUME-PAST3-08,07,09
66	VOLUME-PAST3-VOLUME-PAST3-08,08,07
67	VOLUME-PAST3-VOLUME-PAST3-08,08,08
68	VOLUME-PAST3-VOLUME-PAST3-08,08,09
69	VOLUME-PAST3-VOLUME-PAST3-08,09,07
70	VOLUME-PAST3-VOLUME-PAST3-08,09,08
71	VOLUME-PAST3-VOLUME-PAST3-08,09,09
72	VOLUME-PAST3-VOLUME-PAST3-09,07,07
73	VOLUME-PAST3-VOLUME-PAST3-09,07,08
74	VOLUME-PAST3-VOLUME-PAST3-09,07,09
75	VOLUME-PAST3-VOLUME-PAST3-09,08,07
76	VOLUME-PAST3-VOLUME-PAST3-09,08,08
77	VOLUME-PAST3-VOLUME-PAST3-09,08,09
78	VOLUME-PAST3-VOLUME-PAST3-09,09,07
79	VOLUME-PAST3-VOLUME-PAST3-09,09,08
80	VOLUME-PAST3-VOLUME-PAST3-09,09,09
81	ADJCLOSE-PAST3-ADJCLOSE-PAST3-10,10,10
82	ADJCLOSE-PAST3-ADJCLOSE-PAST3-10,10,11
83	ADJCLOSE-PAST3-ADJCLOSE-PAST3-10,11,10
84	ADJCLOSE-PAST3-ADJCLOSE-PAST3-10,11,11
85	ADJCLOSE-PAST3-ADJCLOSE-PAST3-11,10,10
86	ADJCLOSE-PAST3-ADJCLOSE-PAST3-11,10,11
87	ADJCLOSE-PAST3-ADJCLOSE-PAST3-11,11,10
88	ADJCLOSE-PAST3-ADJCLOSE-PAST3-11,11,11
89	ADJCLOSE-PAST3-ADJCLOSE-PAST3-11,11,12
90	ADJCLOSE-PAST3-ADJCLOSE-PAST3-11,12,11
91	ADJCLOSE-PAST3-ADJCLOSE-PAST3-11,12,12

92	ADJCLOSE-PAST3-ADJCLOSE-PAST3-12,11,11
93	ADJCLOSE-PAST3-ADJCLOSE-PAST3-12,11,12
94	ADJCLOSE-PAST3-ADJCLOSE-PAST3-12,12,11
95	ADJCLOSE-PAST3-ADJCLOSE-PAST3-12,12,12
96	ADJHIGH-PAST3-ADJHIGH-PAST3-13,13,13
97	ADJHIGH-PAST3-ADJHIGH-PAST3-13,13,14
98	ADJHIGH-PAST3-ADJHIGH-PAST3-13,14,13
99	ADJHIGH-PAST3-ADJHIGH-PAST3-13,14,14
100	ADJHIGH-PAST3-ADJHIGH-PAST3-14,13,13
101	ADJHIGH-PAST3-ADJHIGH-PAST3-14,14,13
102	ADJHIGH-PAST3-ADJHIGH-PAST3-14,14,14
103	ADJHIGH-PAST3-ADJHIGH-PAST3-14,14,15
104	ADJHIGH-PAST3-ADJHIGH-PAST3-14,15,14
105	ADJHIGH-PAST3-ADJHIGH-PAST3-14,15,15
106	ADJHIGH-PAST3-ADJHIGH-PAST3-15,14,14
107	ADJHIGH-PAST3-ADJHIGH-PAST3-15,14,15
108	ADJHIGH-PAST3-ADJHIGH-PAST3-15,15,14
109	ADJHIGH-PAST3-ADJHIGH-PAST3-15,15,15
110	ADJLOW-PAST3-ADJLOW-PAST3-16,16,16
111	ADJLOW-PAST3-ADJLOW-PAST3-16,16,17
112	ADJLOW-PAST3-ADJLOW-PAST3-16,17,16
113	ADJLOW-PAST3-ADJLOW-PAST3-16,17,17
114	ADJLOW-PAST3-ADJLOW-PAST3-17,16,16
115	ADJLOW-PAST3-ADJLOW-PAST3-17,16,17
116	ADJLOW-PAST3-ADJLOW-PAST3-17,17,16
117	ADJLOW-PAST3-ADJLOW-PAST3-17,17,17
118	ADJLOW-PAST3-ADJLOW-PAST3-17,17,18
119	ADJLOW-PAST3-ADJLOW-PAST3-17,18,17
120	ADJLOW-PAST3-ADJLOW-PAST3-17,18,18
121	ADJLOW-PAST3-ADJLOW-PAST3-18,17,17
122	ADJLOW-PAST3-ADJLOW-PAST3-18,17,18
123	ADJLOW-PAST3-ADJLOW-PAST3-18,18,17
124	ADJLOW-PAST3-ADJLOW-PAST3-18,18,18
125	ADJOPEN-PAST3-ADJOPEN-PAST3-19,19,19
126	ADJOPEN-PAST3-ADJOPEN-PAST3-19,19,20
127	ADJOPEN-PAST3-ADJOPEN-PAST3-19,20,19
128	ADJOPEN-PAST3-ADJOPEN-PAST3-19,20,20
129	ADJOPEN-PAST3-ADJOPEN-PAST3-20,19,19
130	ADJOPEN-PAST3-ADJOPEN-PAST3-20,19,20
131	ADJOPEN-PAST3-ADJOPEN-PAST3-20,20,19
132	ADJOPEN-PAST3-ADJOPEN-PAST3-20,20,20
133	ADJOPEN-PAST3-ADJOPEN-PAST3-20,20,21
134	ADJOPEN-PAST3-ADJOPEN-PAST3-20,21,20
135	ADJOPEN-PAST3-ADJOPEN-PAST3-20,21,21
136	ADJOPEN-PAST3-ADJOPEN-PAST3-21,20,20
137	ADJOPEN-PAST3-ADJOPEN-PAST3-21,20,21
138	ADJOPEN-PAST3-ADJOPEN-PAST3-21,21,20
139	ADJOPEN-PAST3-ADJOPEN-PAST3-21,21,21
140	ADJVOLUME-PAST3-ADJVOLUME-PAST3-22,22,22
141	ADJVOLUME-PAST3-ADJVOLUME-PAST3-22,22,23
142	ADJVOLUME-PAST3-ADJVOLUME-PAST3-22,22,24
143	ADJVOLUME-PAST3-ADJVOLUME-PAST3-22,23,22
144	ADJVOLUME-PAST3-ADJVOLUME-PAST3-22,23,23
145	ADJVOLUME-PAST3-ADJVOLUME-PAST3-22,23,24
146	ADJVOLUME-PAST3-ADJVOLUME-PAST3-22,24,22
147	ADJVOLUME-PAST3-ADJVOLUME-PAST3-22,24,23
148	ADJVOLUME-PAST3-ADJVOLUME-PAST3-22,24,24
149	ADJVOLUME-PAST3-ADJVOLUME-PAST3-23,22,22
150	ADJVOLUME-PAST3-ADJVOLUME-PAST3-23,22,23
151	ADJVOLUME-PAST3-ADJVOLUME-PAST3-23,22,24
152	ADJVOLUME-PAST3-ADJVOLUME-PAST3-23,23,22
153	ADJVOLUME-PAST3-ADJVOLUME-PAST3-23,23,23
154	ADJVOLUME-PAST3-ADJVOLUME-PAST3-23,23,24
155	ADJVOLUME-PAST3-ADJVOLUME-PAST3-23,24,22
156	ADJVOLUME-PAST3-ADJVOLUME-PAST3-23,24,23
157	ADJVOLUME-PAST3-ADJVOLUME-PAST3-23,24,24
158	ADJVOLUME-PAST3-ADJVOLUME-PAST3-24,22,22
159	ADJVOLUME-PAST3-ADJVOLUME-PAST3-24,22,23
160	ADJVOLUME-PAST3-ADJVOLUME-PAST3-24,22,24
161	ADJVOLUME-PAST3-ADJVOLUME-PAST3-24,23,22
162	ADJVOLUME-PAST3-ADJVOLUME-PAST3-24,23,23
163	ADJVOLUME-PAST3-ADJVOLUME-PAST3-24,23,24
164	ADJVOLUME-PAST3-ADJVOLUME-PAST3-24,24,22
165	ADJVOLUME-PAST3-ADJVOLUME-PAST3-24,24,23
166	ADJVOLUME-PAST3-ADJVOLUME-PAST3-24,24,24
167	HIGH-PAST3-POINT1-HIGH-PAST3-Point1-1/3-{672.3000000, 986.9100000}
168	HIGH-PAST3-POINT1-HIGH-PAST3-Point1-2/3-{986.9100000, 1199.0000000}
169	HIGH-PAST3-POINT1-HIGH-PAST3-Point1-3/3-{1199.0000000, 2123.5469000}
170	HIGH-PAST3-POINT2-HIGH-PAST3-Point2-1/3-{672.3000000, 986.9100000}
171	HIGH-PAST3-POINT2-HIGH-PAST3-Point2-2/3-{986.9100000, 1199.0000000}
172	HIGH-PAST3-POINT2-HIGH-PAST3-Point2-3/3-{1199.0000000, 2123.5469000}
173	HIGH-PAST3-POINT3-HIGH-PAST3-Point3-1/3-{672.3000000, 986.9100000}
174	HIGH-PAST3-POINT3-HIGH-PAST3-Point3-2/3-{986.9100000, 1199.0000000}
175	HIGH-PAST3-POINT3-HIGH-PAST3-Point3-3/3-{1199.0000000, 2123.5469000}
176	LOW-PAST3-POINT1-LOW-PAST3-Point1-1/3-{663.2840000, 972.2500000}
177	LOW-PAST3-POINT1-LOW-PAST3-Point1-2/3-{972.2500000, 1181.1200000}
178	LOW-PAST3-POINT1-LOW-PAST3-Point1-3/3-{1181.1200000, 2078.5400000}
179	LOW-PAST3-POINT2-LOW-PAST3-Point2-1/3-{663.2840000, 972.2500000}
180	LOW-PAST3-POINT2-LOW-PAST3-Point2-2/3-{972.2500000, 1181.1200000}
181	LOW-PAST3-POINT2-LOW-PAST3-Point2-3/3-{1181.1200000, 2078.5400000}
182	LOW-PAST3-POINT3-LOW-PAST3-Point3-1/3-{663.2840000, 972.2500000}
183	LOW-PAST3-POINT3-LOW-PAST3-Point3-2/3-{972.2500000, 1181.1200000}
184	LOW-PAST3-POINT3-LOW-PAST3-Point3-3/3-{1181.1200000, 2078.5400000}
185	VOLUME-PAST3-POINT1-VOLUME-PAST3-Point1-1/3-{346753.0000000, 1267649.0000000}

186	VOLUME-PAST3-POINT1-VOLUME-PAST3-Point1-2/3-{1267649.0000000, 1679384.0000000}
187	VOLUME-PAST3-POINT1-VOLUME-PAST3-Point1-3/3-{1679384.0000000, 6207027.0000000}
188	VOLUME-PAST3-POINT2-VOLUME-PAST3-Point2-1/3-{346753.0000000, 1267649.0000000}
189	VOLUME-PAST3-POINT2-VOLUME-PAST3-Point2-2/3-{1267649.0000000, 1679384.0000000}
190	VOLUME-PAST3-POINT2-VOLUME-PAST3-Point2-3/3-{1679384.0000000, 6207027.0000000}
191	VOLUME-PAST3-POINT3-VOLUME-PAST3-Point3-1/3-{346753.0000000, 1267649.0000000}
192	VOLUME-PAST3-POINT3-VOLUME-PAST3-Point3-2/3-{1267649.0000000, 1679384.0000000}
193	VOLUME-PAST3-POINT3-VOLUME-PAST3-Point3-3/3-{1679384.0000000, 6207027.0000000}
194	ADJCLOSE-PAST3-POINT1-ADJCLOSE-PAST3-Point1-1/3-{668.2600000, 978.8900000}
195	ADJCLOSE-PAST3-POINT1-ADJCLOSE-PAST3-Point1-2/3-{978.8900000, 1189.1300000}
196	ADJCLOSE-PAST3-POINT1-ADJCLOSE-PAST3-Point1-3/3-{1189.1300000, 2098.0000000}
197	ADJCLOSE-PAST3-POINT2-ADJCLOSE-PAST3-Point2-1/3-{668.2600000, 978.8900000}
198	ADJCLOSE-PAST3-POINT2-ADJCLOSE-PAST3-Point2-2/3-{978.8900000, 1189.1300000}
199	ADJCLOSE-PAST3-POINT2-ADJCLOSE-PAST3-Point2-3/3-{1189.1300000, 2098.0000000}
200	ADJCLOSE-PAST3-POINT3-ADJCLOSE-PAST3-Point3-1/3-{668.2600000, 978.8900000}
201	ADJCLOSE-PAST3-POINT3-ADJCLOSE-PAST3-Point3-2/3-{978.8900000, 1189.1300000}
202	ADJCLOSE-PAST3-POINT3-ADJCLOSE-PAST3-Point3-3/3-{1189.1300000, 2098.0000000}
203	ADJHIGH-PAST3-POINT1-ADJHIGH-PAST3-Point1-1/3-{672.3000000, 986.9100000}
204	ADJHIGH-PAST3-POINT1-ADJHIGH-PAST3-Point1-2/3-{986.9100000, 1199.0000000}
205	ADJHIGH-PAST3-POINT1-ADJHIGH-PAST3-Point1-3/3-{1199.0000000, 2123.5469000}
206	ADJHIGH-PAST3-POINT2-ADJHIGH-PAST3-Point2-1/3-{672.3000000, 986.9100000}
207	ADJHIGH-PAST3-POINT2-ADJHIGH-PAST3-Point2-2/3-{986.9100000, 1199.0000000}
208	ADJHIGH-PAST3-POINT2-ADJHIGH-PAST3-Point2-3/3-{1199.0000000, 2123.5469000}
209	ADJHIGH-PAST3-POINT3-ADJHIGH-PAST3-Point3-1/3-{672.3000000, 986.9100000}
210	ADJHIGH-PAST3-POINT3-ADJHIGH-PAST3-Point3-2/3-{986.9100000, 1199.0000000}
211	ADJHIGH-PAST3-POINT3-ADJHIGH-PAST3-Point3-3/3-{1199.0000000, 2123.5469000}
212	ADJLOW-PAST3-POINT1-ADJLOW-PAST3-Point1-1/3-{663.2840000, 972.2500000}
213	ADJLOW-PAST3-POINT1-ADJLOW-PAST3-Point1-2/3-{972.2500000, 1181.1200000}
214	ADJLOW-PAST3-POINT1-ADJLOW-PAST3-Point1-3/3-{1181.1200000, 2078.5400000}
215	ADJLOW-PAST3-POINT2-ADJLOW-PAST3-Point2-1/3-{663.2840000, 972.2500000}
216	ADJLOW-PAST3-POINT2-ADJLOW-PAST3-Point2-2/3-{972.2500000, 1181.1200000}
217	ADJLOW-PAST3-POINT2-ADJLOW-PAST3-Point2-3/3-{1181.1200000, 2078.5400000}
218	ADJLOW-PAST3-POINT3-ADJLOW-PAST3-Point3-1/3-{663.2840000, 972.2500000}
219	ADJLOW-PAST3-POINT3-ADJLOW-PAST3-Point3-2/3-{972.2500000, 1181.1200000}
220	ADJLOW-PAST3-POINT3-ADJLOW-PAST3-Point3-3/3-{1181.1200000, 2078.5400000}
221	ADJOPEN-PAST3-POINT1-ADJOPEN-PAST3-Point1-1/3-{671.0000000, 980.0000000}
222	ADJOPEN-PAST3-POINT1-ADJOPEN-PAST3-Point1-2/3-{980.0000000, 1190.9600000}
223	ADJOPEN-PAST3-POINT1-ADJOPEN-PAST3-Point1-3/3-{1190.9600000, 2105.9100000}
224	ADJOPEN-PAST3-POINT2-ADJOPEN-PAST3-Point2-1/3-{671.0000000, 980.0000000}
225	ADJOPEN-PAST3-POINT2-ADJOPEN-PAST3-Point2-2/3-{980.0000000, 1190.9600000}
226	ADJOPEN-PAST3-POINT2-ADJOPEN-PAST3-Point2-3/3-{1190.9600000, 2105.9100000}
227	ADJOPEN-PAST3-POINT3-ADJOPEN-PAST3-Point3-1/3-{671.0000000, 980.0000000}
228	ADJOPEN-PAST3-POINT3-ADJOPEN-PAST3-Point3-2/3-{980.0000000, 1190.9600000}
229	ADJOPEN-PAST3-POINT3-ADJOPEN-PAST3-Point3-3/3-{1190.9600000, 2105.9100000}
230	ADJVOLUME-PAST3-POINT1-ADJVOLUME-PAST3-Point1-1/3-{346753.0000000, 1267649.0000000}
231	ADJVOLUME-PAST3-POINT1-ADJVOLUME-PAST3-Point1-2/3-{1267649.0000000, 1679384.0000000}
232	ADJVOLUME-PAST3-POINT1-ADJVOLUME-PAST3-Point1-3/3-{1679384.0000000, 6207027.0000000}
233	ADJVOLUME-PAST3-POINT2-ADJVOLUME-PAST3-Point2-1/3-{346753.0000000, 1267649.0000000}
234	ADJVOLUME-PAST3-POINT2-ADJVOLUME-PAST3-Point2-2/3-{1267649.0000000, 1679384.0000000}
235	ADJVOLUME-PAST3-POINT2-ADJVOLUME-PAST3-Point2-3/3-{1679384.0000000, 6207027.0000000}
236	ADJVOLUME-PAST3-POINT3-ADJVOLUME-PAST3-Point3-1/3-{346753.0000000, 1267649.0000000}
237	ADJVOLUME-PAST3-POINT3-ADJVOLUME-PAST3-Point3-2/3-{1267649.0000000, 1679384.0000000}
238	**ADJVOLUME-PAST3-POINT3-ADJVOLUME-PAST3-Point3-3/3-{1679384.0000000, 6207027.0000000}**

Таблица 18 – Обучающая выборка (фрагмент)

NAME_OBJ	N2	N3	N4	N5	N6	N7	N8	N9	N10	N11
2016-02-16 00:00:00+00:00	1	4	1	4	9	10	13	16	19	24
2016-02-17 00:00:00+00:00	1	4	1	4	9	10	13	16	19	24
2016-02-18 00:00:00+00:00	1	4	1	4	9	10	13	16	19	24
2016-02-19 00:00:00+00:00	1	4	1	4	8	10	13	16	19	23
2016-02-22 00:00:00+00:00	1	4	1	4	9	10	13	16	19	24
2016-02-23 00:00:00+00:00	1	4	1	4	9	10	13	16	19	24
2016-02-24 00:00:00+00:00	1	4	1	4	9	10	13	16	19	24
2016-02-25 00:00:00+00:00	1	4	1	4	8	10	13	16	19	23
2016-02-26 00:00:00+00:00	1	4	1	4	9	10	13	16	19	24
2016-02-29 00:00:00+00:00	1	4	1	4	9	10	13	16	19	24
2016-03-01 00:00:00+00:00	1	4	1	4	9	10	13	16	19	24
2016-03-02 00:00:00+00:00	1	4	1	4	8	10	13	16	19	23
2016-03-03 00:00:00+00:00	1	4	1	4	9	10	13	16	19	24
2016-03-04 00:00:00+00:00	1	4	1	4	9	10	13	16	19	24
2016-03-07 00:00:00+00:00	1	4	1	4	9	10	13	16	19	24
2016-03-08 00:00:00+00:00	1	4	1	4	9	10	13	16	19	24
2016-03-09 00:00:00+00:00	1	4	1	4	8	10	13	16	19	23
2016-03-10 00:00:00+00:00	1	4	1	4	9	10	13	16	19	24
2016-03-11 00:00:00+00:00	1	4	1	4	9	10	13	16	19	24
2016-03-14 00:00:00+00:00	1	4	1	4	9	10	13	16	19	24
2016-03-15 00:00:00+00:00	1	4	1	4	9	10	13	16	19	24
2016-03-16 00:00:00+00:00	1	4	1	4	8	10	13	16	19	23
2016-03-17 00:00:00+00:00	1	4	1	4	9	10	13	16	19	24
2016-03-18 00:00:00+00:00	1	4	1	4	9	10	13	16	19	24
2016-03-21 00:00:00+00:00	1	4	1	4	9	10	13	16	19	24
2016-03-22 00:00:00+00:00	1	4	1	4	8	10	13	16	19	23
2016-03-23 00:00:00+00:00	1	4	1	4	8	10	13	16	19	23
2016-03-24 00:00:00+00:00	1	4	1	4	8	10	13	16	19	23
2016-03-28 00:00:00+00:00	1	4	1	4	8	10	13	16	19	23
2016-03-29 00:00:00+00:00	1	4	1	4	9	10	13	16	19	24
2016-03-30 00:00:00+00:00	1	4	1	4	9	10	13	16	19	24

Обучающая выборка (таблица 18), по сути, представляет собой нормализованные исходные данные, т.е. таблицу исходных данных

(таблица 15), закодированную с помощью классификационных и описательных шкал и градаций (таблицы 16 и 17).

На рисунке 8 мы видим, что в обучающей выборке присутствуют коды градаций не только базовых классификационных и описательных шкал, но и сценарных шкал, и шкал, отражающих значения точек сценариев. Все они формируются в системе «Эйдос» автоматически по заданным параметрам непосредственно на основе исходных данных (таблица 2).

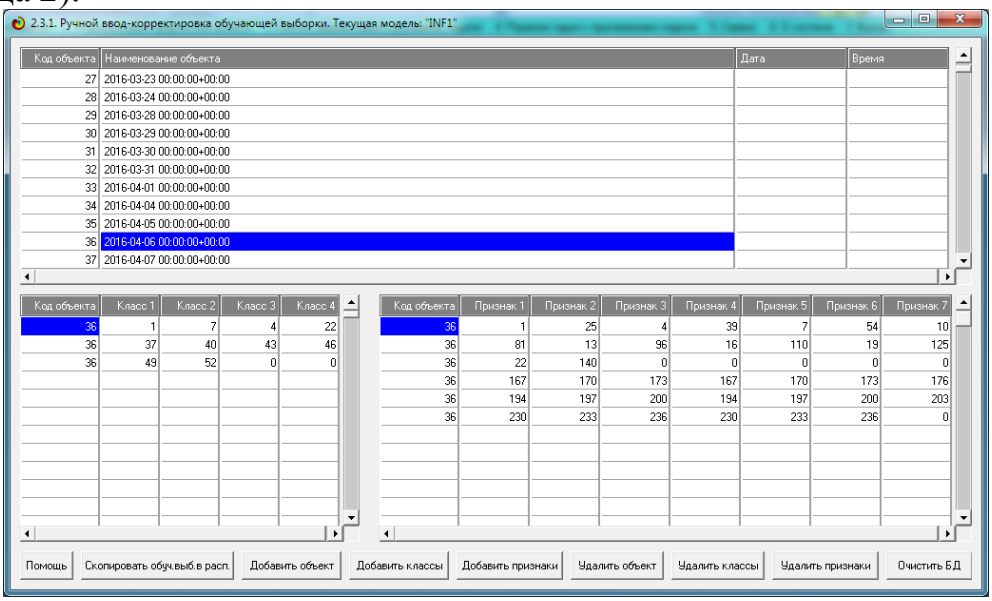


Рисунок 15. Экранная форма режима 2.3.1 системы «Эйдос»: обучающая выборка

Таким образом, в результате формализации предметной области созданы все необходимые и достаточные условия для выполнения следующего этапа сценарного АСК-анализа: т.е. для синтеза и верификации моделей.

13.3.3.3. Будущие и прошлые сценарии изменения значений градаций базовых шкал

Будущие сценарии изменения значений градаций базовых классификационных шкал в графическом виде можно получить, кликнув по самой провой кнопке на экранной форме, приведенной на рисунке 13. Сами изображения будущих сценариев приведены на рисунке 16.

Прошлые сценарии изменения значений градаций базовых описательных шкал в графическом виде можно получить кликнув по самой провой кнопке на экранной форме, приведенной на рисунке 14. Сами изображения прошлых сценариев приведены на рисунке 17.

Все приведенные сценарии формируются по заданным в API-2.3.2.2 параметрам (рисунок 9) полностью автоматически. Эти сценарии являются градациями соответствующих классификационных и описательных шкал,

которые формируются также автоматически по этим параметрам (таблицы 16, 17). Все сценарии автоматически кодируются и учитываются в обучающей выборке (таблица 18, рисунок 15).

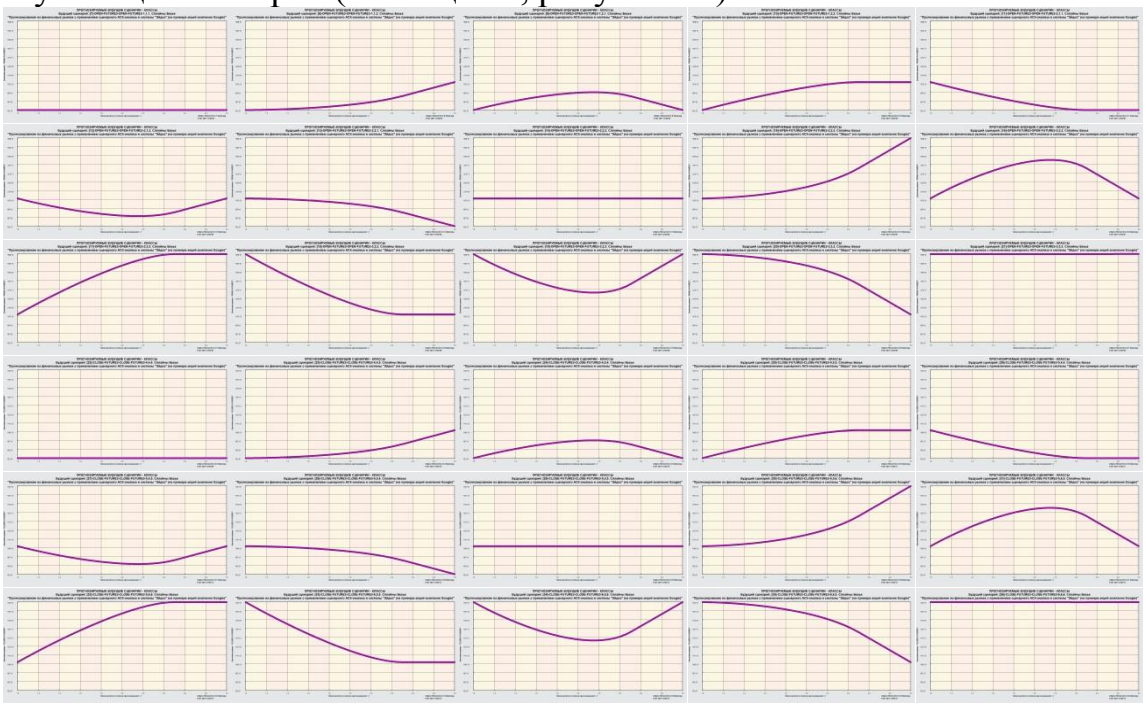
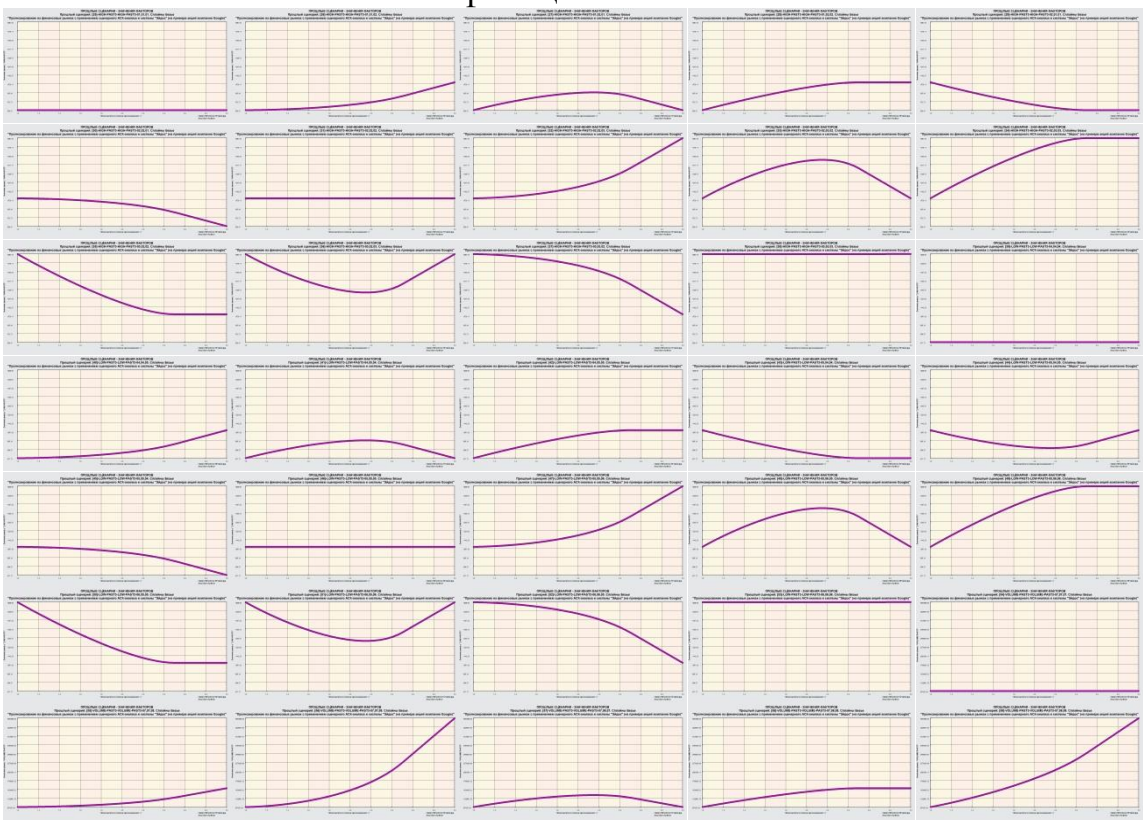
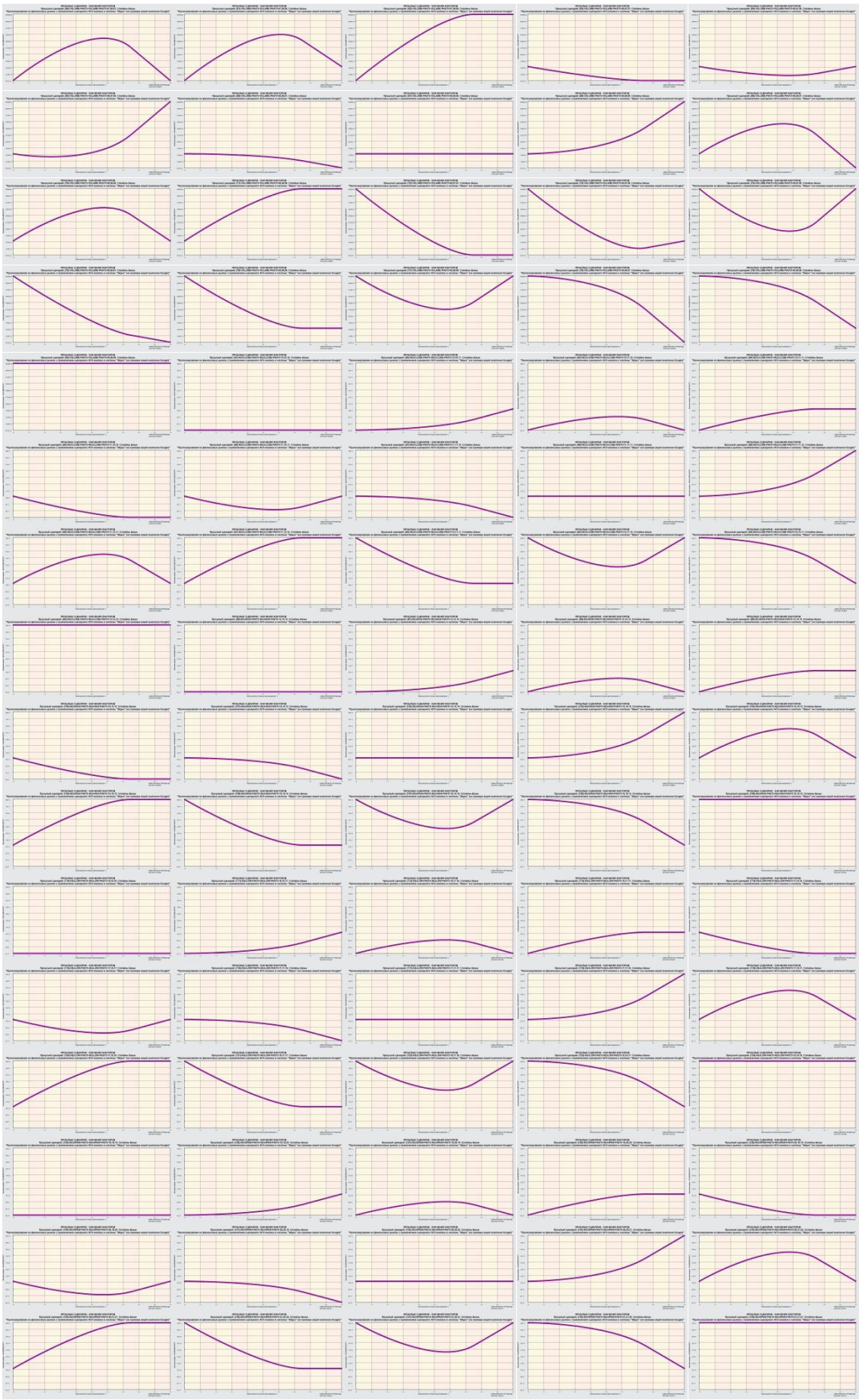


Рисунок 16. Будущие сценарии изменения значений градаций базовых классификационных шкал[31]



[31] Не смотря на малый размер рисунков в работе они вполне читабельны при просмотре текста работы в увеличенном масштабе, например при масштабе 200% или 500%.



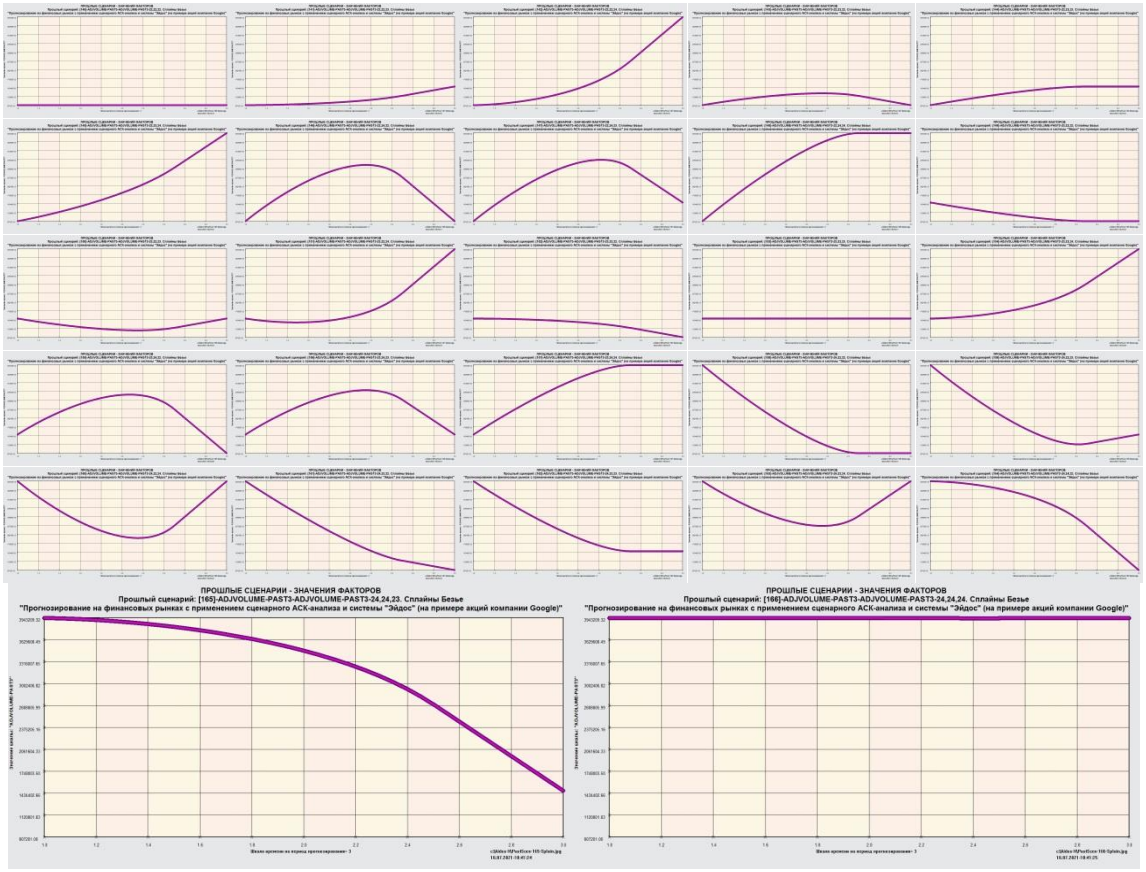


Рисунок 17. Прошлые сценарии изменения значений градаций базовых описательных шкал[32]

13.3.4. Задача 3: синтез и верификация моделей и выбор наиболее достоверной модели

13.3.4.1. Синтез и верификация статистических и системно-когнитивных моделей

Синтез и верификация статистических и системно-когнитивных моделей (СК-моделей) моделей осуществляется в режиме 3.5 системы «Эйдос» (рисунки 1 и 11). Математические модели, на основе которых рассчитываются статистические и СК-модели, приведены в работе [10].

Обратим внимание на то, что на рисунке 9 в правом нижнем углу окна задана опция: «Расчеты проводить на графическом процессоре (GPU)». Из рисунка 11 видно, что весь процесс синтеза и верификации моделей занял 11 минут 1 секунду. Отметим, что при синтезе и верификации моделей использовался графический процессор (GPU) видеокарты. Точнее использовалось 1500 шейдерных процессоров видеокарты NVIDIA GeForce GTX 770. Для расчета 10 выходных форм по результатам распознавания использовался центральный процессор (CPU)

[32] Не смотря на малый размер рисунков в работе они вполне читабельны при просмотре текста работы в увеличенном масштабе, например при масштабе 200% или 500%. Отметим также, что автор не форматировал размеры рисунков вручную, а использовал для этого стандартные возможности ворда.

i7. В основном время было затрачено именно на расчет выходных форм. На центральном процессоре (CPU) выполнение этих операций занимает значительно большее время (на некоторых задачах это происходит в десятки, сотни и даже тысячи раз дольше). Таким образом, неграфические вычисления на графических процессорах видеокарты делает возможной обработку больших объемов исходных данных за разумное время. В процессе синтеза и верификации моделей осуществляется также расчет 10 выходных форм и оценка достоверности моделей путем решения задачи идентификации объектов обучающей выборки, на что уходит более 99% времени исполнения.

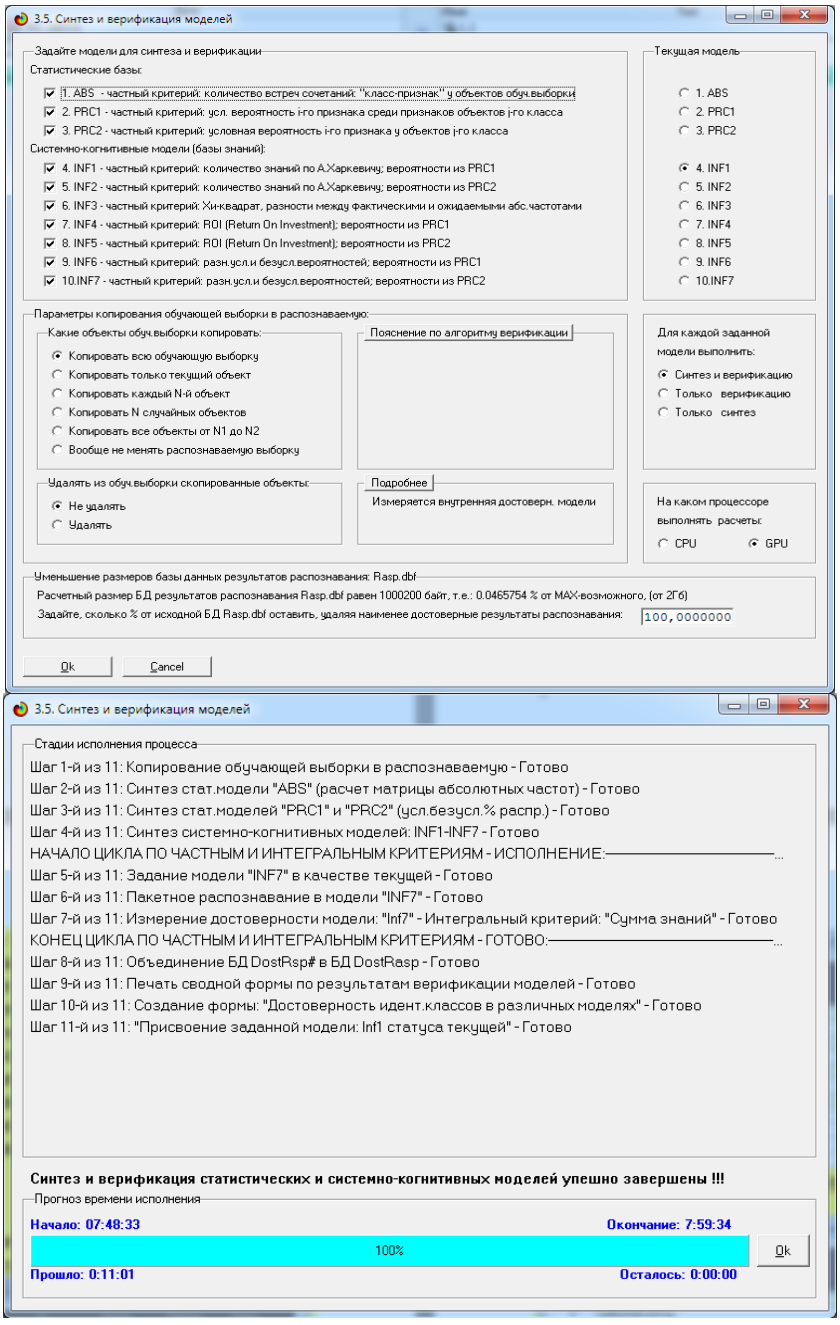
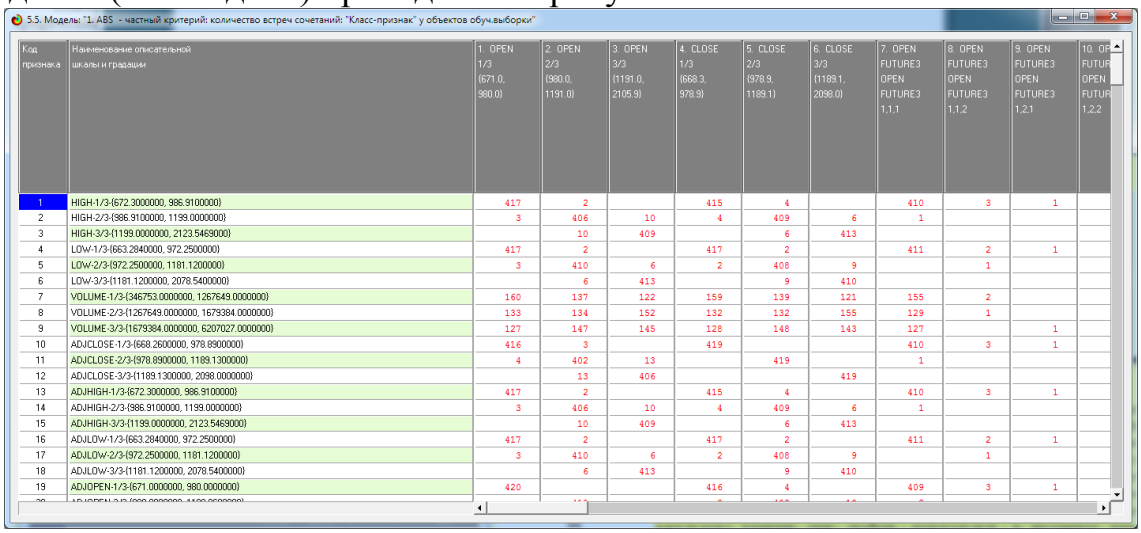


Рисунок 18. Экранные формы режима синтеза и верификации моделей системы «Эйдос» (режим 3.5)

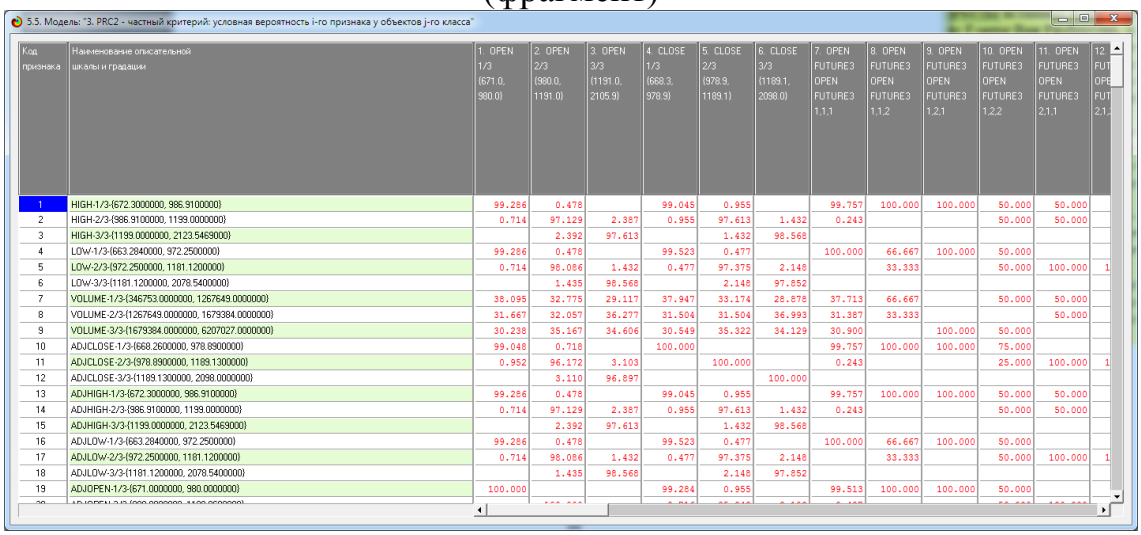
Некоторые из созданных статистических и системно-когнитивных моделей (СК-модели) приведены на рисунках 12 – 15.



5.5. Модель: "1. ABS - частный критерий: количество встреч сочетаний: "Класс-признак" у объектов обуч.выборки"

Код признака	Наименование описательной шкалы и градации	1. OPEN 1/3 (671.0, 980.0)	2. OPEN 2/3 (980.0, 1191.0)	3. OPEN 3/3 (1191.0, 2105.9)	4. CLOSE 1/3 (668.3, 978.9)	5. CLOSE 2/3 (978.9, 1189.1)	6. CLOSE 3/3 (1189.1, 2098.0)	7. OPEN FUTURE3 OPEN FUTURE3 1,1,1	8. OPEN FUTURE3 OPEN FUTURE3 1,1,2	9. OPEN FUTURE3 OPEN FUTURE3 1,2,1	10. OP FUTUR OPEN FUTUR 1,2,2
1	HIGH-1/3-(672.3000000, 986.9100000)	417	2		415	4		410	3	1	
2	HIGH-2/3-(986.9100000, 1199.0000000)	3	406	10	4	409	6	1			
3	HIGH-3/3-(1199.0000000, 2123.5469000)		10	409		6	413				
4	LOW-1/3-(663.2840000, 972.2500000)	417	2		417	2		411	2	1	
5	LOW-2/3-(972.2500000, 1181.1200000)	3	410	6	2	408	9		1		
6	LOW-3/3-(1181.1200000, 2078.5400000)		6	413		9	410				
7	VOLUME-1/3-(346753.0000000, 1267649.0000000)	160	137	122	159	139	121	155	2		
8	VOLUME-2/3-(1267649.0000000, 1679384.0000000)	133	134	152	132	132	155	129	1		
9	VOLUME-3/3-(1679384.0000000, 6207027.0000000)	127	147	145	128	148	143	127		1	
10	ADJCLOSE-1/3-(668.2600000, 978.8900000)	416	3		419			410	3	1	
11	ADJCLOSE-2/3-(978.8900000, 1189.1300000)	4	402	13		419		1			
12	ADJCLOSE-3/3-(1189.1300000, 2098.0000000)		13	406			419				
13	ADJHIGH-1/3-(672.3000000, 986.9100000)	417	2		415	4		410	3	1	
14	ADJHIGH-2/3-(986.9100000, 1199.0000000)	3	406	10	4	409	6	1			
15	ADJHIGH-3/3-(1199.0000000, 2123.5469000)		10	409		6	413				
16	ADJLOW-1/3-(663.2840000, 972.2500000)	417	2		417	2		411	2	1	
17	ADJLOW-2/3-(972.2500000, 1181.1200000)	3	410	6	2	408	9		1		
18	ADJLOW-3/3-(1181.1200000, 2078.5400000)		6	413		9	410				
19	ADJOPEN-1/3-(671.0000000, 980.0000000)	420			416	4		409	3	1	

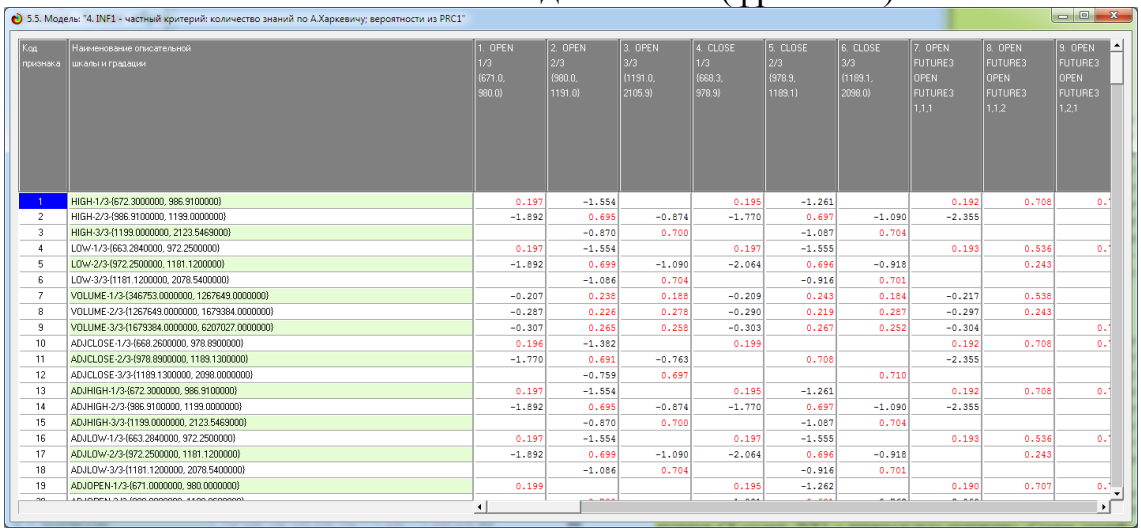
Рисунок 19. Матрица абсолютных частот: статистическая модель ABS (фрагмент)



5.5. Модель: "3. PRC2 - частный критерий: условная вероятность i-го признака у объектов j-го класса"

Код признака	Наименование описательной шкалы и градации	1. OPEN 1/3 (671.0, 980.0)	2. OPEN 2/3 (980.0, 1191.0)	3. OPEN 3/3 (1191.0, 2105.9)	4. CLOSE 1/3 (668.3, 978.9)	5. CLOSE 2/3 (978.9, 1189.1)	6. CLOSE 3/3 (1189.1, 2098.0)	7. OPEN FUTURE3 OPEN FUTURE3 1,1,1	8. OPEN FUTURE3 OPEN FUTURE3 1,1,2	9. OPEN FUTURE3 OPEN FUTURE3 1,2,1	10. OPEN FUTURE3 OPEN FUTURE3 1,2,2	11. OPEN FUTURE3 OPEN FUTURE3 2,1,1	12 FUT OPE FUT 2,1
1	HIGH-1/3-(672.3000000, 986.9100000)	99.286	0.478		99.045	0.955		99.757	100.000	100.000	50.000	50.000	
2	HIGH-2/3-(986.9100000, 1199.0000000)	0.714	97.129	2.387	0.955	97.613	1.432	0.243			50.000	50.000	
3	HIGH-3/3-(1199.0000000, 2123.5469000)		2.392	97.613		1.432	98.568						
4	LOW-1/3-(663.2840000, 972.2500000)	99.286	0.478		99.523	0.477		100.000	66.667	100.000	50.000		
5	LOW-2/3-(972.2500000, 1181.1200000)	0.714	98.086	1.432	0.477	97.375	2.148		33.333		50.000	100.000	1
6	LOW-3/3-(1181.1200000, 2078.5400000)		1.435	98.568		2.148	97.852						
7	VOLUME-1/3-(346753.0000000, 1267649.0000000)	38.095	32.775	29.117	37.947	33.174	28.878	37.713	66.667		50.000	50.000	
8	VOLUME-2/3-(1267649.0000000, 1679384.0000000)	31.667	32.057	36.277	31.504	31.504	36.993	31.387	33.333			50.000	
9	VOLUME-3/3-(1679384.0000000, 6207027.0000000)	30.238	35.167	34.606	30.549	35.322	34.129	30.900		100.000	50.000		
10	ADJCLOSE-1/3-(668.2600000, 978.8900000)	99.048	0.718		100.000			99.757	100.000	100.000	75.000		
11	ADJCLOSE-2/3-(978.8900000, 1189.1300000)	0.952	96.172	3.103		100.000		0.243			25.000	100.000	1
12	ADJCLOSE-3/3-(1189.1300000, 2098.0000000)		3.110	96.897			100.000						
13	ADJHIGH-1/3-(672.3000000, 986.9100000)	99.286	0.478		99.045	0.955		99.757	100.000	100.000	50.000	50.000	
14	ADJHIGH-2/3-(986.9100000, 1199.0000000)	0.714	97.129	2.387	0.955	97.613	1.432	0.243			50.000	50.000	
15	ADJHIGH-3/3-(1199.0000000, 2123.5469000)		2.392	97.613		1.432	98.568						
16	ADJLOW-1/3-(663.2840000, 972.2500000)	99.286	0.478		99.523	0.477		100.000	66.667	100.000	50.000		
17	ADJLOW-2/3-(972.2500000, 1181.1200000)	0.714	98.086	1.432	0.477	97.375	2.148		33.333		50.000	100.000	1
18	ADJLOW-3/3-(1181.1200000, 2078.5400000)		1.435	98.568		2.148	97.852						
19	ADJOPEN-1/3-(671.0000000, 980.0000000)	100.000			99.284	0.955		99.513	100.000	100.000	50.000		

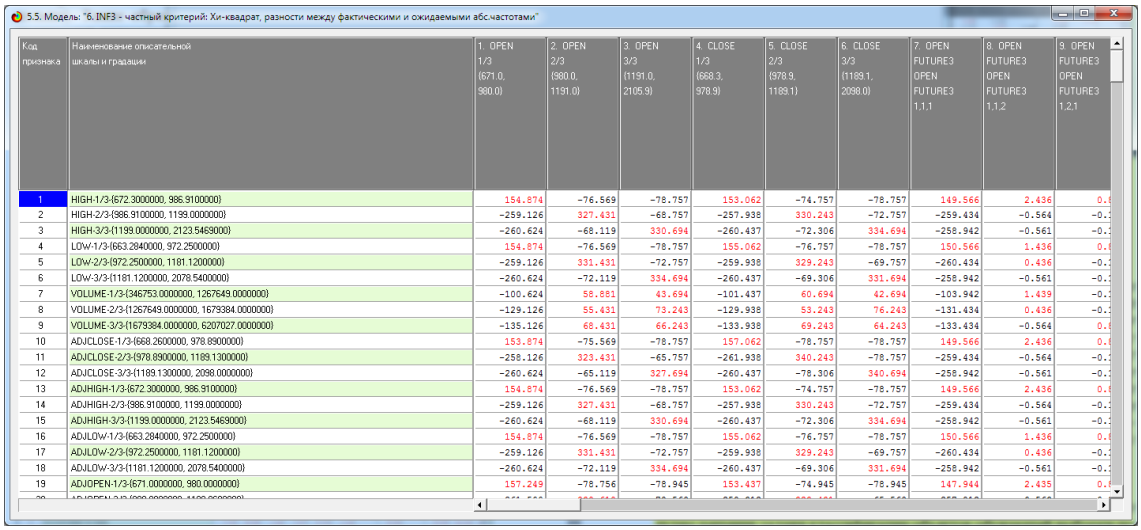
Рисунок 20. Матрица условных и безусловных процентных распределений: статистическая модель PRC2 (фрагмент)



5.5. Модель: "4. INF1 - частный критерий: количество знаний по А.Харкевичу: вероятности из PRC1"

Код признака	Наименование описательной шкалы и градации	1. OPEN 1/3 (671.0, 980.0)	2. OPEN 2/3 (980.0, 1191.0)	3. OPEN 3/3 (1191.0, 2105.9)	4. CLOSE 1/3 (668.3, 978.9)	5. CLOSE 2/3 (978.9, 1189.1)	6. CLOSE 3/3 (1189.1, 2098.0)	7. OPEN FUTURE3 OPEN FUTURE3 1,1,1	8. OPEN FUTURE3 OPEN FUTURE3 1,1,2	9. OPEN FUTURE3 OPEN FUTURE3 1,2,1
1	HIGH-1/3-(672.3000000, 986.9100000)	0.197	-1.554		0.195	-1.261		0.192	0.708	0.
2	HIGH-2/3-(986.9100000, 1199.0000000)	-1.892	0.695	-0.874	-1.770	0.697	-1.090	-2.355		
3	HIGH-3/3-(1199.0000000, 2123.5469000)		-0.870	0.700		-1.087	0.704			
4	LOW-1/3-(663.2840000, 972.2500000)	0.197	-1.554		0.197	-1.555		0.193	0.536	0.
5	LOW-2/3-(972.2500000, 1181.1200000)	-1.892	0.699	-1.090	-2.064	0.696	-0.918		0.243	
6	LOW-3/3-(1181.1200000, 2078.5400000)		-1.086	0.704		-0.916	0.701			
7	VOLUME-1/3-(346753.0000000, 1267649.0000000)	-0.207	0.238	0.188	-0.209	0.243	0.184	-0.217	0.538	
8	VOLUME-2/3-(1267649.0000000, 1679384.0000000)	-0.287	0.226	0.278	-0.290	0.219	0.287	-0.297	0.243	
9	VOLUME-3/3-(1679384.0000000, 6207027.0000000)	-0.307	0.265	0.258	-0.303	0.267	0.252	-0.304		0.
10	ADJCLOSE-1/3-(668.2600000, 978.8900000)	0.196	-1.382		0.199			0.192	0.708	0.
11	ADJCLOSE-2/3-(978.8900000, 1189.1300000)	-1.770	0.691	-0.763		0.708		-2.355		
12	ADJCLOSE-3/3-(1189.1300000, 2098.0000000)		-0.759	0.697			0.710			
13	ADJHIGH-1/3-(672.3000000, 986.9100000)	0.197	-1.554		0.195	-1.261		0.192	0.708	0.
14	ADJHIGH-2/3-(986.9100000, 1199.0000000)	-1.892	0.695	-0.874	-1.770	0.697	-1.090	-2.355		
15	ADJHIGH-3/3-(1199.0000000, 2123.5469000)		-0.870	0.700		-1.087	0.704			
16	ADJLOW-1/3-(663.2840000, 972.2500000)	0.197	-1.554		0.197	-1.555		0.193	0.536	0.
17	ADJLOW-2/3-(972.2500000, 1181.1200000)	-1.892	0.699	-1.090	-2.064	0.696	-0.918		0.243	
18	ADJLOW-3/3-(1181.1200000, 2078.5400000)		-1.086	0.704		-0.916	0.701			
19	ADJOPEN-1/3-(671.0000000, 980.0000000)	0.199			0.195	-1.262		0.190	0.707	0.

Рисунок 21. Матрица информативностей: СК-модель INF1 (фрагмент)



5.5. Модель: "6.INF3 - частный критерий: Хи-квадрат, разности между фактическими и ожидаемыми абс.частотами"

Код признака	Наименование описательной шкалы и градации	1. OPEN 1/3 (671.0, 980.0)	2. OPEN 2/3 (980.0, 1191.0)	3. OPEN 3/3 (1191.0, 2105.9)	4. CLOSE 1/3 (668.3, 978.9)	5. CLOSE 2/3 (978.9, 1189.1)	6. CLOSE 3/3 (1189.1, 2098.0)	7. OPEN FUTURE3 OPEN FUTURE3 1.1.1	8. OPEN FUTURE3 OPEN FUTURE3 1.1.2	9. OPEN FUTURE3 OPEN FUTURE3 1.2.1
1	HIGH-1/3-(672.3000000, 986.9100000)	154.874	-76.569	-78.757	153.062	-74.757	-78.757	149.566	2.436	0.0
2	HIGH-2/3-(986.9100000, 1199.0000000)	-259.126	327.431	-68.757	-257.938	330.243	-72.757	-259.434	-0.564	-0.0
3	HIGH-3/3-(1199.0000000, 2123.5469000)	-260.624	-68.119	330.694	-260.437	-72.306	334.694	-258.942	-0.561	-0.0
4	LOW-1/3-(663.2840000, 972.2500000)	154.874	-76.569	-78.757	155.062	-76.757	-78.757	150.566	1.436	0.0
5	LOW-2/3-(972.2500000, 1181.1200000)	-259.126	331.431	-72.757	-259.938	329.243	-69.757	-260.434	0.436	-0.0
6	LOW-3/3-(1181.1200000, 2078.5400000)	-260.624	-72.119	334.694	-260.437	-69.306	331.694	-258.942	-0.561	-0.0
7	VOLUME-1/3-(346753.0000000, 1267649.0000000)	-100.624	58.881	43.694	-101.437	60.694	42.694	-103.942	1.439	-0.0
8	VOLUME-2/3-(1267649.0000000, 1673384.0000000)	-129.126	55.431	73.243	-129.938	53.243	76.243	-131.434	0.436	-0.0
9	VOLUME-3/3-(1673384.0000000, 6207027.0000000)	-135.126	68.431	66.243	-133.938	69.243	64.243	-133.434	-0.564	0.0
10	ADJCLOSE-1/3-(668.2600000, 978.8900000)	153.874	-75.569	-78.757	157.062	-78.757	-78.757	149.566	2.436	0.0
11	ADJCLOSE-2/3-(978.8900000, 1189.1300000)	-258.126	323.431	-65.757	-261.938	340.243	-78.757	-259.434	-0.564	-0.0
12	ADJCLOSE-3/3-(1189.1300000, 2098.0000000)	-260.624	-65.119	327.694	-260.437	-78.306	340.694	-258.942	-0.561	-0.0
13	ADJHIGH-1/3-(672.3000000, 986.9100000)	154.874	-76.569	-78.757	153.062	-74.757	-78.757	149.566	2.436	0.0
14	ADJHIGH-2/3-(986.9100000, 1199.0000000)	-259.126	327.431	-68.757	-257.938	330.243	-72.757	-259.434	-0.564	-0.0
15	ADJHIGH-3/3-(1199.0000000, 2123.5469000)	-260.624	-68.119	330.694	-260.437	-72.306	334.694	-258.942	-0.561	-0.0
16	ADJLOW-1/3-(663.2840000, 972.2500000)	154.874	-76.569	-78.757	155.062	-76.757	-78.757	150.566	1.436	0.0
17	ADJLOW-2/3-(972.2500000, 1181.1200000)	-259.126	331.431	-72.757	-259.938	329.243	-69.757	-260.434	0.436	-0.0
18	ADJLOW-3/3-(1181.1200000, 2078.5400000)	-260.624	-72.119	334.694	-260.437	-69.306	331.694	-258.942	-0.561	-0.0
19	ADJOPEN-1/3-(671.0000000, 980.0000000)	157.249	-78.756	-78.945	153.437	-74.945	-78.945	147.944	2.435	0.0

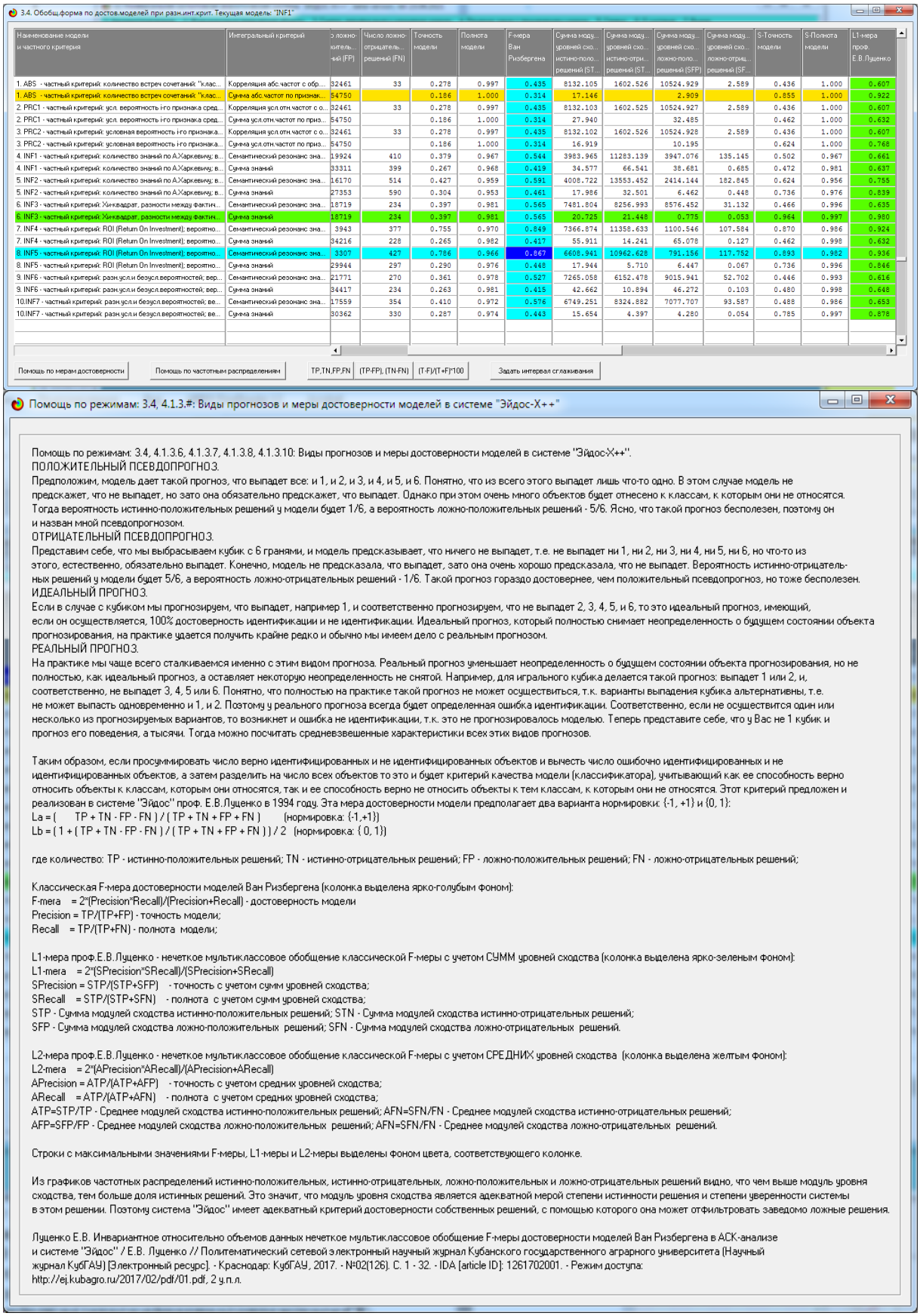
Рисунок 22. Матрица хи-квадрат: СК-модель INF3 (фрагмент)

Отметим, что в АСК-анализе и СК-моделях степень выраженности различных свойств объектов наблюдения рассматривается с единственной точки зрения: с точки зрения того, какое ***количество информации*** содержится в них о том, к каким обобщающим категориям (классам) будут принадлежать или не принадлежать эти объекты.

Поэтому не играет никакой роли, в каких единицах измерения измеряются те или иные свойства объектов наблюдения, а также в каких единицах измеряются результаты влияния этих свойств, натуральных, в процентах или стоимостных [4]. Это и есть решение проблемы сопоставимости в АСК-анализе и системе «Эйдос», отличающее их от других интеллектуальных технологий.

13.3.4.2. Оценка достоверности моделей

Оценка достоверности моделей в системе «Эйдос» осуществляется путем решения задачи классификации объектов обучающей выборки по обобщенным образам классов и подсчета количества истинных и ложных положительных и отрицательных решений по F-мере Ван Ризбергена, а также по критериям L1- L2-мерам проф. Е.В.Луценко, которые предложены для того, чтобы смягчить или полностью преодолеть некоторые недостатки F-меры [11]. В режиме 3.4 системы «Эйдос» изучается достоверность каждой частной модели в соответствии с этими мерами достоверности (рисунок 16). Из рисунка 16 мы видим, что в данном интеллектуальном приложении по F-критерию Ван Ризбергена наиболее достоверной является СК-модель INF5 с интегральным критерием «Семантический резонанс знаний» (F=0,867 при максимуме 1,000), что является неплохим результатом для моделируемой предметной области. ***Это подтверждает наличие и адекватное отражение в СК-модели INF5 сильных причинно-следственных зависимостей между динамикой различных характеристик финансового рынка и курсами акций компании Гугл.***



Помощь по режимам: 3.4, 4.1.3.6, 4.1.3.7, 4.1.3.8, 4.1.3.10: Виды прогнозов и меры достоверности моделей в системе "Эйдос-X++".

ПОЛОЖИТЕЛЬНЫЙ ПСЕВДОПРОГНОЗ.

Предположим, модель дает такой прогноз, что выпадет все: и 1, и 2, и 3, и 4, и 5, и 6. Понятно, что из всего этого выпадет лишь что-то одно. В этом случае модель не предскажет, что не выпадет, но зато она обязательно предскажет, что выпадет. Однако при этом очень много объектов будет отнесено к классам, к которым они не относятся. Тогда вероятность истинно-положительных решений у модели будет 1/6, а вероятность ложно-положительных решений - 5/6. Ясно, что такой прогноз бесполезен, поэтому он и назван мной псевдопрогнозом.

ОТРИЦАТЕЛЬНЫЙ ПСЕВДОПРОГНОЗ.

Представим себе, что мы выбрасываем кубик с 6 гранями, и модель предсказывает, что ничего не выпадет, т.е. не выпадет ни 1, ни 2, ни 3, ни 4, ни 5, ни 6, но что-то из этого, естественно, обязательно выпадет. Конечно, модель не предсказала, что выпадет, зато она очень хорошо предсказала, что не выпадет. Вероятность истинно-отрицательных решений у модели будет 5/6, а вероятность ложно-отрицательных решений - 1/6. Такой прогноз гораздо достовернее, чем положительный псевдопрогноз, но тоже бесполезен.

ИДЕАЛЬНЫЙ ПРОГНОЗ.

Если в случае с кубиком мы прогнозируем, что выпадет, например 1, и соответственно прогнозируем, что не выпадет 2, 3, 4, 5, и 6, то это идеальный прогноз, имеющий, если он осуществляется, 100% достоверность идентификации и не идентификации. Идеальный прогноз, который полностью снимает неопределенность о будущем состоянии объекта прогнозирования, на практике удается получить крайне редко и обычно мы имеем дело с реальным прогнозом.

РЕАЛЬНЫЙ ПРОГНОЗ.

На практике мы чаще всего сталкиваемся именно с этим видом прогноза. Реальный прогноз уменьшает неопределенность о будущем состоянии объекта прогнозирования, но не полностью, как идеальный прогноз, а оставляет некоторую неопределенность не снятой. Например, для игрального кубика делается такой прогноз: выпадет 1 или 2, и, соответственно, не выпадет 3, 4, 5 или 6. Понятно, что полностью на практике такой прогноз не может осуществиться, т.к. варианты выпадения кубика альтернативны, т.е. не может выпасть одновременно и 1, и 2. Поэтому у реального прогноза всегда будет определенная ошибка идентификации. Соответственно, если не осуществится один или несколько из прогнозируемых вариантов, то возникнет и ошибка не идентификации, т.к. это не прогнозировалось моделью. Теперь представите себе, что у Вас не 1 кубик и прогноз его поведения, а тысячи. Тогда можно посчитать средневзвешенные характеристики всех этих видов прогнозов.

Таким образом, если просуммировать число верно идентифицированных и не идентифицированных объектов и вычесть число ошибочно идентифицированных и не идентифицированных объектов, а затем разделить на число всех объектов то это и будет критерий качества модели (классификатора), учитывающий как ее способность верно относить объекты к классам, которым они относятся, так и ее способность верно не относить объекты к тем классам, к которым они не относятся. Этот критерий предложен и реализован в системе "Эйдос" проф. Е.В.Луценко в 1994 году. Эта мера достоверности модели предполагает два варианта нормировки: {-1, +1} и {0, 1}:

La = (TP + TN - FP - FN) / (TP + TN + FP + FN) (нормировка: {-1,+1})

Lb = (1 + (TP + TN - FP - FN) / (TP + TN + FP + FN)) / 2 (нормировка: { 0, 1})

где количество: TP - истинно-положительных решений; TN - истинно-отрицательных решений; FP - ложно-положительных решений; FN - ложно-отрицательных решений;

Классическая F-мера достоверности моделей Ван Ризбергена (колонка выделена ярко-голубым фоном):

F-mera = 2*(Precision*Recall)/(Precision+Recall) - достоверность модели

Precision = TP/(TP+FP) - точность модели;

Recall = TP/(TP+FN) - полнота модели;

L1-мера проф.Е.В.Луценко - нечеткое мультиклассовое обобщение классической F-меры с учетом СУММ уровней сходства (колонка выделена ярко-зеленым фоном):

L1-mera = 2*(SPrecision*SRecall)/(SPrecision+SRecall)

SPrecision = STP/(STP+SFP) - точность с учетом сумм уровней сходства;

SRecall = STP/(STP+SFN) - полнота с учетом сумм уровней сходства;

STP - Сумма модулей сходства истинно-положительных решений; STN - Сумма модулей сходства истинно-отрицательных решений;

SFP - Сумма модулей сходства ложно-положительных решений; SFN - Сумма модулей сходства ложно-отрицательных решений.

L2-мера проф.Е.В.Луценко - нечеткое мультиклассовое обобщение классической F-меры с учетом СРЕДНИХ уровней сходства (колонка выделена желтым фоном):

L2-mera = 2*(APrecision*ARecall)/(APrecision+ARecall)

APrecision = ATP/(ATP+AFP) - точность с учетом средних уровней сходства;

ARecall = ATP/(ATP+AFN) - полнота с учетом средних уровней сходства;

ATP=STP/TP - Среднее модулей сходства истинно-положительных решений; AFN=SFN/FN - Среднее модулей сходства истинно-отрицательных решений;

AFP=SFP/FP - Среднее модулей сходства ложно-положительных решений; AFN=SFN/FN - Среднее модулей сходства ложно-отрицательных решений.

Строки с максимальными значениями F-меры, L1-меры и L2-меры выделены фоном цвета, соответствующего колонке.

Из графиков частотных распределений истинно-положительных, истинно-отрицательных, ложно-положительных и ложно-отрицательных решений видно, что чем выше модуль уровня сходства, тем больше доля истинных решений. Это значит, что модуль уровня сходства является адекватной мерой степени истинности решения и степени уверенности системы в этом решении. Поэтому система "Эйдос" имеет адекватный критерий достоверности собственных решений, с помощью которого она может отфильтровать заведомо ложные решения.

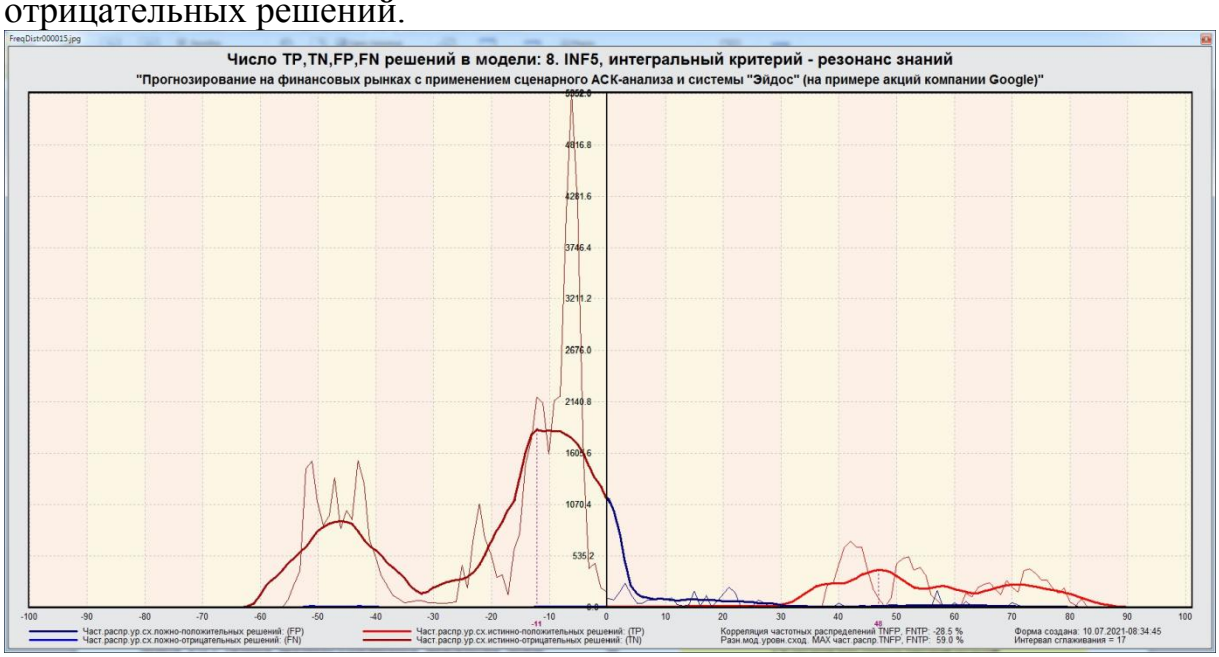
Луценко Е.В. Инвариантное относительно объемов данных нечеткое мультиклассовое обобщение F-меры достоверности моделей Ван Ризбергена в АСК-анализе и системе "Эйдос" / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. - Краснодар: КубГАУ, 2017. - №02(126). С. 1 - 32. - IDA [article ID]: 1261702001. - Режим доступа: http://ej.kubagro.ru/2017/02/pdf/01.pdf, 2 у.п.л.

Рисунок 23. Экранная форма с информацией о достоверности моделей по F-критерию Ван Ризбергена и help данного режима

На рисунке 17 приведены:

– частотное распределения числа истинных и ложных положительных и отрицательных решений по прогнозированию курсов открытия и закрытия акций Гугл и их динамики при разных уровнях сходства и различия в СК-модели INF5 по данным обучающей выборки;

– разность количества истинных и ложных положительных и отрицательных решений.



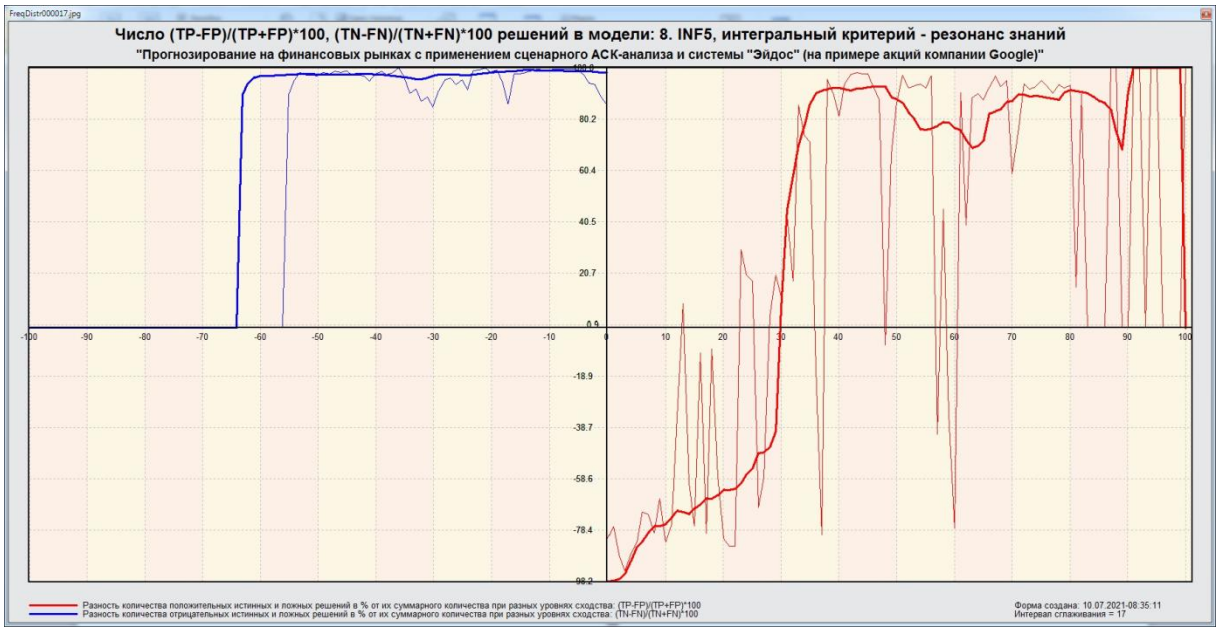


Рисунок 24. Частотные распределения числа истинных и ложных положительных и отрицательных решений и их разности в СК-модели Inf3

Рисунок 17 содержит изображения частотных распределений количества истинных и ложных положительных и отрицательных решений

и их разности в зависимости от уровня сходства. Из этого рисунка мы видим, что:

для положительных решений (т.е. когда уровень сходства объекта с классом положительный):

– при уровнях сходства объекта с классом от 0% до 30% количество ложных решений превосходит количество истинных решений;

– при уровнях сходства объекта с классом выше 35% количество истинных решений значительно превосходит число ложных решений.

для отрицательных решений (т.е. когда уровень сходства объекта с классом отрицательный) количество истинных решений всегда, т.е. всех уровнях сходства объекта с классом, значительно превосходит число ложных решений.

Поэтому выберем СК-модель INF5 в качестве текущей для решения большинства задач.

Отметим также, что из второго рисунка 17 видно, что при увеличении уровня сходства объекта с классом закономерно растет и доля истинных решений среди всех решений, а доля ложных решений уменьшается. Из этого можно обоснованно сделать очень важный вывод: ***уровень сходства объекта с классом, т.е. значение интегрального критерия, является адекватной мерой степени истинности решения.*** Этот вывод подтверждается на огромном количестве решенных в системе «Эйдос» задач из самых различных предметных областей.

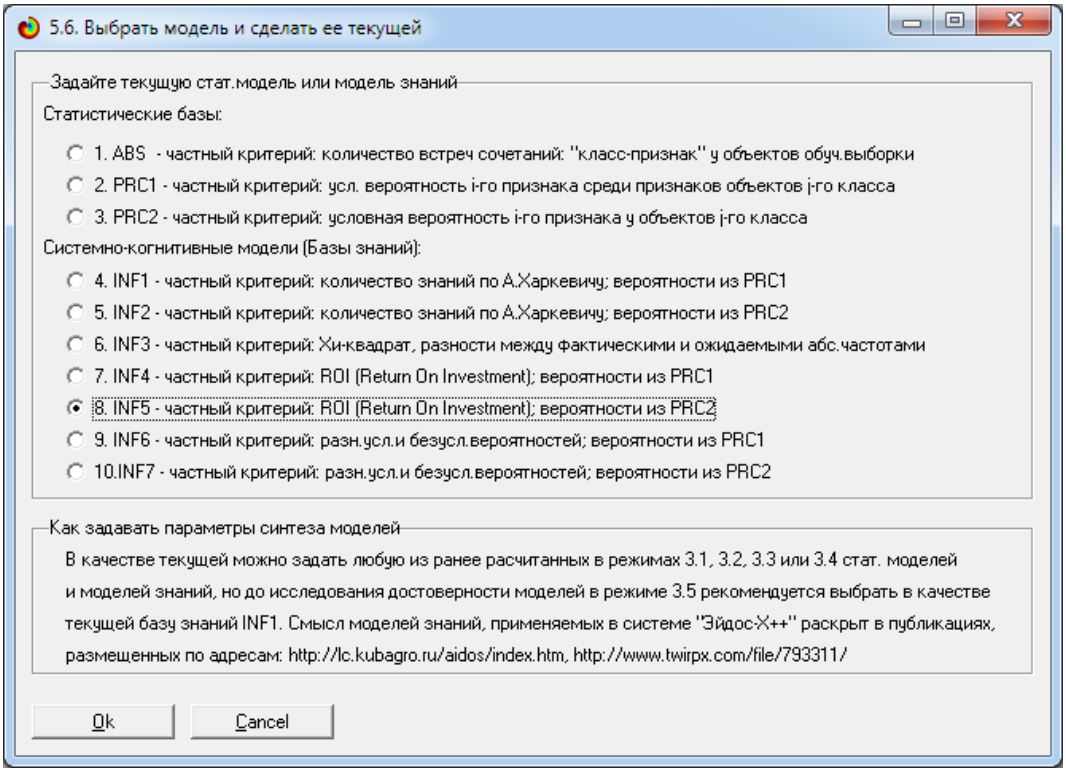
Это означает, что ***в системе «Эйдос» есть достоверный внутренний критерий степени истинности решений задач, предлагаемых системой на основе созданных в ней моделей. Таким образом, система «Эйдос» не просто идентифицирует, но и оценивает достоверность идентификации, не просто прогнозирует, но и оценивает достоверность прогнозирования, не просто предлагает решение, но и оценивает эффективность этого решения, и т.д.***

Таким образом, система Эйдос не только прогнозирует значения будущих параметров, но и адекватно оценивает достоверность их прогнозирования. Наличие в системе «Эйдос» внутреннего достоверного критерия достоверности прогнозирования позволяет прогнозировать наступление точки ***бифуркации***, точки неопределенности. ***В точках бифуркации резко уменьшается достоверность прогнозирования и возрастает разброс точечных прогнозов с различных позиций во времени. Фактически это означает, что можно либо достоверно прогнозировать, что произойдет, либо достоверно прогнозировать, что мы не можем достоверно прогнозировать, т.е. достоверно прогнозировать точку бифуркации***. Об этом есть в работе [21]: http://lc.kubagro.ru/aidos/aidos02/7.4.htm. В этой монографии 2002 года описаны результаты, полученные в 1994 году.

13.3.4.3. Задание текущей модели

В системе «Эйдос» большинство задач решается сразу для всех моделей. Однако задача идентификации (распознавания, классификации, диагностики) и задача прогнозирования решаются только в модели, заданной в качестве текущей. Это сделано потому, что эти задачи являются наиболее трудоемкими в вычислительном отношении и их решение может занимать довольно продолжительное время. Эта вычислительная сложность вязана с тем, что при решении этих задач каждый объект обучающей выборки сравнивается с каждым из классов по всем признакам. Даже при использовании графического процессора для расчетов, а это возможно в системе «Эйдос», время распознавания может быть довольно заметным при очень большом количестве объектов обучающей выборки, очень большом количестве классов и очень большом количестве признаков. А после самого решения задачи по результатам ее решения рассчитывается еще 10 выходных форм, и это делается (в текущей версии системы «Эйдос») на центральном процессоре и занимает также заметное время, которое составляет 99% времени решения этих задач. Но не рассчитывать этих выходных форм нельзя, т.к. именно в их расчете состоит смысл решения этих задач.

Поэтому зададим наиболее достоверную модель INF5 в качестве текущей. Для этого выполним режим 5.6 (рисунки 1 и 20).



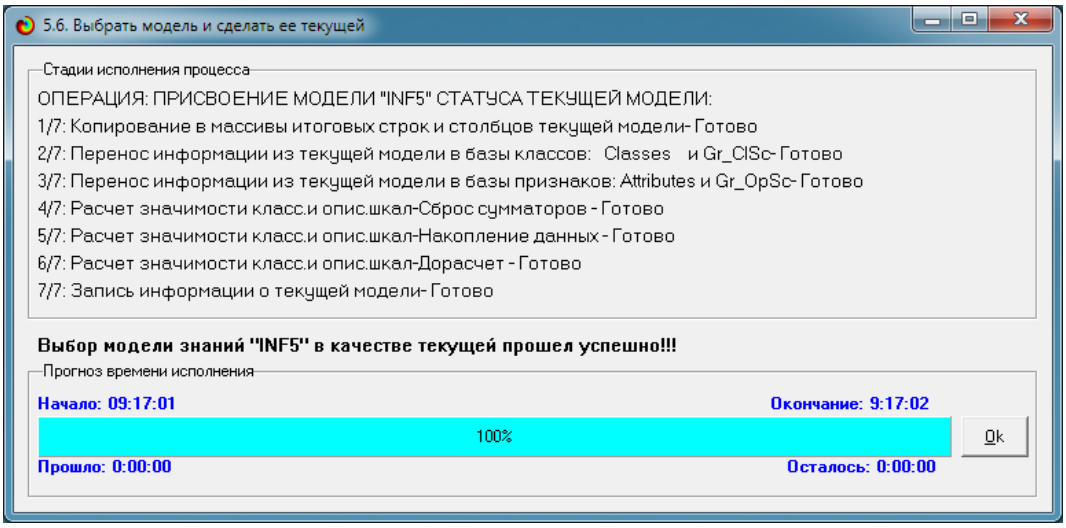


Рисунок 25. Экранные формы присвоения наиболее достоверной СК-модели INF5 статуса текущей модели

Из второй экранной формы на рисунке 18 видно, что весь процесс присвоения наиболее достоверной СК-модели INF5 статуса текущей модели занял менее половины секунды.

13.3.5. Задача 4: решение различных задач в наиболее достоверной модели

13.3.5.1. Подзадача 4.1. Прогнозирование (диагностика, классификация, распознавание, идентификация)

Решим задачу системной идентификации 1257 объектов наблюдения с 54 классами. Эту задачу решим в наиболее достоверной СК-модели INF5 на графическом процессоре (GPU) (рисунок 19).

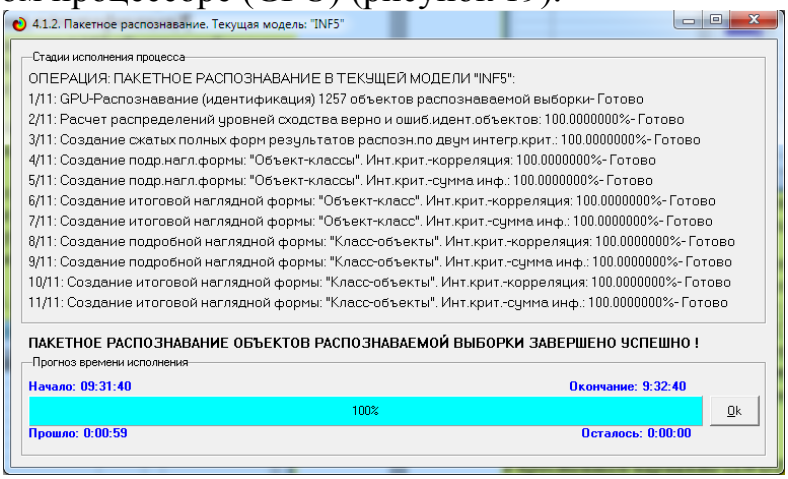


Рисунок 26. Экранные формы, которые отображают процесс решения задачи системной идентификации в текущей модели

Из рисунка 19 видно, что процесс идентификации занял 59 секунд.

Для самого прогнозирования использовался графический процессор (GPU), а точнее 1500 шейдерных процессоров видеокарты NVIDIA

GeForce GTX 770. Для расчета 10 выходных форм по результатам прогнозирования использовался центральный процессор (CPU) i7. В основном время было затрачено именно на расчет этих выходных форм. Эти формы отражают результаты прогнозирования в различных разрезах и обобщениях:

В связи с ограниченностью объема данной работы приведем лишь одну из этих 10 выходных форм: 4.1.3.1 (рисунок 20).

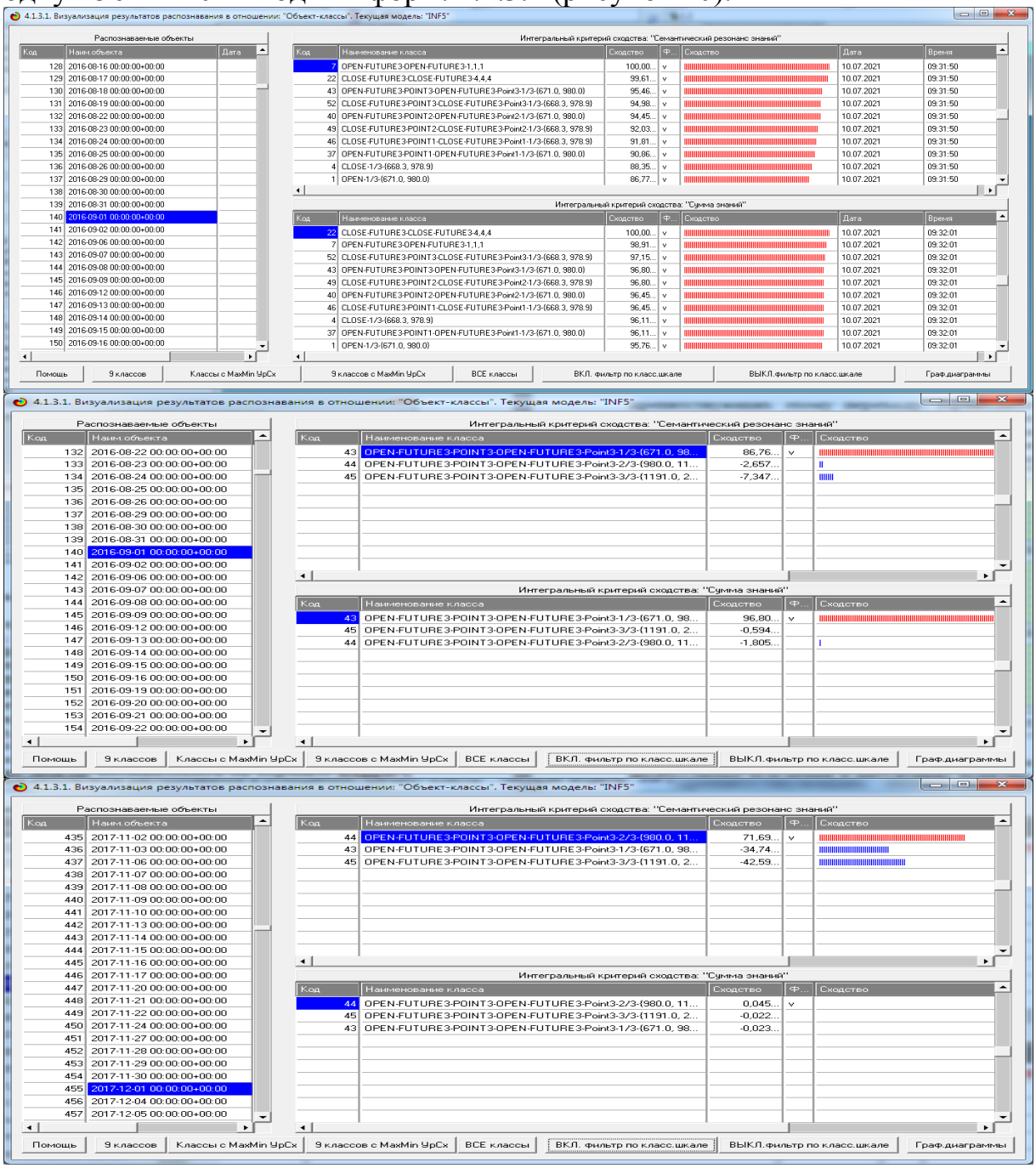


Рисунок 27. Выходные формы по результатам прогнозирования

Символ «√» стоит против тех результатов идентификации, которые подтвердились на опыте, т.е. соответствуют факту.

Из рисунка 22 видно, что результаты идентификации являются отличными, естественно при учете информации из рисунка 17 о том, что ***достоверные прогнозы в данной модели имеют уровень сходства выше 35% по интегральному критерию «Резонанс знаний» (верхнее правое окно в экранных формах на рисунке 20), т.е., по сути, результаты с более низким уровнем сходства надо просто игнорировать.***

На рисунке 22 во всех скришотах, кроме первого, включен фильтр по одной из классификационных шкал с кодами: 43, 44, 45, 52, 53, 54, отражающей значение в 3-й точке сценария: «Значение в третьей точке сценария» (см. таблицу 5).

Это и есть решение задачи, поставленной на портале Kaggle, только не для 30-й точки сценариев, а для 3-й.

Для получения средневзвешенных сценариев кликаем по самой правой кнопке экранной формы, приведенной на рисунке 22: «Графические диаграммы» и появившейся экранной форме задаем птичками какие формы получить и записать (рисунок 22а):

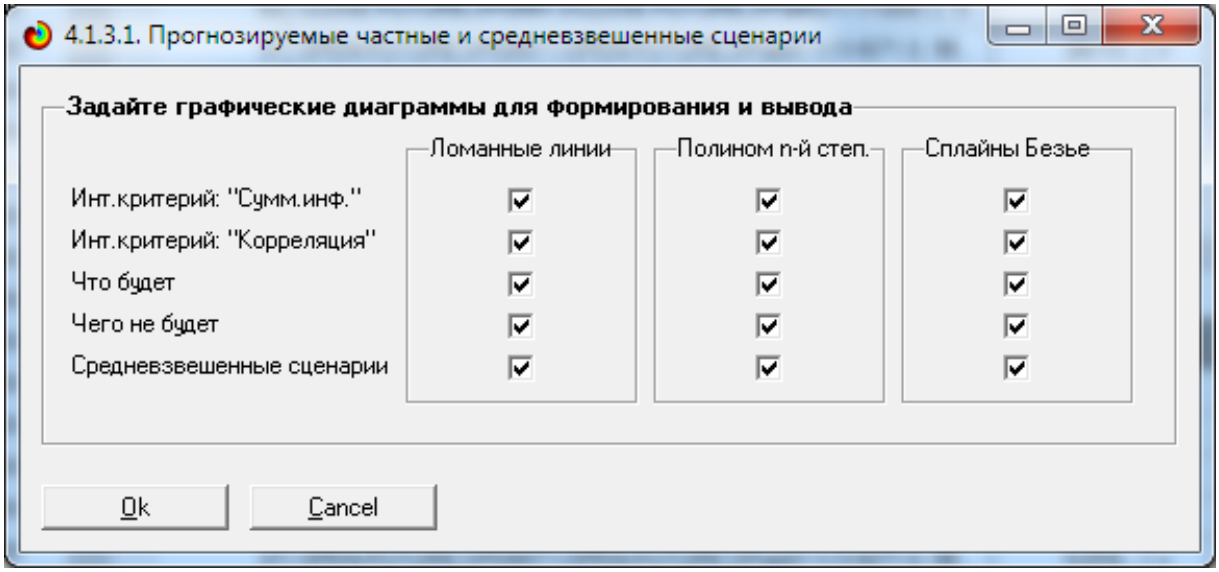
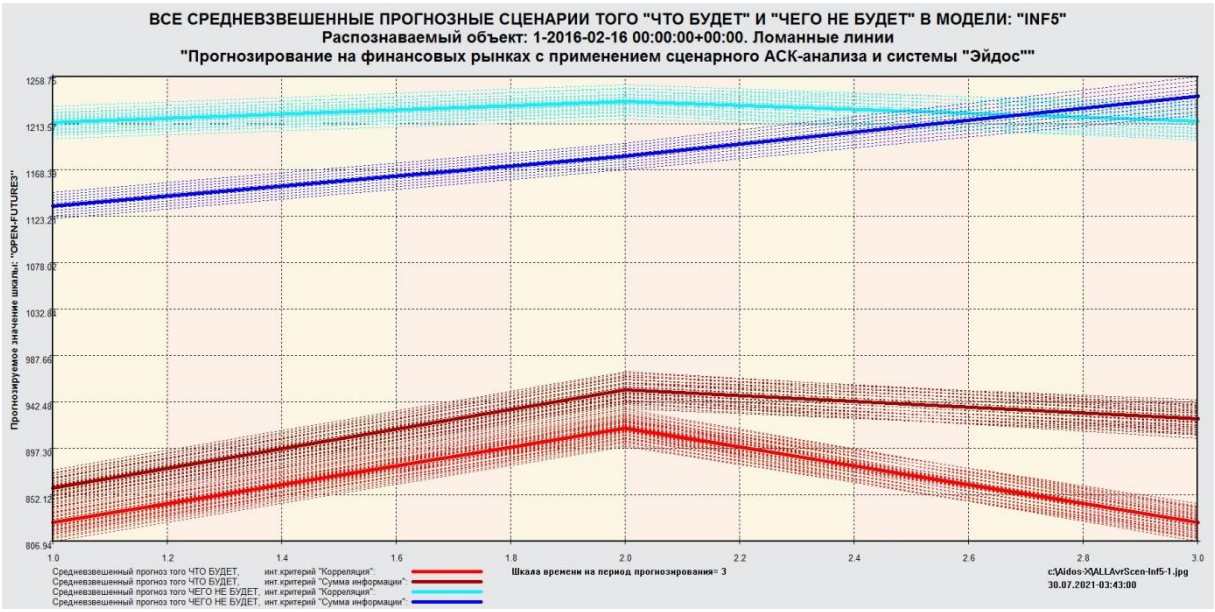
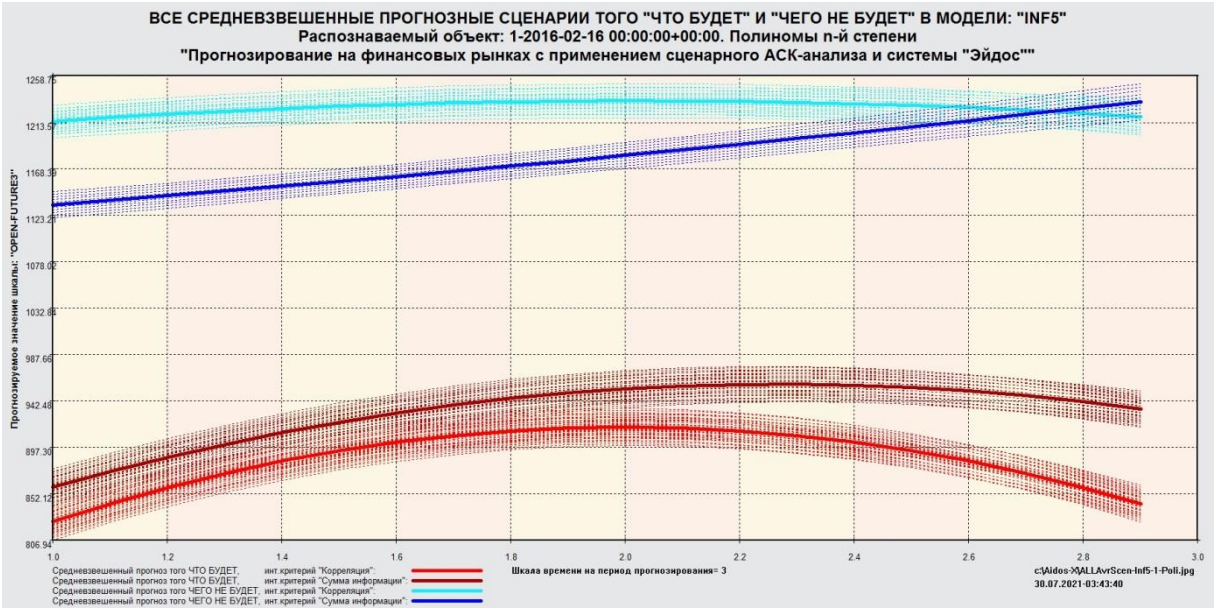
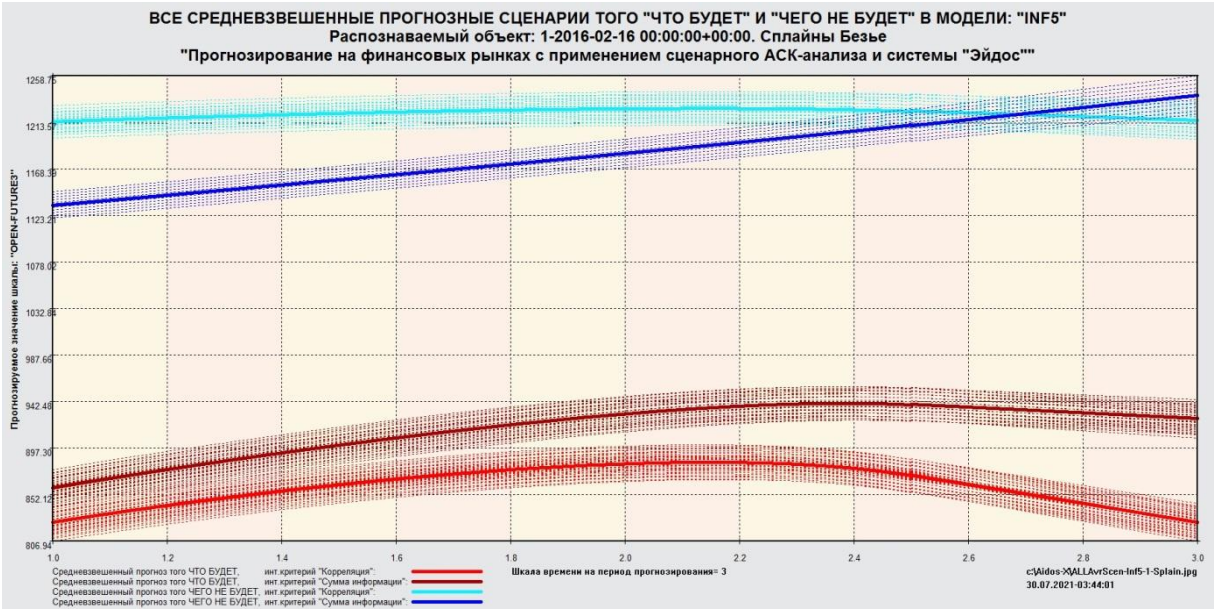


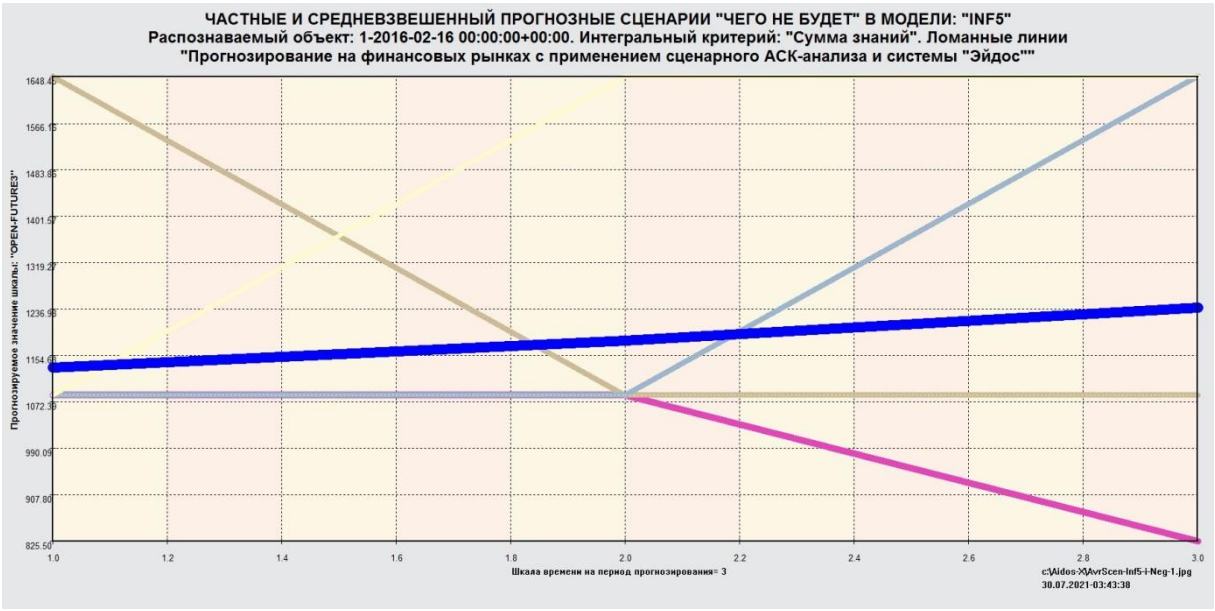
Рисунок 28. Задание графических диаграмм по результатам распознавания для формирования и вывода

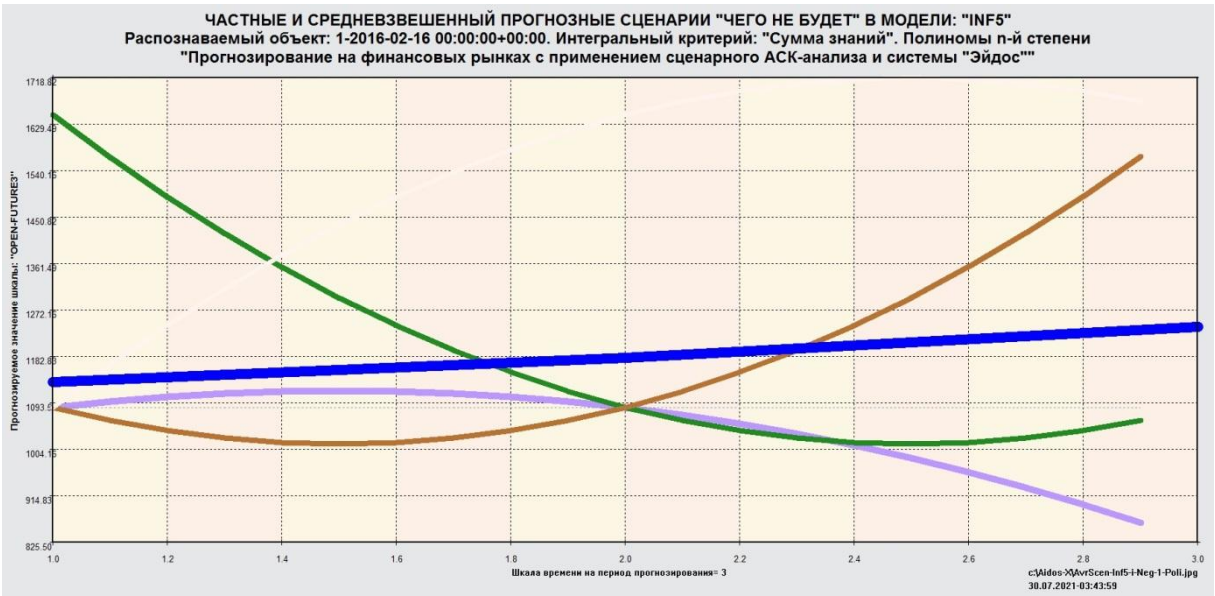
В результате были сформированы и записаны в виде файлов следующие диаграммы (рисунок 22б):

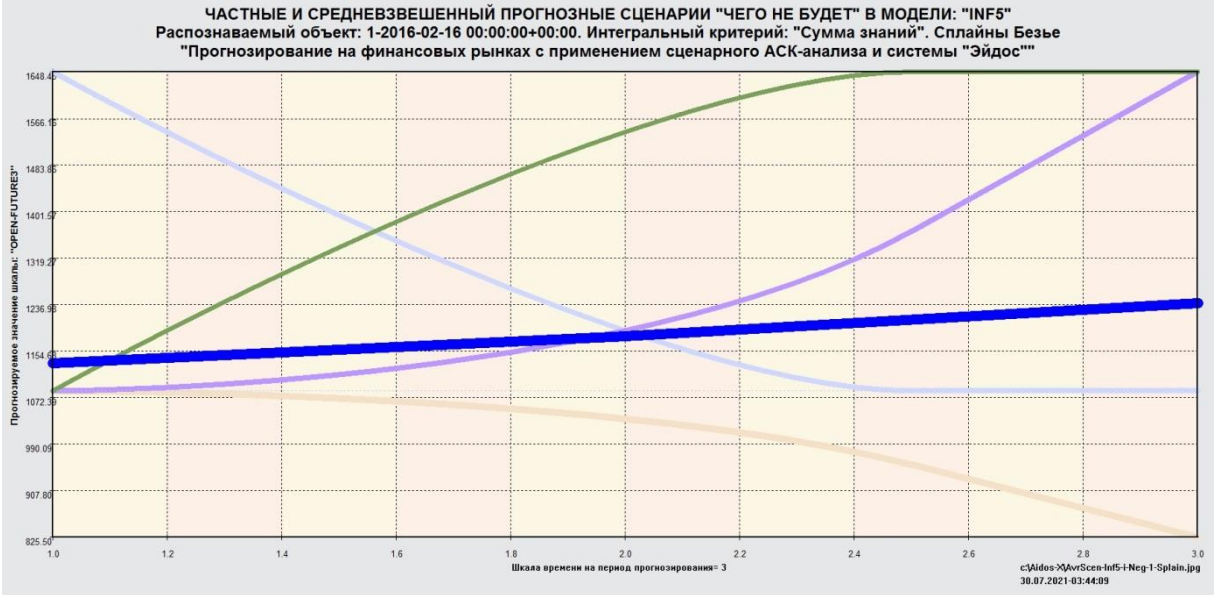

ВСЕ СРЕДНЕВЗВЕШЕННЫЕ ПРОГНОЗНЫЕ СЦЕНАРИИ ТОГО "ЧТО БУДЕТ" И "ЧЕГО НЕ БУДЕТ" В МОДЕЛИ: "INF5"
Распознаваемый объект: 1-2016-02-16 00:00:00+00:00. Ломанные линии
"Прогнозирование на финансовых рынках с применением сценарного АСК-анализа и системы "Эйдос""
Прогнозируемое значение шкалы: "OPEN-FUTURES"
1258.75
1213.57
1168.39
1123.21
1078.02
1032.84
987.66
942.48
897.30
852.12
806.94
1.0
1.2
1.4
1.6
1.8
2.0
2.2
2.4
2.6
2.8
3.0
Средневзвешенный прогноз того ЧТО БУДЕТ, инт.критерий "Корреляция":
Средневзвешенный прогноз того ЧТО БУДЕТ, инт.критерий "Сумма информации":
Средневзвешенный прогноз того ЧЕГО НЕ БУДЕТ, инт.критерий "Корреляция":
Средневзвешенный прогноз того ЧЕГО НЕ БУДЕТ, инт.критерий "Сумма информации":
Шкала времени на период прогнозирования= 3
c:\Aidos-X\ALLAvrScen-Inf5-1.jpg
30.07.2021-03:43:00

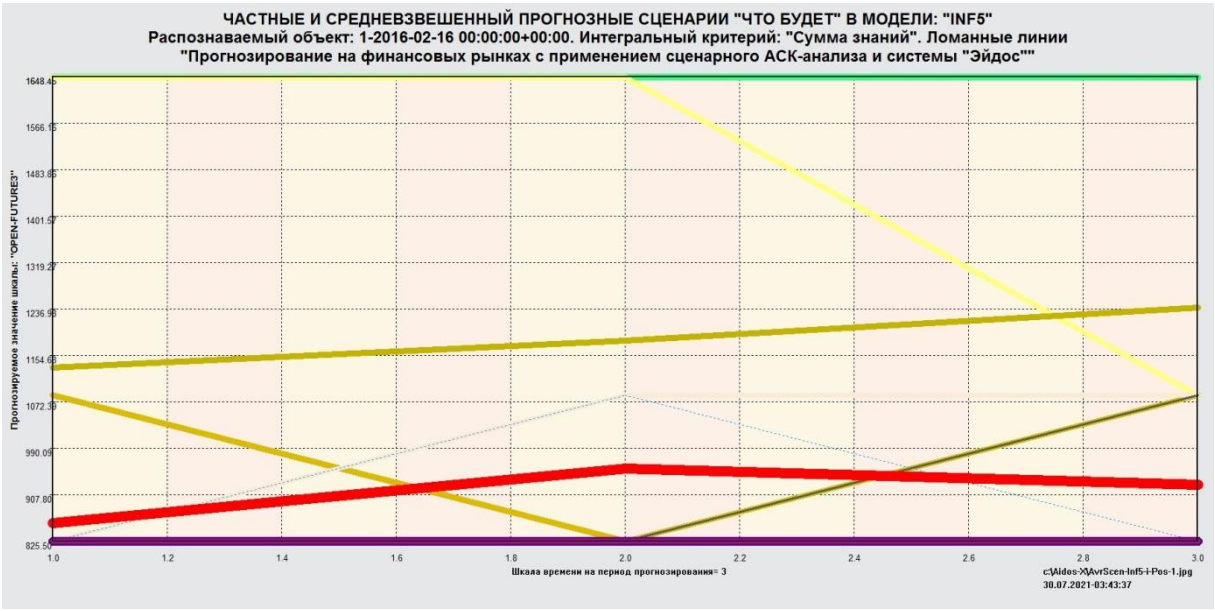

ВСЕ СРЕДНЕВЗВЕШЕННЫЕ ПРОГНОЗНЫЕ СЦЕНАРИИ ТОГО "ЧТО БУДЕТ" И "ЧЕГО НЕ БУДЕТ" В МОДЕЛИ: "INF5"
Распознаваемый объект: 1-2016-02-16 00:00:00+00:00. Полиномы n-й степени
"Прогнозирование на финансовых рынках с применением сценарного АСК-анализа и системы "Эйдос""
Прогнозируемое значение шкалы: "OPEN-FUTURES"
1258.75
1213.57
1168.39
1123.21
1078.02
1032.84
987.66
942.48
897.30
852.12
806.94
1.0
1.2
1.4
1.6
1.8
2.0
2.2
2.4
2.6
2.8
3.0
Средневзвешенный прогноз того ЧТО БУДЕТ, инт.критерий "Корреляция":
Средневзвешенный прогноз того ЧТО БУДЕТ, инт.критерий "Сумма информации":
Средневзвешенный прогноз того ЧЕГО НЕ БУДЕТ, инт.критерий "Корреляция":
Средневзвешенный прогноз того ЧЕГО НЕ БУДЕТ, инт.критерий "Сумма информации":
Шкала времени на период прогнозирования= 3
c:\Aidos-X\ALLAvrScen-Inf5-1-Poli.jpg
30.07.2021-03:43:40

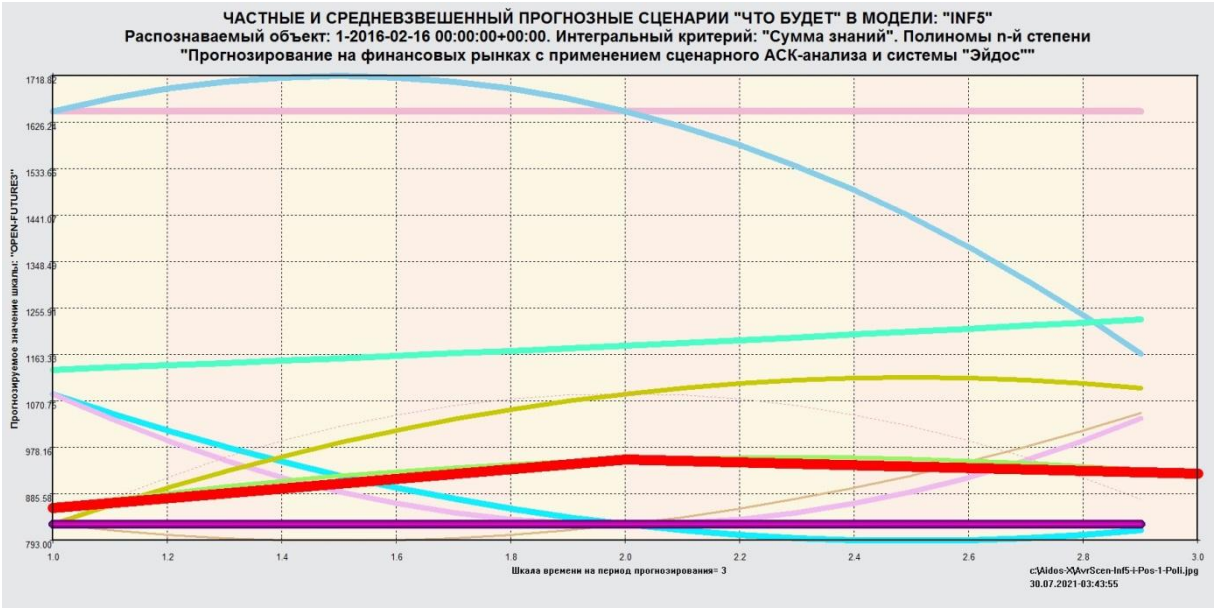

ВСЕ СРЕДНЕВЗВЕШЕННЫЕ ПРОГНОЗНЫЕ СЦЕНАРИИ ТОГО "ЧТО БУДЕТ" И "ЧЕГО НЕ БУДЕТ" В МОДЕЛИ: "INF5"
Распознаваемый объект: 1-2016-02-16 00:00:00+00:00. Сплайны Безье
"Прогнозирование на финансовых рынках с применением сценарного АСК-анализа и системы "Эйдос""
Прогнозируемое значение шкалы: "OPEN-FUTURES"
1258.75
1213.57
1168.39
1123.21
1078.02
1032.84
987.66
942.48
897.30
852.12
806.94
1.0
1.2
1.4
1.6
1.8
2.0
2.2
2.4
2.6
2.8
3.0
Средневзвешенный прогноз того ЧТО БУДЕТ, инт.критерий "Корреляция":
Средневзвешенный прогноз того ЧТО БУДЕТ, инт.критерий "Сумма информации":
Средневзвешенный прогноз того ЧЕГО НЕ БУДЕТ, инт.критерий "Корреляция":
Средневзвешенный прогноз того ЧЕГО НЕ БУДЕТ, инт.критерий "Сумма информации":
Шкала времени на период прогнозирования= 3
c:\Aidos-X\ALLAvrScen-Inf5-1-Splain.jpg
30.07.2021-03:44:01

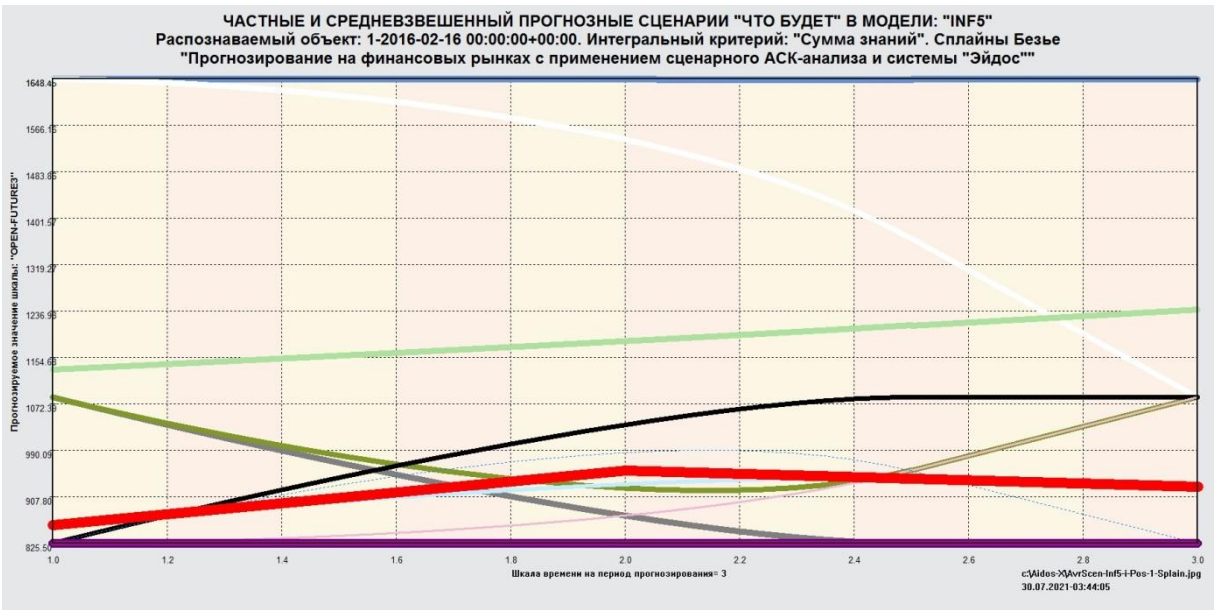

ЧАСТНЫЕ И СРЕДНЕВЗВЕШЕННЫЙ ПРОГНОЗНЫЕ СЦЕНАРИИ "ЧЕГО НЕ БУДЕТ" В МОДЕЛИ: "INF5"
Распознаваемый объект: 1-2016-02-16 00:00:00+00:00. Интегральный критерий: "Сумма знаний". Ломанные линии
"Прогнозирование на финансовых рынках с применением сценарного АСК-анализа и системы "Эйдос""
Прогнозируемое значение шкалы: "OPEN-FUTURES"
1648.45
1566.15
1483.85
1401.57
1319.27
1236.98
1154.68
1072.39
990.09
907.80
825.50
1.0
1.2
1.4
1.6
1.8
2.0
2.2
2.4
2.6
2.8
3.0
Шкала времени на период прогнозирования= 3
c:\Aidos-X\AvrScen-Inf5-i-Neg-1.jpg
30.07.2021-03:43:38

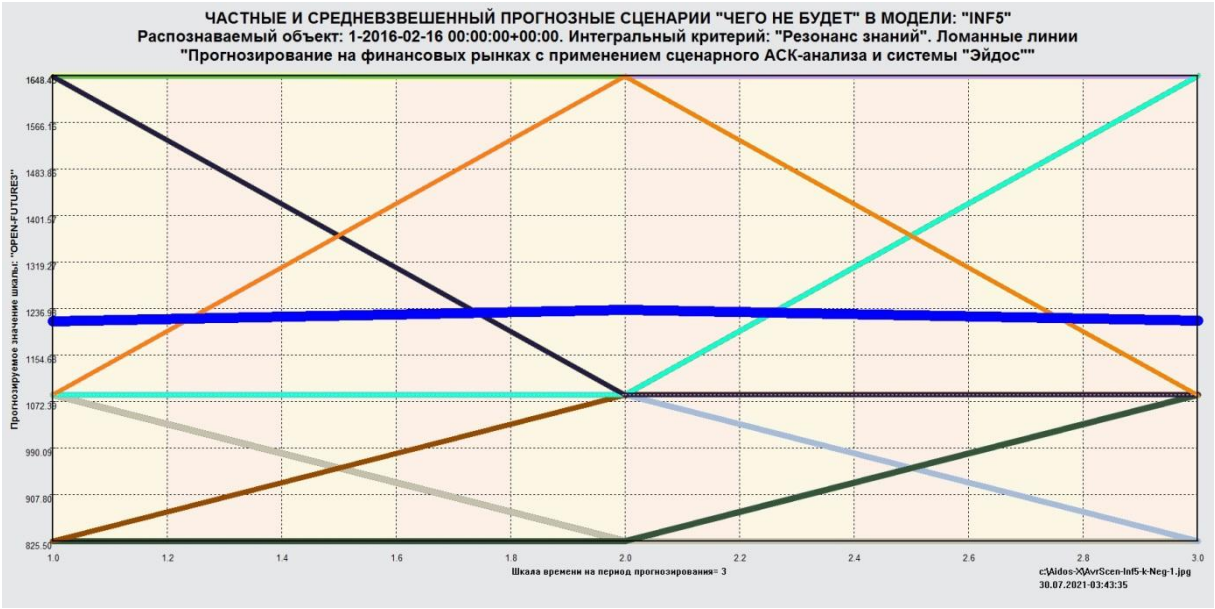

ЧАСТНЫЕ И СРЕДНЕВЗВЕШЕННЫЙ ПРОГНОЗНЫЕ СЦЕНАРИИ "ЧЕГО НЕ БУДЕТ" В МОДЕЛИ: "INF5"
Распознаваемый объект: 1-2016-02-16 00:00:00+00:00. Интегральный критерий: "Сумма знаний". Полиномы n-й степени
"Прогнозирование на финансовых рынках с применением сценарного АСК-анализа и системы "Эйдос""
Прогнозируемое значение шкалы: "OPEN-FUTURES"
1718.82
1629.48
1540.15
1450.82
1361.49
1272.15
1182.82
1093.49
1004.15
914.83
825.50
1.0
1.2
1.4
1.6
1.8
2.0
2.2
2.4
2.6
2.8
3.0
Шкала времени на период прогнозирования= 3
c:\Aidos-X\AvrScen-Inf5-i-Neg-1-Poli.jpg
30.07.2021-03:43:59

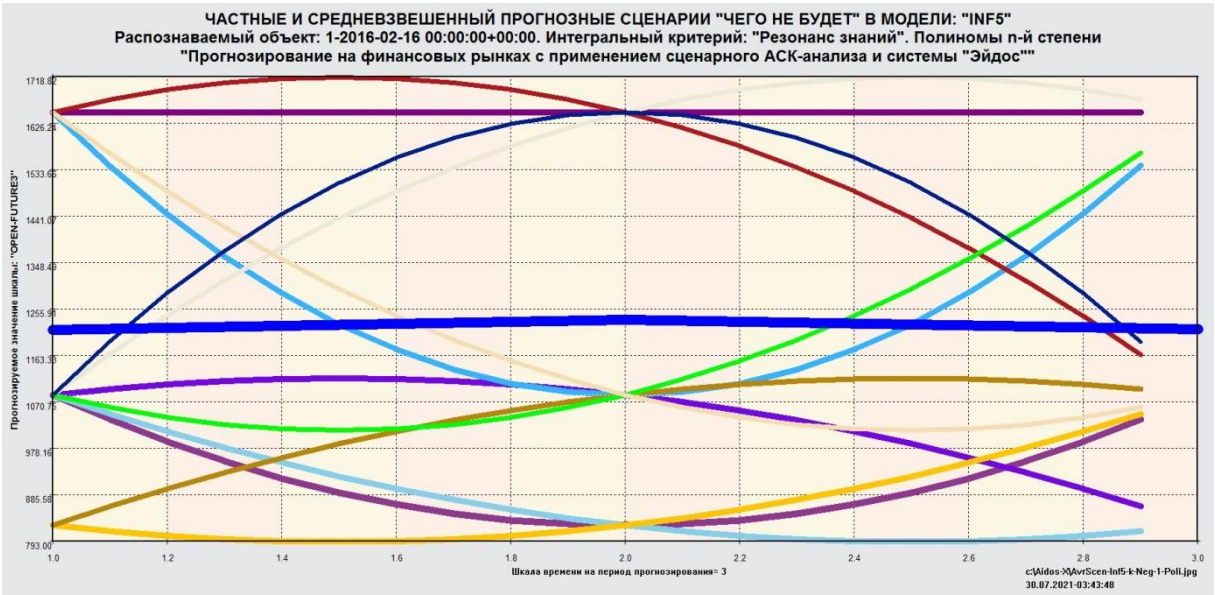

ЧАСТНЫЕ И СРЕДНЕВЗВЕШЕННЫЙ ПРОГНОЗНЫЕ СЦЕНАРИИ "ЧЕГО НЕ БУДЕТ" В МОДЕЛИ: "INF5"
Распознаваемый объект: 1-2016-02-16 00:00:00+00:00. Интегральный критерий: "Сумма знаний". Сплайны Безье
"Прогнозирование на финансовых рынках с применением сценарного АСК-анализа и системы "Эйдос""
Прогнозируемое значение шкалы: "OPEN-FUTURES"
1648.45
1566.15
1483.85
1401.57
1319.27
1236.98
1154.68
1072.39
990.09
907.80
825.50
1.0
1.2
1.4
1.6
1.8
2.0
2.2
2.4
2.6
2.8
3.0
Шкала времени на период прогнозирования= 3
c:\Aidos-X\AvrScen-Inf5-i-Neg-1-Splain.jpg
30.07.2021-03:44:09

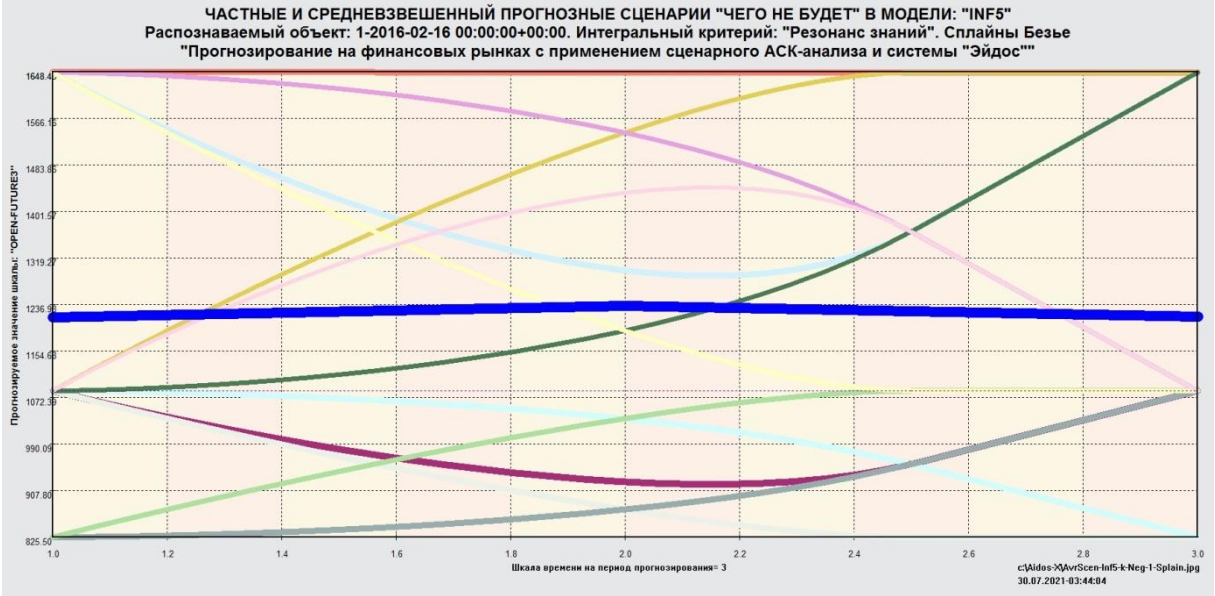

ЧАСТНЫЕ И СРЕДНЕВЗВЕШЕННЫЙ ПРОГНОЗНЫЕ СЦЕНАРИИ "ЧТО БУДЕТ" В МОДЕЛИ: "INF5"
Распознаваемый объект: 1-2016-02-16 00:00:00+00:00. Интегральный критерий: "Сумма знаний". Ломанные линии
"Прогнозирование на финансовых рынках с применением сценарного АСК-анализа и системы "Эйдос""
Прогнозируемое значение шкалы: "OPEN-FUTURES"
1648.45
1566.13
1483.85
1401.57
1319.27
1236.99
1154.69
1072.39
990.09
907.80
825.50
1.0
1.2
1.4
1.6
1.8
2.0
2.2
2.4
2.6
2.8
3.0
Шкала времени на период прогнозирования= 3
c:\Aidos-X\AvrScen-Inf5-i-Pos-1.jpg
30.07.2021-03:43:37


ЧАСТНЫЕ И СРЕДНЕВЗВЕШЕННЫЙ ПРОГНОЗНЫЕ СЦЕНАРИИ "ЧТО БУДЕТ" В МОДЕЛИ: "INF5"
Распознаваемый объект: 1-2016-02-16 00:00:00+00:00. Интегральный критерий: "Сумма знаний". Полиномы n-й степени
"Прогнозирование на финансовых рынках с применением сценарного АСК-анализа и системы "Эйдос""
Прогнозируемое значение шкалы: "OPEN-FUTURES"
1718.82
1626.28
1533.65
1441.07
1348.49
1255.91
1163.33
1070.75
978.16
885.58
793.00
1.0
1.2
1.4
1.6
1.8
2.0
2.2
2.4
2.6
2.8
3.0
Шкала времени на период прогнозирования= 3
c:\Aidos-X\AvrScen-Inf5-i-Pos-1-Poli.jpg
30.07.2021-03:43:55


ЧАСТНЫЕ И СРЕДНЕВЗВЕШЕННЫЙ ПРОГНОЗНЫЕ СЦЕНАРИИ "ЧТО БУДЕТ" В МОДЕЛИ: "INF5"
Распознаваемый объект: 1-2016-02-16 00:00:00+00:00. Интегральный критерий: "Сумма знаний". Сплайны Безье
"Прогнозирование на финансовых рынках с применением сценарного АСК-анализа и системы "Эйдос""
Прогнозируемое значение шкалы: "OPEN-FUTURES"
1648.45
1566.13
1483.85
1401.57
1319.27
1236.99
1154.69
1072.39
990.09
907.80
825.50
1.0
1.2
1.4
1.6
1.8
2.0
2.2
2.4
2.6
2.8
3.0
Шкала времени на период прогнозирования= 3
c:\Aidos-X\AvrScen-Inf5-i-Pos-1-Splain.jpg
30.07.2021-03:44:05


ЧАСТНЫЕ И СРЕДНЕВЗВЕШЕННЫЙ ПРОГНОЗНЫЕ СЦЕНАРИИ "ЧЕГО НЕ БУДЕТ" В МОДЕЛИ: "INF5"
Распознаваемый объект: 1-2016-02-16 00:00:00+00:00. Интегральный критерий: "Резонанс знаний". Ломанные линии
"Прогнозирование на финансовых рынках с применением сценарного АСК-анализа и системы "Эйдос""
Прогнозируемое значение шкалы: "OPEN-FUTURE3"
Шкала времени на период прогнозирования= 3
c:\Aidos-X\AvrScen-Inf5-k-Neg-1.jpg
30.07.2021-03:43:35


ЧАСТНЫЕ И СРЕДНЕВЗВЕШЕННЫЙ ПРОГНОЗНЫЕ СЦЕНАРИИ "ЧЕГО НЕ БУДЕТ" В МОДЕЛИ: "INF5"
Распознаваемый объект: 1-2016-02-16 00:00:00+00:00. Интегральный критерий: "Резонанс знаний". Полиномы n-й степени
"Прогнозирование на финансовых рынках с применением сценарного АСК-анализа и системы "Эйдос""
Прогнозируемое значение шкалы: "OPEN-FUTURE3"
Шкала времени на период прогнозирования= 3
c:\Aidos-X\AvrScen-Inf5-k-Neg-1-Poli.jpg
30.07.2021-03:43:48


ЧАСТНЫЕ И СРЕДНЕВЗВЕШЕННЫЙ ПРОГНОЗНЫЕ СЦЕНАРИИ "ЧЕГО НЕ БУДЕТ" В МОДЕЛИ: "INF5"
Распознаваемый объект: 1-2016-02-16 00:00:00+00:00. Интегральный критерий: "Резонанс знаний". Сплайны Безье
"Прогнозирование на финансовых рынках с применением сценарного АСК-анализа и системы "Эйдос""
Прогнозируемое значение шкалы: "OPEN-FUTURE3"
Шкала времени на период прогнозирования= 3
c:\Aidos-X\AvrScen-Inf5-k-Neg-1-Splain.jpg
30.07.2021-03:44:04

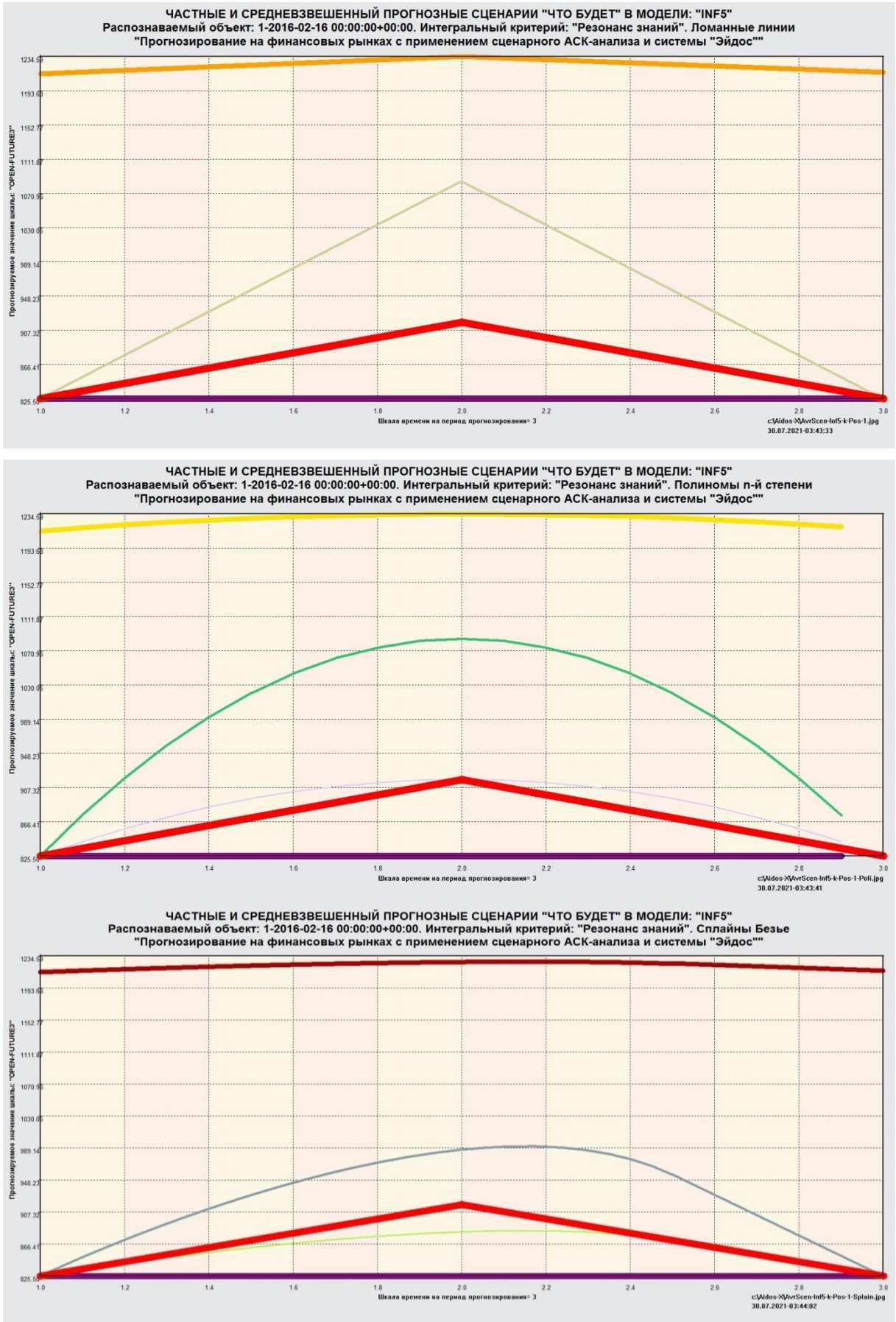


Рисунок 29. Графические диаграммы по результатам распознавания

Толщина линий прогнозируемых сценариев соответствует степени сходства ситуации на момент прогнозирования с обобщенным образом класса соответствующего сценария. Средневзвешенный сценарий получен путем суммирования прогнозируемых сценариев с их весами, как описано в предыдущей главе и в работе [6].

13.3.5.2. Подзадача 4.2. Поддержка принятия решений в простейшем варианте (SWOT-анализ)

При принятии решений определяется сила и направление влияния значений факторов на принадлежность состояний объекта моделирования к тем или иным классам, соответствующим различным будущим состояниям. В простейшем варианте принятие решений это, по сути, решение задачи SWOT-анализа [12]. Применительно к задаче, решаемой в данной работе, SWOT-анализ показывает степень влияния различных значений характеристик финансового рынка на курсы открытия и закрытия акций компании Гугл и динамику этих курсов. В системе «Эйдос» в режиме 4.4.8 поддерживается решение этой задачи. При этом ***выявляется система детерминации заданного класса***, т.е. система значений факторов, обуславливающих переход объекта моделирования и управления в состояние, соответствующее данному классу, а также препятствующих этому переходу. Приводится также степень влияния значений факторов на результат. На рисунках 21 приведены примеры некоторых SWOT-диаграмм, наглядно отражающих силу и направление влияния различных значений характеристик финансового рынка на курсы открытия и закрытия акций компании Гугл и на динамику этих курсов:

Экранные формы, приведенные на рисунках 23, содержат все необходимые пояснения и интуитивно понятны.

Отметим также, что система «Эйдос» обеспечивала решение этой задачи ***всегда***, т.е. даже в самых ранних DOS-версиях и в реализациях системы «Эйдос» на других языках и типах компьютеров. Например, первый акт внедрения системы «Эйдос», где об этом упоминается в явном виде, датируется 1987 годом, а первый подобный расчет относится к 1981 году. Но тогда SWOT-диаграммы назывались позитивным и негативным информационными портретами классов.

Информация о системе значений факторов, обуславливающих переход объекта моделирования в различные будущие состояния, соответствующие классам, может быть приведена не только в диаграммах, показанных на рисунках 21, но и во многих других табличных и графических выходных формах, которые в данной работе не приводятся только из-за ограничений на ее объем. В частности в этих формах может быть выведена значительно более полная информация (в т. ч. вообще вся имеющая в модели). Подобная подробная информация содержится в базах данных, расположенных по пути: \Aidos-

X\AID_DATA\A0000001\System\SWOTCls####Inf5.DBF, где: «####» – код класса с ведущими нулями. Эти базы открываются в MS Excel.

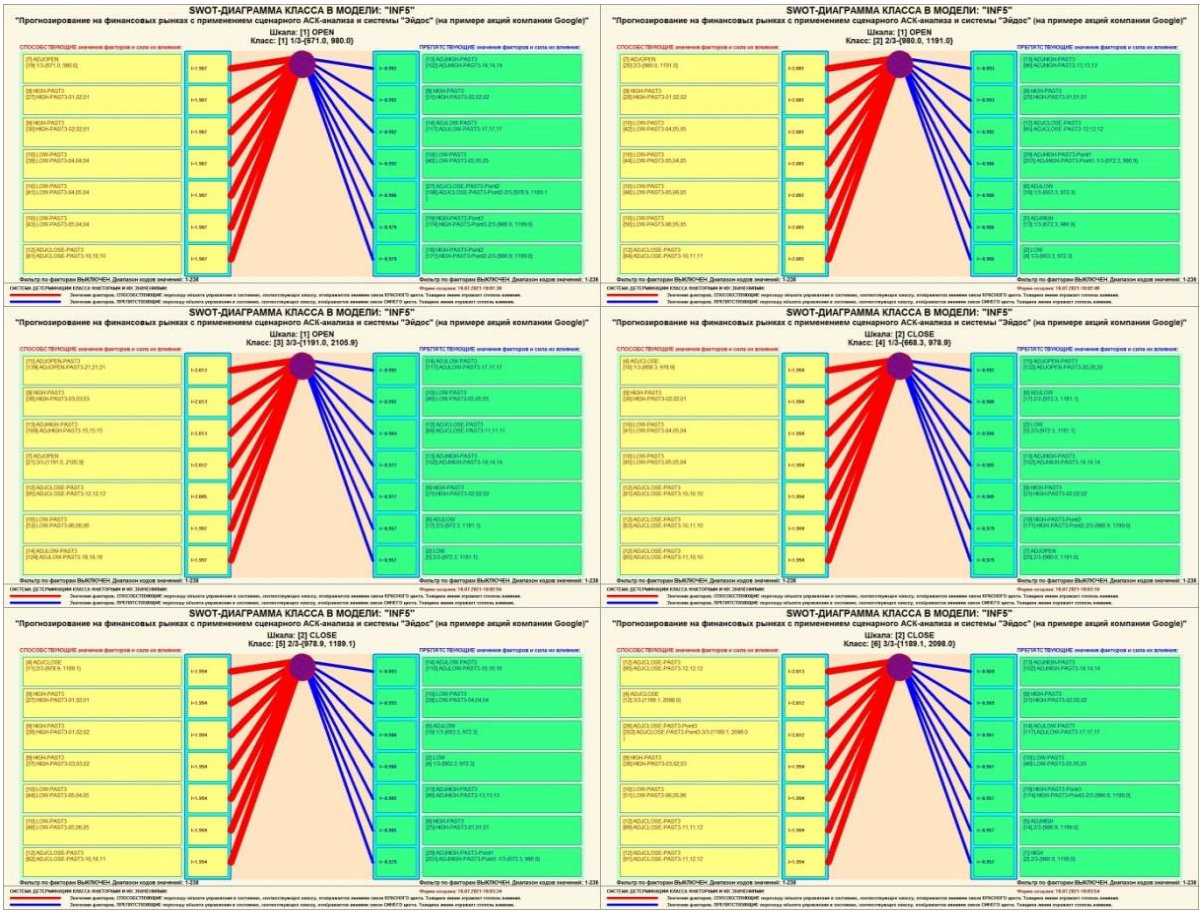


Рисунок 30. SWOT-диаграммы детерминации курсов открытия и закрытия акций компании Гугл и динамику этих курсов[33]

На рисунке 24 приведены примеры нескольких инвертированных SWOT-диаграмм (предложены автором [12]), отражающих силу и направление влияния различных характеристик финансового рынка на курсы открытия и закрытия акций компании Гугл и на динамику этих курсов.

Из инвертированных SWOT-диаграмм, приведенных на рисунке 22, видно, как влияют различные значения характеристик финансового рынка на курсы открытия и закрытия и динамику курсов компании Гугл.

Отметим, что аналогичные инвертированные SWOT-диаграммы могут быть получены для всех характеристик финансового рынка и здесь они не приводятся только из-за ограничений на объем работы. Но они могут получены любым желающим, если он скачает систему «Эйдос» с сайта ее автора и разработчика проф.Е.В.Луценко по ссылке:

[33] Не смотря на малый размер рисунков в работе они вполне читабельны при просмотре текста работы в увеличенном масштабе, например при масштабе 200% или 500%.

http://lc.kubagro.ru/aidos/_Aidos-X.htm, установит ее на своем компьютере, а затем в режиме 1.3 установит интеллектуальное облачное Эйдос-приложение №**295**, просчитает модели в режиме 3.5 и перейдет в режим 4.4.9.

УТВЕРЖДАЮ
Заведующий Краснодарским сектором ИСИ АН СССР, к.ф.н.
_____ А.А.Хагуров
_____ 1987г.

УТВЕРЖДАЮ
Директор Северо-Кавказского филиала ВНИЦ "АИУС-агроресурсы", к.э.н.
_____ Э.М.Трахов
19.05 1987г.

А К Т

Настоящий акт составлен комиссией в составе: Кириченко М.М., Ляшко Г.А., Самсонов Г.А., Коренец В.И., Луценко Е.В. в том, что в соответствии с договором о научно-техническом сотрудничестве между Сезеро-Кавказским филиалом ВНИЦ "АИУС-агроресурсы" и Краснодарским сектором Института социологических исследований АН СССР Северо-Кавказским филиалом ВНИЦ "АИУС-агроресурсы" выполнены следующие работы:

- осуществлена постановка задачи: "Обработка на ЭВМ социологических анкет Крайагропрома";
- разработаны математическая модель и программное обеспечение подсистемы распознавания образов, позволяющие решать данную задачу в среде персональной технологической системы ВЕГА-М;
- на профессиональной персональной ЭВМ "Искра-226" осуществлены расчёты по задаче в объёме:

Входная информация составила 425 анкет по 9-ти предприятиям.
Выходная информация - 4 вида выходных форм объёмом 90 листов формата А3 и 20 листов формата А4 содержит:

- процентное распределение ответов в разрезе по социальным типам корреспондентов;
- распределение информативностей признаков (в битах) для распознавания социальных типов корреспондентов;
- позитивные и негативные информационные портреты 30-ти социальных типов на языке 212 признаков;
- обобщённая характеристика информативности признаков для выбора такого минимального набора признаков, который содержит максимум информации о распознаваемых объектах (оптимизация анкет).

Работы выполнены на высоком научно-методическом уровне и в срок.

От ИСИ АН СССР:

Мл.научный сотрудник
_____ М.М.Кириченко
19.05 1987г.

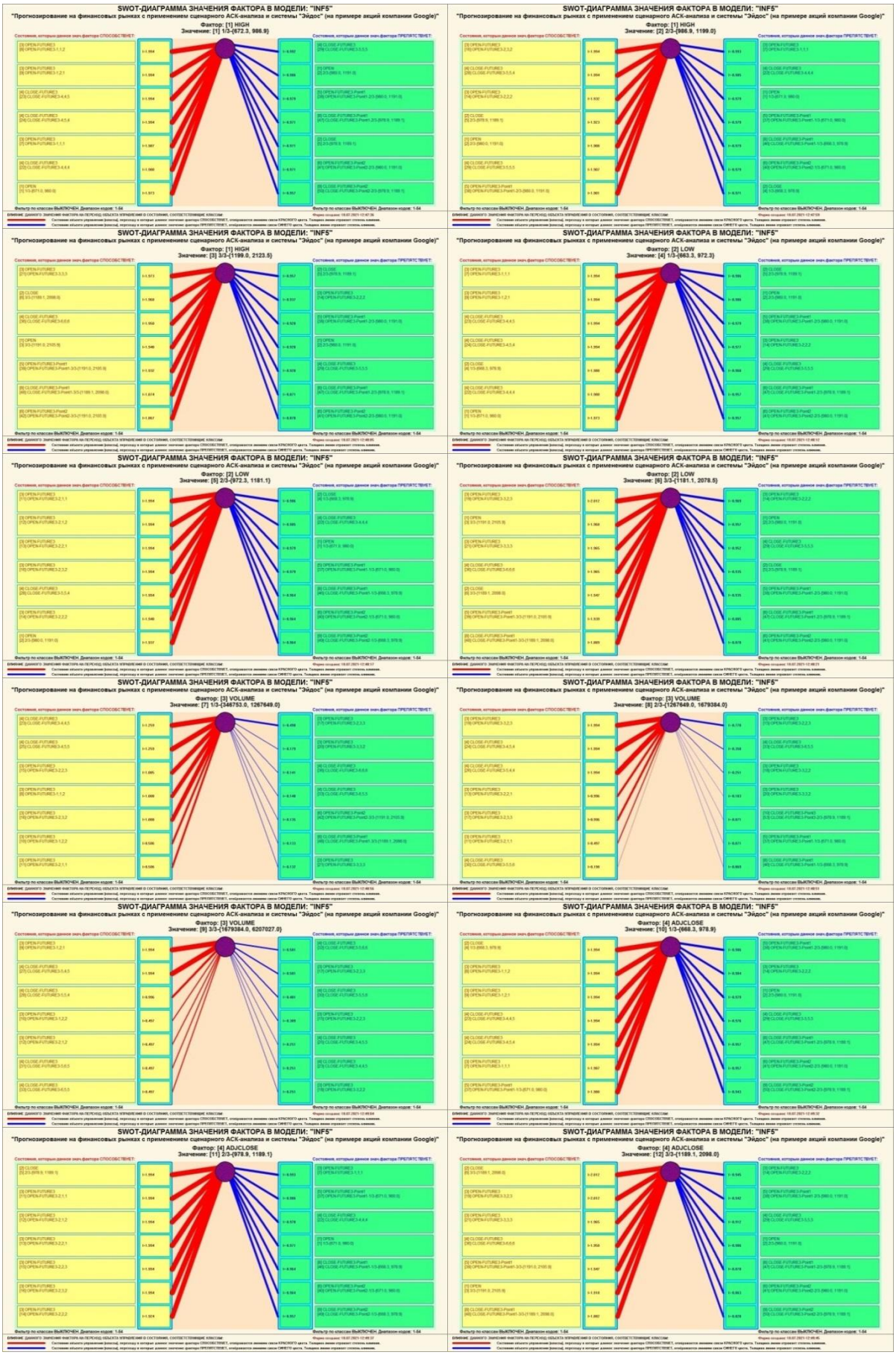
Мл.научный сотрудник
_____ Г.А.Ляшко
19.05 1987г.

От СКФ ВНИЦ "АИУС-агроресурсы":

Зав.отделом аэрокосмических и тематических изысканий №4, к.э.н.
_____ Г.А.Самсонов
19.05. 1987г.

Главный конструктор проекта
_____ В.И.Коренец
19.05.87г. 1987г.

Главный конструктор проекта
_____ Е.В.Луценко
19.05.87 1987г.


SWOT-ДИАГРАММА ЗНАЧЕНИЯ ФАКТОРА В МОДЕЛИ: "INF5"
"Прогнозирование на финансовых рынках с применением сценарного АСК-анализа и системы "Эйдос" (на примере акций компании Google)"
Фактор: [1] HIGH
Значение: [1] 1/3-(672.3, 986.9)
SWOT-ДИАГРАММА ЗНАЧЕНИЯ ФАКТОРА В МОДЕЛИ: "INF5"
"Прогнозирование на финансовых рынках с применением сценарного АСК-анализа и системы "Эйдос" (на примере акций компании Google)"
Фактор: [1] HIGH
Значение: [2] 2/3-(986.9, 1199.0)
SWOT-ДИАГРАММА ЗНАЧЕНИЯ ФАКТОРА В МОДЕЛИ: "INF5"
"Прогнозирование на финансовых рынках с применением сценарного АСК-анализа и системы "Эйдос" (на примере акций компании Google)"
Фактор: [1] HIGH
Значение: [3] 3/3-(1199.0, 2123.5)
SWOT-ДИАГРАММА ЗНАЧЕНИЯ ФАКТОРА В МОДЕЛИ: "INF5"
"Прогнозирование на финансовых рынках с применением сценарного АСК-анализа и системы "Эйдос" (на примере акций компании Google)"
Фактор: [2] LOW
Значение: [4] 1/3-(663.3, 972.3)
SWOT-ДИАГРАММА ЗНАЧЕНИЯ ФАКТОРА В МОДЕЛИ: "INF5"
"Прогнозирование на финансовых рынках с применением сценарного АСК-анализа и системы "Эйдос" (на примере акций компании Google)"
Фактор: [2] LOW
Значение: [5] 2/3-(972.3, 1181.1)
SWOT-ДИАГРАММА ЗНАЧЕНИЯ ФАКТОРА В МОДЕЛИ: "INF5"
"Прогнозирование на финансовых рынках с применением сценарного АСК-анализа и системы "Эйдос" (на примере акций компании Google)"
Фактор: [2] LOW
Значение: [6] 3/3-(1181.1, 2078.5)
SWOT-ДИАГРАММА ЗНАЧЕНИЯ ФАКТОРА В МОДЕЛИ: "INF5"
"Прогнозирование на финансовых рынках с применением сценарного АСК-анализа и системы "Эйдос" (на примере акций компании Google)"
Фактор: [3] VOLUME
Значение: [7] 1/3-(346753.0, 1267649.0)
SWOT-ДИАГРАММА ЗНАЧЕНИЯ ФАКТОРА В МОДЕЛИ: "INF5"
"Прогнозирование на финансовых рынках с применением сценарного АСК-анализа и системы "Эйдос" (на примере акций компании Google)"
Фактор: [3] VOLUME
Значение: [8] 2/3-(1267649.0, 1679384.0)
SWOT-ДИАГРАММА ЗНАЧЕНИЯ ФАКТОРА В МОДЕЛИ: "INF5"
"Прогнозирование на финансовых рынках с применением сценарного АСК-анализа и системы "Эйдос" (на примере акций компании Google)"
Фактор: [3] VOLUME
Значение: [9] 3/3-(1679384.0, 6207027.0)
SWOT-ДИАГРАММА ЗНАЧЕНИЯ ФАКТОРА В МОДЕЛИ: "INF5"
"Прогнозирование на финансовых рынках с применением сценарного АСК-анализа и системы "Эйдос" (на примере акций компании Google)"
Фактор: [4] ADJCLOSE
Значение: [10] 1/3-(668.3, 978.9)
SWOT-ДИАГРАММА ЗНАЧЕНИЯ ФАКТОРА В МОДЕЛИ: "INF5"
"Прогнозирование на финансовых рынках с применением сценарного АСК-анализа и системы "Эйдос" (на примере акций компании Google)"
Фактор: [4] ADJCLOSE
Значение: [11] 2/3-(978.9, 1189.1)
SWOT-ДИАГРАММА ЗНАЧЕНИЯ ФАКТОРА В МОДЕЛИ: "INF5"
"Прогнозирование на финансовых рынках с применением сценарного АСК-анализа и системы "Эйдос" (на примере акций компании Google)"
Фактор: [4] ADJCLOSE
Значение: [12] 3/3-(1189.1, 2098.0)

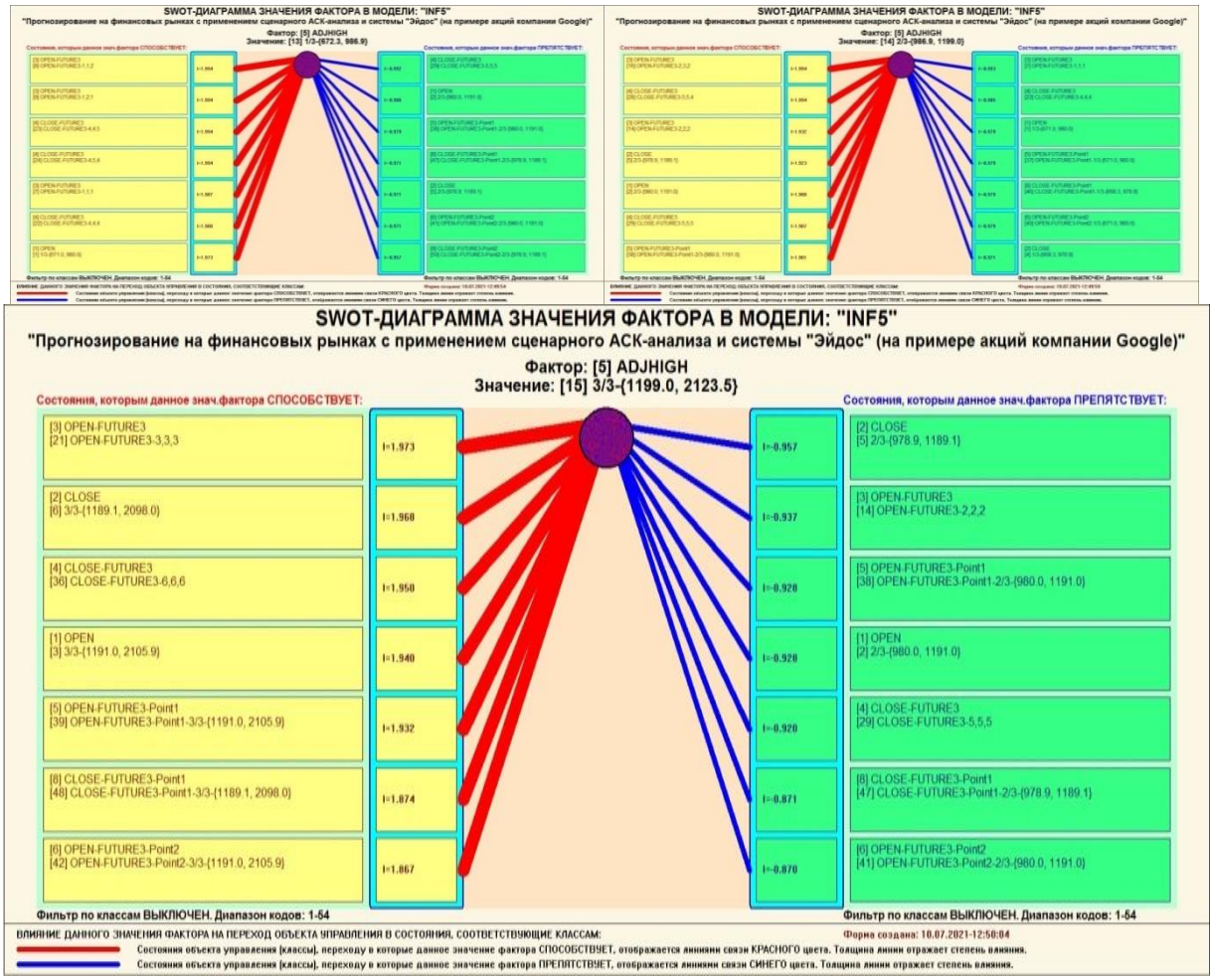


Рисунок 31. Примеры SWOT-диаграмм, отражающих силу и направление влияния различных значений характеристик финансового рынка на курсы открытия и закрытия и динамику курсов компании Гугл[34]

В заключение отметим, что SWOT-анализ является широко известным и общепризнанным метод стратегического планирования. Однако это не мешает тому, что он подвергается критике, часто вполне справедливой, обоснованной и хорошо аргументированной. В результате критического рассмотрения SWOT-анализа в полном соответствии с методологией SWOT-анализа выявлено довольно много его слабых и сильных сторон.

В частности, по мнению автора, основным недостатком SWOT-анализа является необходимость привлечения экспертов как для выбора самой системы факторов, так и для и оценки силы и направления влияния этих факторов на результат.

Ясно, что эксперты это делают неформализуемым путем на основе своего опыта, интуиции и профессиональной компетенции, т.е. грубо

[34] Не смотря на малый размер рисунков в работе они вполне читабельны при просмотре текста работы в увеличенном масштабе, например при масштабе 200% или 500%.

говоря «от фонаря». Если честно, чаще всего этими экспертами являются сами авторы работ, обычно студенты, магистранты и аспиранты, которых трудно заподозрить в том, что они реально являются экспертами в какой-либо предметной области (кроме одной).

Возможности привлечения экспертов имеют свои естественные ограничения, финансовые временные, организационные и другие. Кроме того часто по различным причинам эксперты не могут или не хотят сообщать свои способы принятия решений.

Иногда даже встречаются ситуации, когда сообщение экспертом когнитологу своего подхода к принятию решений можно считать чистосердечным признанием, смягчающим наказание по определенным статьям.

Таким образом, возникает проблема проведения SWOT-анализа без привлечения экспертов. Эта проблема решается путем автоматизации путем автоматизации функций экспертов в SWOT-анализе, т.е. путем создания непосредственно на основе эмпирических данных моделей, обесчещивающих измерения силы и направления влияния факторов на результаты. Подобная технология разработана давно, ей уже более 30 лет, но, к сожалению, единственная система, в которой это реализовано, сравнительно малоизвестна (это интеллектуальная система «Эйдос»).

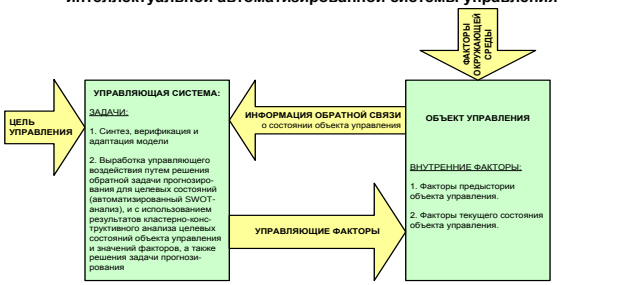
13.3.5.3. Подзадача 4.2. Развитый алгоритм принятия решений

В предыдущем разделе кратко описан вариант приятия решений путем применения когнитивного автоматизированного SWOT-анализа. Однако по трем основным причинам SWOT-анализ можно рассматривать как метод принятия решений только лишь в очень упрощенной форме:

1) В SWOT-анализе рассматривается лишь одно целевое будущее состояние, а их может быть очень много. Например, эффективность фирмы можно измерять в *натуральном и стоимостном выражении* и по каждому из этих вариантов может быть очень много показателей (количество и качество различных видов продукции, прибыль и рентабельность и др.);
2) Неизвестно, корректно ли поставлены цели управления, т.е. достижимы ли целевые состояния одновременно, т.е. являются ли они совместимыми по системе обуславливающих значений факторов (системе детерминации), или они являются недостижимыми одновременно, альтернативными.
3) Все значения факторов, рекомендуемые в WSOT-анализе, необходимо использовать для достижения целевого состояния. Однако некоторые из них может не быть физической или финансовой возможности использовать. Что в этом случае делать не совсем понятно.

В развитом алгоритме принятия решений в интеллектуальных системах управления на основе АСК-анализа и системы «Эйдос» все эти проблемы решены (рисунок 25).

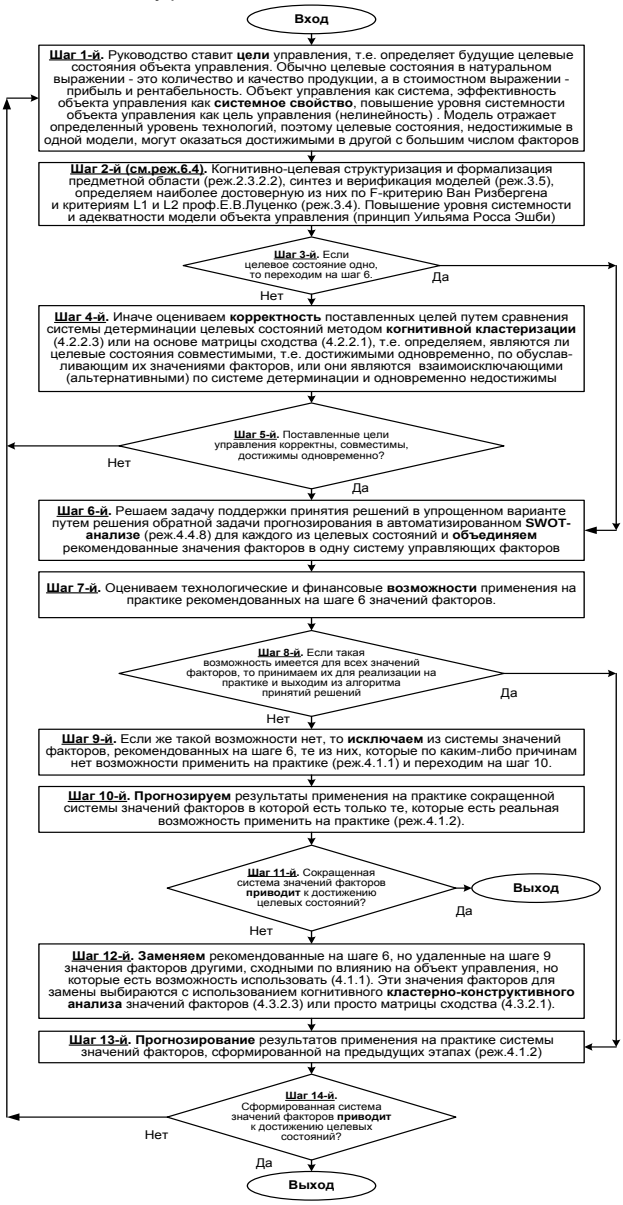
Принципиальная схема замкнутой адаптивной интеллектуальной автоматизированной системы управления



Луценко Е.В. Автоматизация Функционально-стоимостного анализа и метода "Директ-костинг" на основе АСК-анализа и системы "Эйдос" (автоматизация управления натуральной и финансовой эффективностью затрат без содержательных технологических и финансово-экономических расчетов на основе информационных и когнитивных технологий и теории управления) / Е.В.Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. - Краснодар: КубГАУ, 2017. - №07(131). С. 1-18.-IDA [article ID): 1311707001.-Режим доступа: http://ej.kubagro.ru/2017/07/pdf/01.pdf, 1,125 у.п.л.

1. Решения всегда принимаются на основе модели. Модели могут быть различной степени формализации: интуитивные (субъективные осознанные и неосознанные) неформализованные модели, вербализованные модели, лингвистические модели (различные структуры текста), алгоритмические модели и модели данных, статистические и информационные модели, математические (аналитические) модели. Формализация нужна чтобы передавать модели людям и технических системам.
2. Виды управления: оперативное, тактическое, стратегическое. Что это значит в экономических и технических системах управления.
3. Различие между АСУ и САУ: участие человека в реальном времени в принятии решений. Кто несет ответственность за ошибочные решения. Адаптивность: принцип дуальности управления Александра Фельдбаума.
4. Критерий различия управляющих факторов от факторов окружающей среды с точки зрения управляющей системы и объекта управления. Иерархическая структура окружающей среды. Мы прогнозируем курс рубля на завтра, а ЦБ принимает решение об этом, для нас это фактор окружающей среды, а для ЦБ - это управляющий фактор.
5. Решение задачи принятия решений путем **многократного** многовариантного решения задачи прогнозирования быстро приводит к **комбинаторному взрыву** при увеличении количества факторов. Обычно в реальных задачах очень большое количество факторов. Поэтому при реальном количестве факторов задача принятия решений может быть решена только путем решения обратной задачи прогнозирования, т.е. SWOT-анализа. **Однако** в SWOT-анализе задается только **одно** целевое состояние и некоторые рекомендуемые значения факторов **не могут** быть применены по технологическим и финансовым причинам. Поэтому необходимо их исключить или заменить на основе результатов кластерно-конструктивного анализа значений факторов и оценить адекватность такого варианта решения путем прогнозирования результатов применения такой измененной системы значений факторов.

Развитый алгоритм принятия решений в адаптивных интеллектуальных системах управления на основе АСК-анализа и системы «Эйдос»



Луценко Е.В. Системное обобщение принципа Эшби и повышение уровня системности модели объекта познания как необходимое условие адекватности процесса его познания / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. - Краснодар: КубГАУ, 2020. - №59(15). С. 109 - 134. - IDA [article ID]: 1632059009. - Режим доступа: http://ej.kubagro.ru/2020/05/pdf/09.pdf, 2,188 у.п.л.

Луценко Е.В. Эффективность объекта управления как его эмерджентное свойство и повышение уровня системности как цель управления / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. - Краснодар: КубГАУ, 2021. - №01(165). С. 77 - 98. - IDA [article ID]: 1652101009. - Режим доступа: http://ej.kubagro.ru/2021/01/pdf/05.pdf, 1,375 у.п.л.

Луценко Е.В. Метризация измерительных шкал различных типов и совместная сопоставимая количественная обработка разнородных факторов в системно-когнитивном анализе и системе «Эйдос» / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. - Краснодар: КубГАУ, 2013. - №08(092). С. 859 - 883. - IDA [article ID]: 0921308058. - Режим доступа: http://ej.kubagro.ru/2013/08/pdf/58.pdf, 1,562 у.п.л.

Луценко Е.В. Инвариантное относительно объемов данных нечеткое мультиклассовое обобщение F-меры достоверности моделей Ван Ризбергена в АСК-анализе и системе «Эйдос» / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. - Краснодар: КубГАУ, 2017. - №02(126). С. 1-32. - IDA [article ID]: 1261702001. - Режим доступа: http://ej.kubagro.ru/2017/02/pdf/01.pdf, 2 у.п.л.

Луценко Е.В. Метод когнитивной кластеризации или кластеризация на основе знаний (кластеризация в системно-когнитивном анализе и интеллектуальной системе «Эйдос») / Е.В. Луценко, В.Е. Коржаков // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. - Краснодар: КубГАУ, 2011. - №07(071). С. 528 - 576. - Шифр Информрегистра: 0421100012\0253, IDA [article ID]: 0711107040. - Режим доступа: http://ej.kubagro.ru/2011/07/pdf/40.pdf, 3,062 у.п.л.

Луценко Е.В. Количественный автоматизированный SWOT- и PEST-анализ средствами АСК-анализа и интеллектуальной системы «Эйдос-Х++» / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. - Краснодар: КубГАУ, 2014. - №07(101). С. 1367 - 1409. - IDA [article ID]: 1011407090. - Режим доступа: http://ej.kubagro.ru/2014/07/pdf/90.pdf, 2,688 у.п.л.

Lutsenko E.V. Scenario and spectral automated system-cognitive analysis // July 2021, DOI: 10.13140/RG.2.2.22981.37608, License CC BY-SA 4.0, https://www.researchgate.net/profile/Eugene-Lutsenko/publication/353405096

Луценко Е.В. Метод когнитивной кластеризации или кластеризация на основе знаний (кластеризация в системно-когнитивном анализе и интеллектуальной системе «Эйдос») / Е.В. Луценко, В.Е. Коржаков // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. - Краснодар: КубГАУ, 2011. - №07(071). С. 528 - 576. - Шифр Информрегистра: 0421100012\0253, IDA [article ID]: 0711107040. - Режим доступа: http://ej.kubagro.ru/2011/07/pdf/40.pdf, 3,062 у.п.л.

Lutsenko E.V. Scenario and spectral automated system-cognitive analysis // July 2021, DOI: 10.13140/RG.2.2.22981.37608, License CC BY-SA 4.0, https://www.researchgate.net/profile/Eugene-Lutsenko/publication/353405096

Рисунок 32. Развитый алгоритм принятия решений в интеллектуальных системах управления на основе АСК-анализа и системы «Эйдос»

Этот алгоритм полностью реализуется средствами системы «Эйдос» и обеспечивает корректное и обоснованное принятие управленческих решений в реальных ситуациях.

Подробное пояснение данного алгоритма (который в принципе и так вполне понятен) не входит в задачи данной работы и дано в других работах автора, например [13], а также в видеозанятиях:

– в Пермском национальном университете:
https://bigbluebutton.pstu.ru/b/w3y-2ir-ukd-bqn

– в Кубанском государственном университете и Кубанском государственном аграрном университете:
https://disk.yandex.ru/d/knISAD5qzV83Ng?w=1.

13.3.5.4. Подзадача 4.3. Исследование моделируемой предметной области путем исследования ее модели

Если модель предметной области достоверна, то исследование модели можно считать исследованием самого моделируемого объекта, т.е. результаты исследования модели корректно относить к самому объекту моделирования, «переносить на него».

В системе «Эйдос» есть довольно много возможностей для такого исследования, но в данной работе из-за ограничений на ее объем мы рассмотрим лишь некоторые из них: когнитивные диаграммы классов и значений факторов, агломеративная когнитивная кластеризация классов и значений факторов, нелокальные нейроны и нейронные сети, 3d-интегральные когнитивные карты, когнитивные функции), исследование силы и направления влияния факторов и степени детерминированности классов, обуславливающими их значениями факторов.

13.3.5.4.1. Когнитивные диаграммы классов

Эти диаграммы отражают сходство/различие классов. Мы получаем их в режимах 4.2.2.1 и 4.2.2.2 (рисунок 26).

Отметим также, что на когнитивной диаграмме, приведенной на рисунке 26, показаны ***количественные*** оценки сходства/различие рисков невозврата ссуды по связанным с ними значениям характеристик ссудополучателей. Важно, что эти результаты сравнения получены с применением системно-когнитивной модели, созданной *непосредственно на основе эмпирических данных*, а не как традиционно делается на основе экспертных оценок неформализуемым путем на основе опыта, интуиции и профессиональной компетенции. Мы ранее уже рассматривали какие проблемы возникают при привлечении экспертов. Здесь же эти проблемы

вообще не возникают, т.к. система «Эйдос» формирует когнитивные диаграммы (по сути это сетевые нечеткие модели представления знаний) на основе моделей, создаваемый непосредственно на основе эмпирических данных.

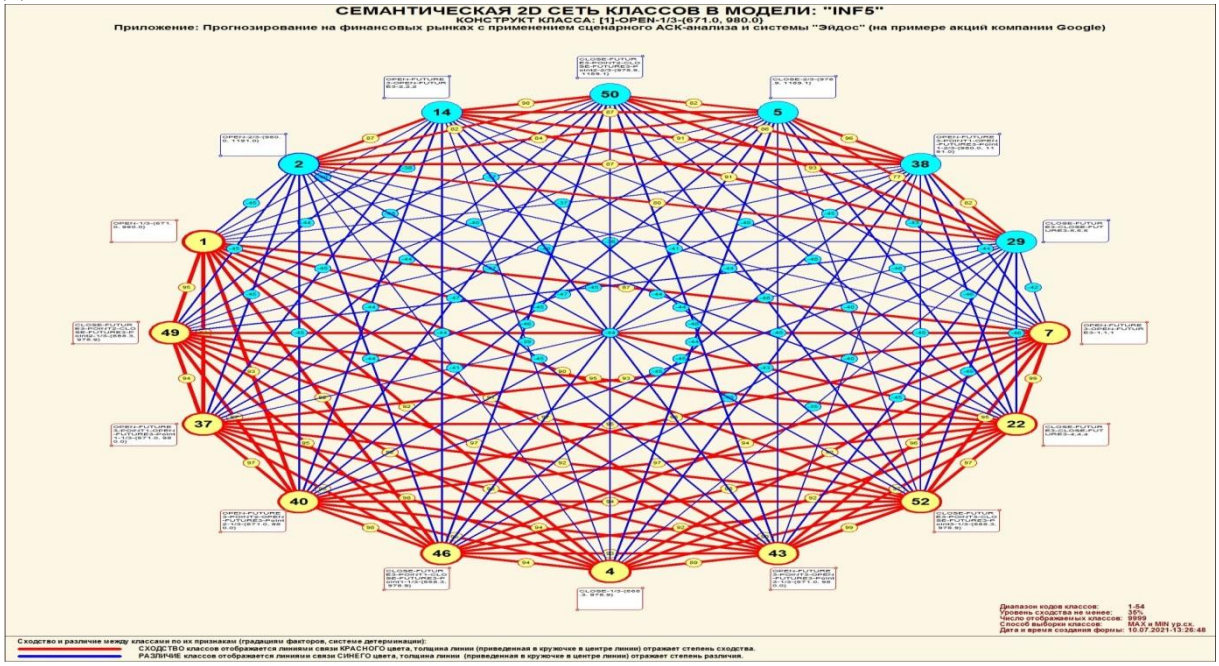


Рисунок 33. Когнитивная диаграмма классов, отражающая сходство/различие классов по их системе детерминации

В системе «Эйдос» есть возможность при необходимости управлять параметрами формирования и вывода изображения, приведенного на рисунке 24. Для этого используется диалоговое окно, приведенное на рисунке 25.

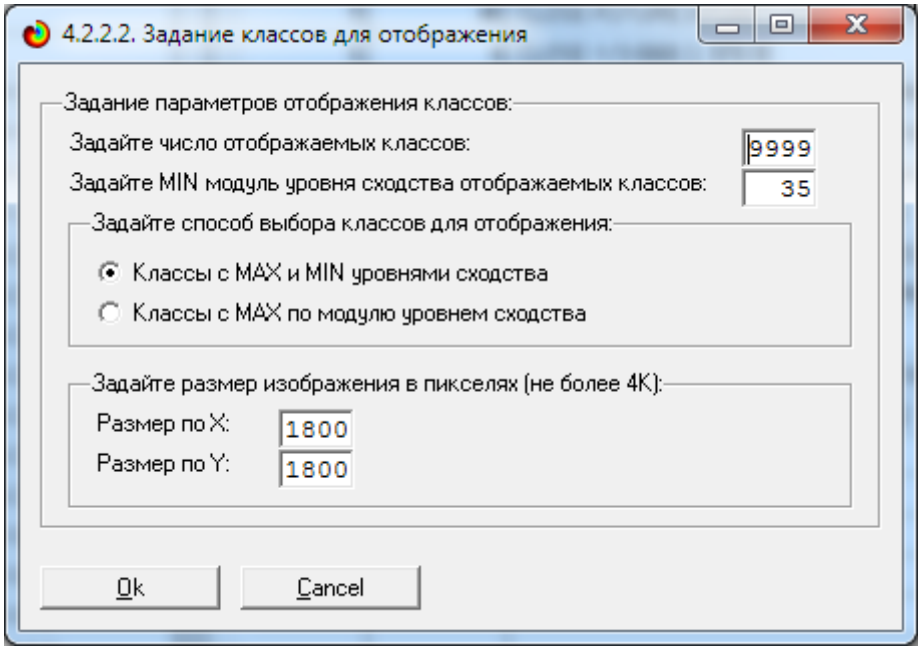


Рисунок 34. Диалоговое окно управления параметрами формирования и вывода изображения когнитивной диаграммы классов

13.3.5.4.2. Агломеративная когнитивная кластеризация классов

Информация о сходстве/различии классов, содержащаяся в матрице сходства, может быть визуализирована не только в форме, когнитивных диаграмм, пример которой приведен на рисунке 24, но и в форме агломеративных дендрограмм, полученных в результате ***когнитивной кластеризации*** (рисунок 26) [14]. На рисунке 27 мы видим график изменения межкластерных расстояний:

Из когнитивной диаграммы на рисунке 24 и дендрограммы когнитивной агломеративной кластеризации классов, приведенной на рисунке 29, мы видим, что определенные классы сходны по детерминирующей их системе значений характеристик финансового рынка, а другие сильно отличаются.

Из рисунков 24 и 26 мы видим также, что все классы образуют два противоположных кластера, являющихся полюсами конструкта, по системе значений обуславливающих их характеристик.

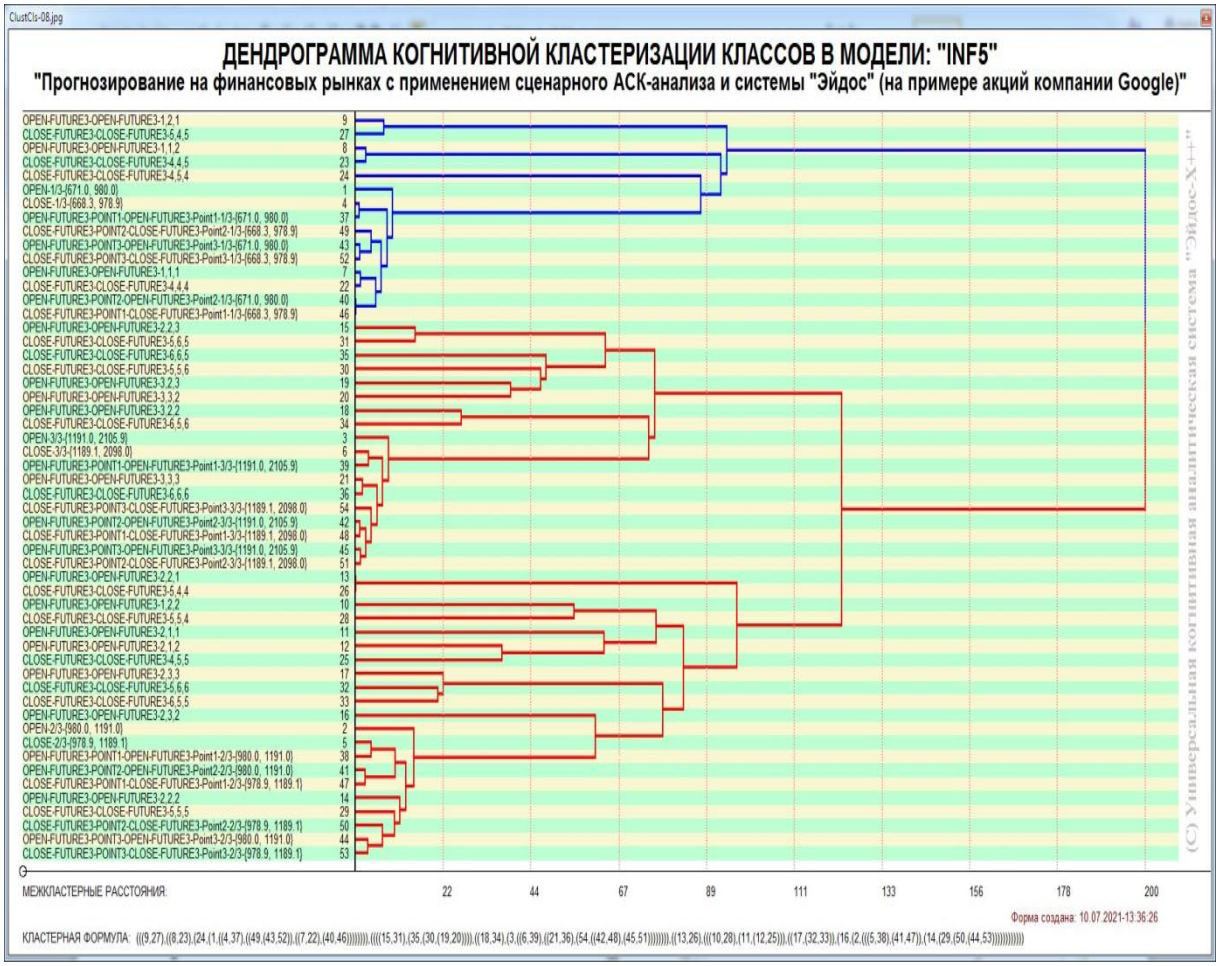


Рисунок 35. Дендрограмма когнитивной агломеративной кластеризации, отражающая сходство/различие классов по системе их детерминации

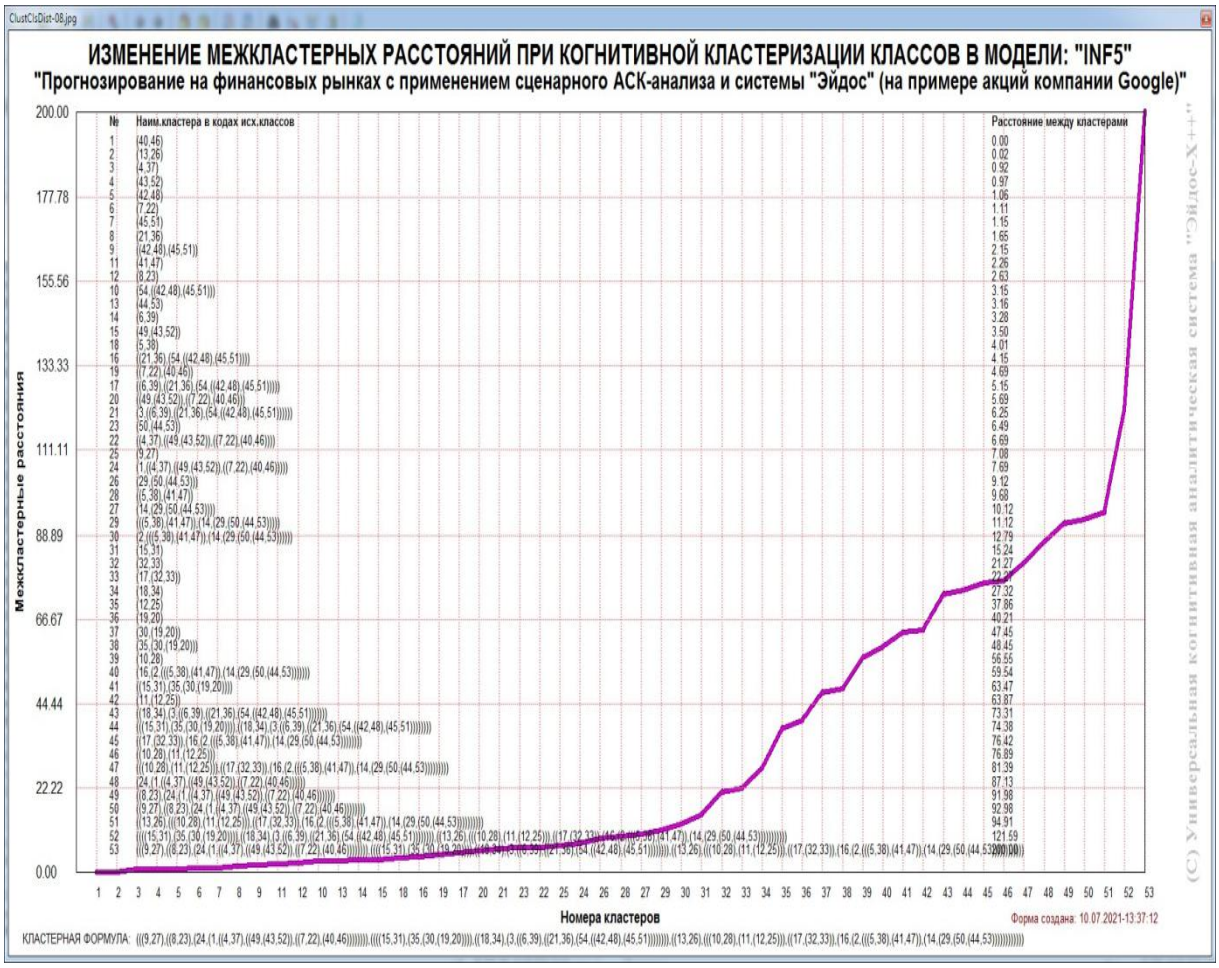


Рисунок 36. График изменения межкластерных расстояний

13.3.5.4.3. Когнитивные диаграммы значений факторов

Эти диаграммы отражают сходство/различие значений характеристик ссудополучателей по их смыслу, т.е. по содержащейся в них информации о риске невозврата ссуды.

Эти диаграммы мы получаем в режимах 4.3.2.1 и 4.3.2.2 (рисунок 28).

Из рисунка 30 видно, что все значения факторов образуют два крупных кластера, противоположных по их смыслу. Эти кластеры образуют полюса конструкта.

Отметим, что на когнитивной диаграмме, приведенной на рисунке 28, показаны ***количественные*** оценки сходства/различия значений факторов, полученные с применением системно-когнитивной модели, созданной *непосредственно на основе эмпирических данных*, а не как традиционно делается на основе экспертных оценок неформализуемым путем на основе опыта, интуиции и профессиональной компетенции. Мы ранее уже рассматривали какие проблемы возникают при привлечении экспертов. Здесь же эти проблемы вообще не возникают, т.к. система «Эйдос» формирует когнитивные диаграммы (по сути это сетевые

нечеткие модели представления знаний) на основе моделей, создаваемый непосредственно на основе эмпирических данных.

Диаграмма, приведенная на рисунке 28, получена при параметрах, приведенных на рисунке 29.

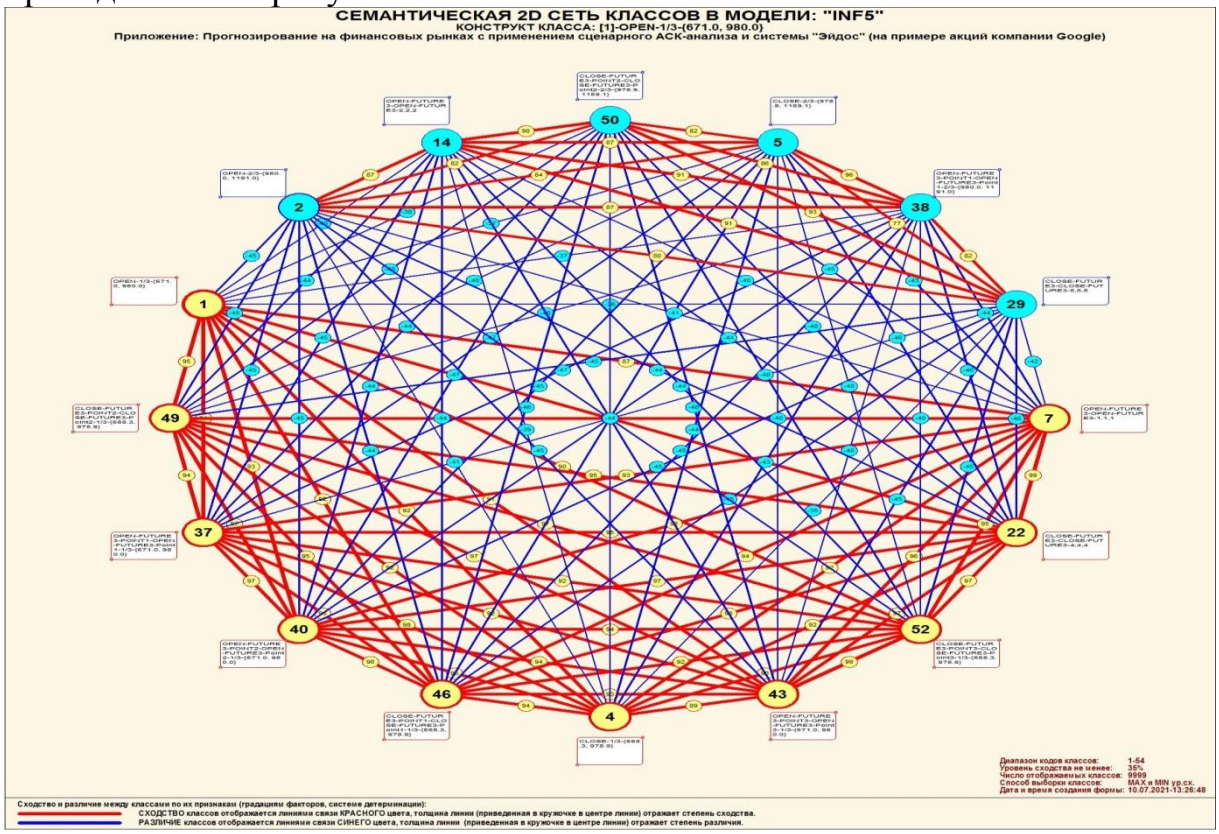


Рисунок 37. Сходство/различие характеристик ссудополучателей по их влиянию на риск невозврата ссуды

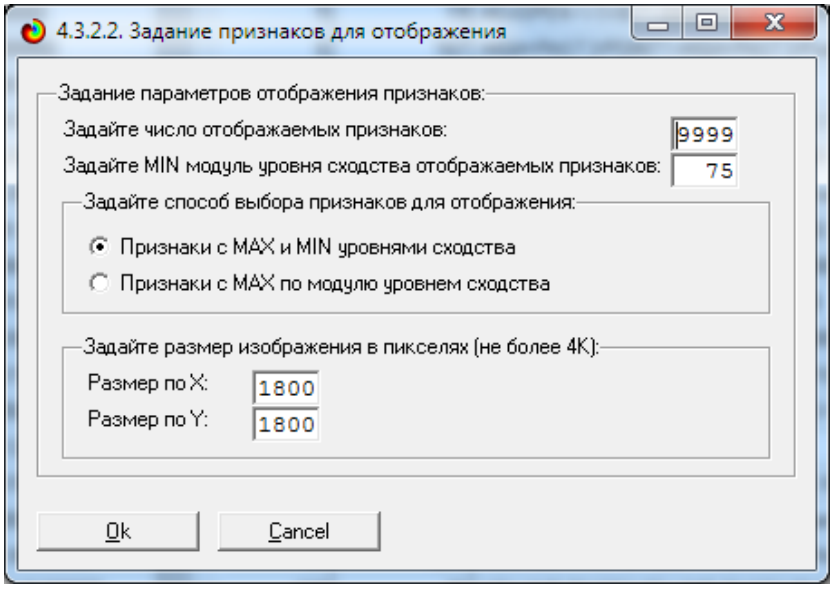


Рисунок 38. Параметры отображения когнитивной диаграммы, приведенной на рисунке 28

13.3.5.4.4. Агломеративная когнитивная кластеризация значений факторов

На рисунке 32 приведена агломеративная дендрограмма когнитивной кластеризации значений факторов [14].

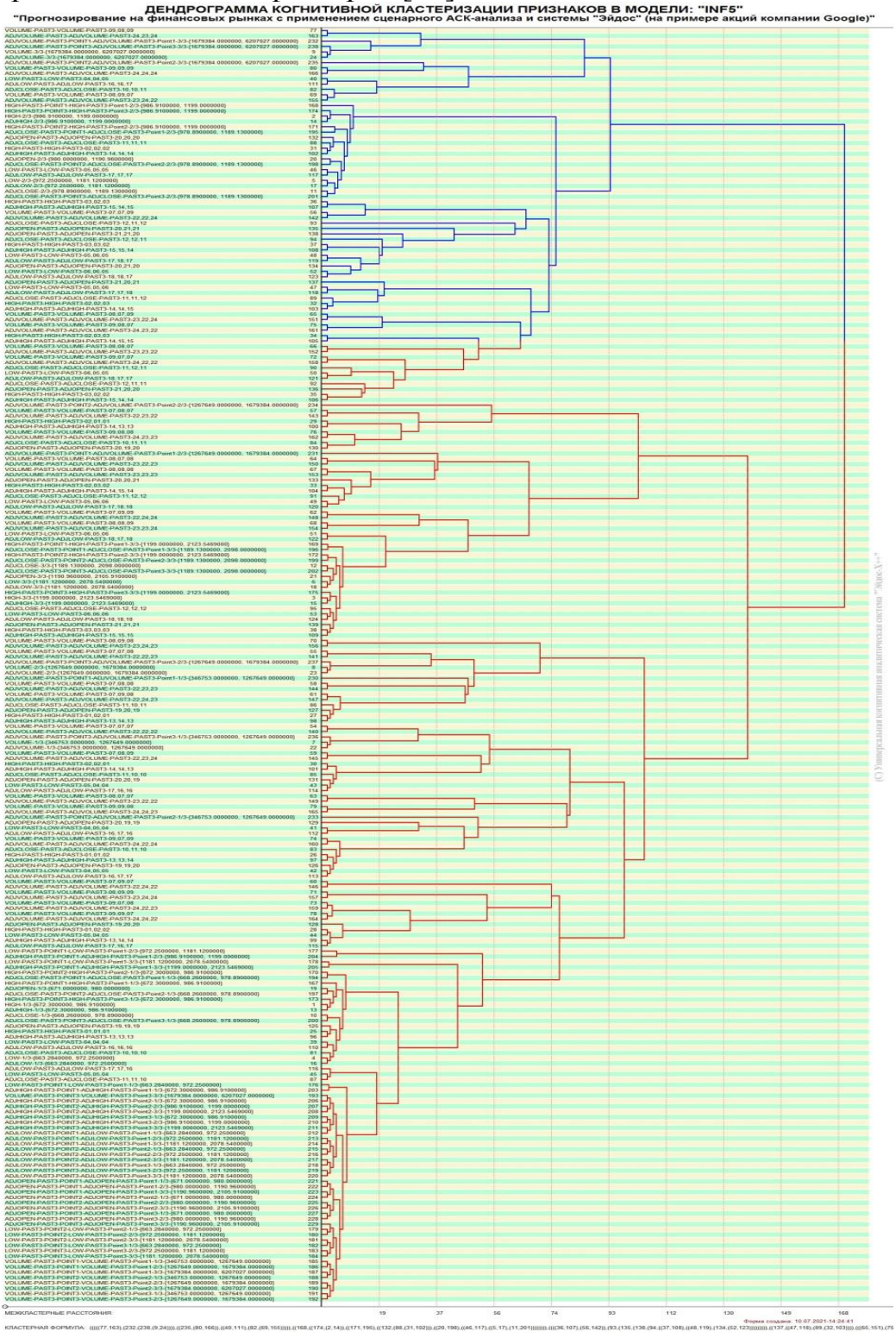


Рисунок 39. Дендрограмма агломеративной когнитивной кластеризации значений характеристик финансового рынка

Эта дендрограмма получена на основе той же матрицы сходства признаков по их смыслу, что и в когнитивных диаграммах, пример которой приведен на рисунке 30. Из дендрограммы на рисунке 32 мы видим, что все значения факторов образуют 2 четко выраженных кластера, объединенных в полюса конструкта (показаны синими и красным цветами). Хорошо видна группировка значений характеристик финансового рынка по их смыслу, т.е. по содержащейся в них информации о курсах акций компании Гугл и их динамике. ***Значения факторов на полюсах конструкта факторов (рисунки 28 30) обуславливают переход объекта моделирования в состояния, соответствующие классам, представленным на полюсах конструкта классов (рисунки 24 и 26).***

На рисунке 31 приведен график межкластерных расстояний значений признаков.

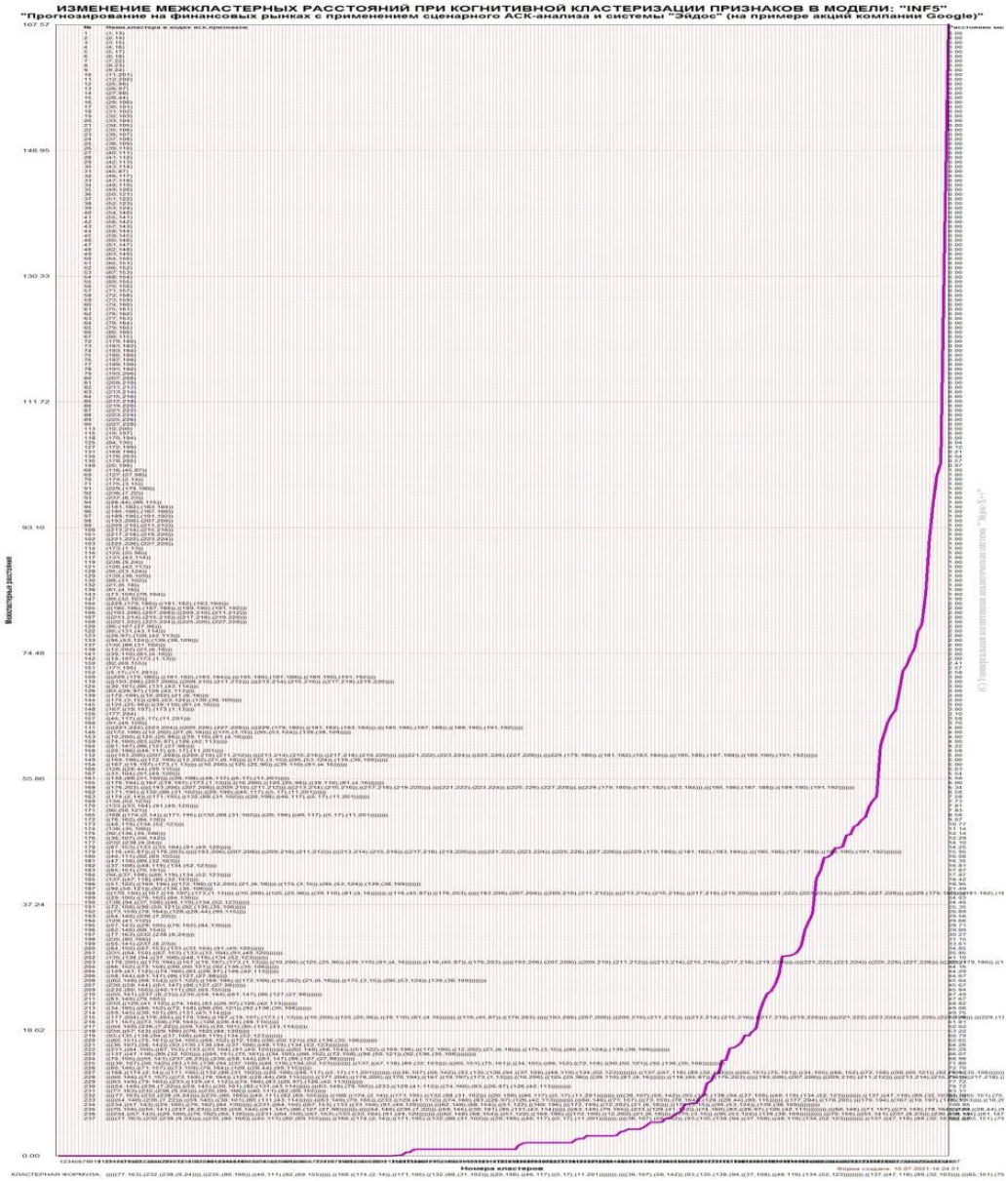


Рисунок 40. График изменения межкластерных расстояний при когнитивной кластеризации значений факторов

13.3.5.4.5. Нелокальные нейроны и нелокальные нейронные сети

На рисунке 32 приведён пример нелокального нейрона, а на рисунке 33 – фрагмент одного слоя нелокальной нейронной сети [15]:

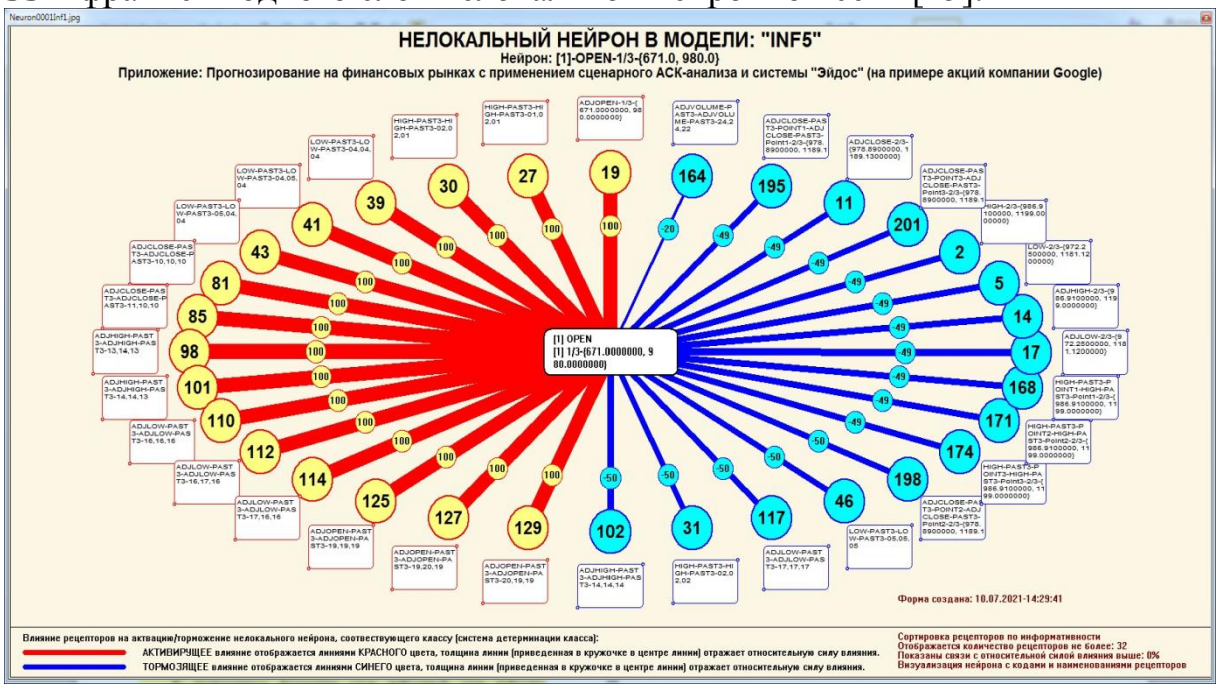


Рисунок 41. Пример нелокального нейрона, отражающего силу и направление влияния значений характеристик финансового рынка на значение курса акций компании Гугл

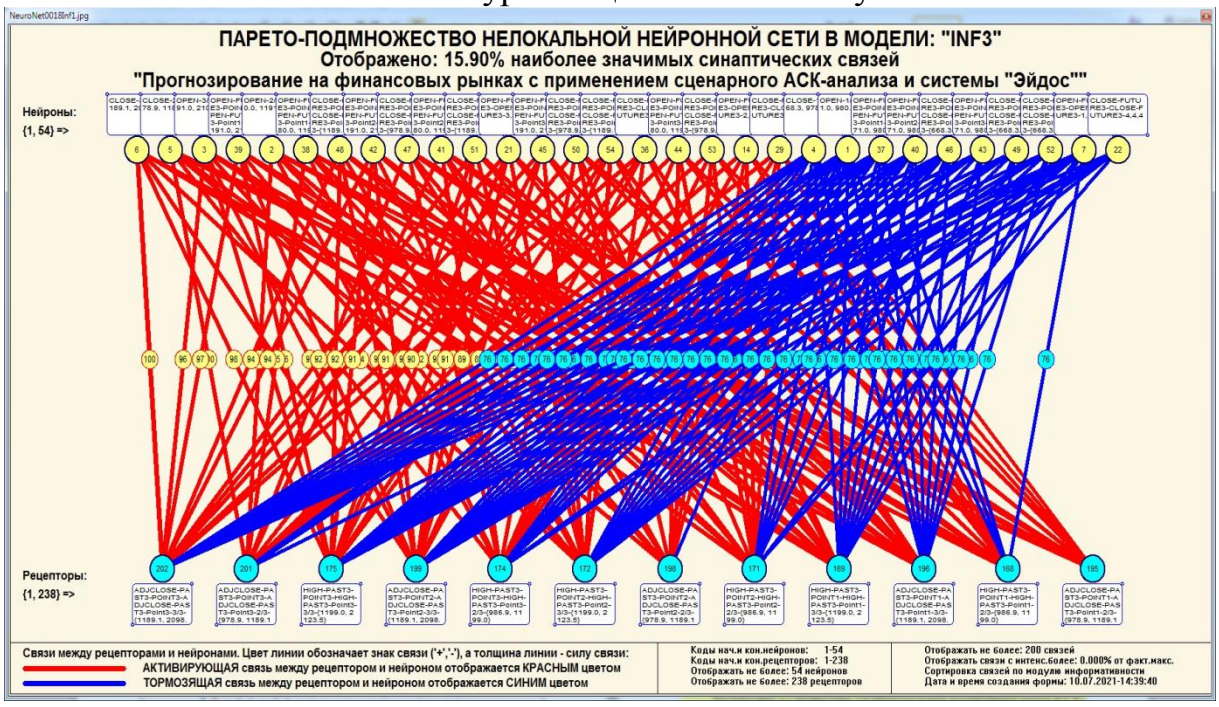


Рисунок 42. Один слой нелокальной нейронной сети, отражающий силу и направление значений характеристик финансового рынка на значения курса акций компании Гугл (фрагмент 15.9%))

В приведенном фрагменте слоя нейронной сети нейроны соответствуют классу (курсам открытия и закрытия, сценариям их изменения, значениям точек на сценариях), а рецепторы –характеристикам финансового рынка.

Нейроны на рисунке 33 расположены слева направо в порядке убывания модуля суммарной силы их детерминации, т.е. слева находятся результаты, наиболее жестко обусловленные действующими на них значениями факторов, а справа – менее жестко обусловленные.

Модель знаний системы «Эйдос» относится к ***нечетким декларативным*** гибридным моделям и объединяет в себе некоторые особенности нейросетевой и фреймовой моделей представления знаний. Классы в этой модели соответствуют нейронам и фреймам, а признаки рецепторам и шпациям (описательные шкалы – слотам).

От фреймовой модели представления знаний модель системы «Эйдос» отличается своей эффективной и простой программной реализацией, полученной за счет того, что разные фреймы отличаются друг от друга не набором слотов и шпаций, а лишь информацией в них. Поэтому в системе «Эйдос» при увеличении числа фреймов само количество баз данных не увеличивается, а увеличивается лишь их размерность.

От нейросетевой модели представления знаний модель системы «Эйдос» отличается тем, что:

1) весовые коэффициенты на рецепторах не подбираются итерационным методом обратного распространения ошибки, а считаются прямым счетом на основе хорошо теоретически обоснованной модели, основанной на теории информации (это напоминает байесовские сети);

2) весовые коэффициенты имеют хорошо теоретически обоснованную содержательную интерпретацию, основанную на теории информации;

3) нейросеть является нелокальной, как сейчас говорят «полносвязной».

13.3.5.4.6. 3d-интегральные когнитивные карты

На рисунке 34 приведен фрагмент 3d-интегральной когнитивной карты в СК-модели IIN5.

3d-интегральная когнитивная карта является отображением на одном рисунке когнитивных диаграмм классов и значений факторов вверху и внизу соответственно (представлены на рисунках 24 и 28) и одного слоя нейронной сети (приведен на рисунке 33).

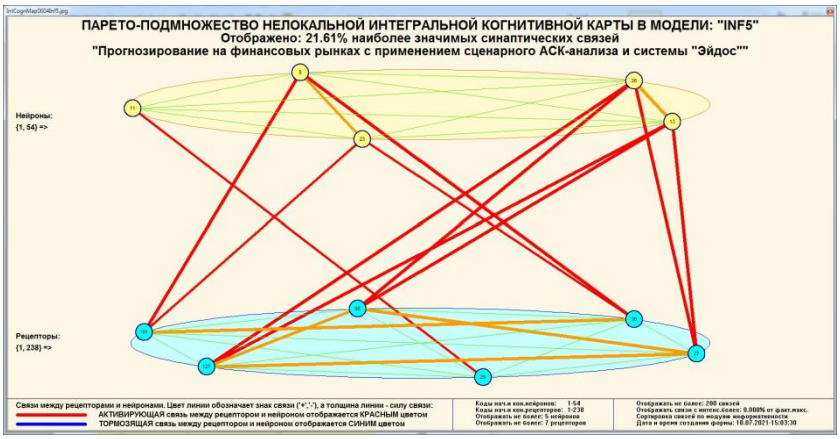


Рисунок 43. 3d-интегральная когнитивная карта в СК-модели INF5

13.3.5.4.7. Когнитивные функции

Вместо описания того, что представляют собой когнитивные функции, приведем help соответствующего режима системы «Эйдос» (рисунок 35) и сошлемся на работы, в которых описан этот подход [10][35].

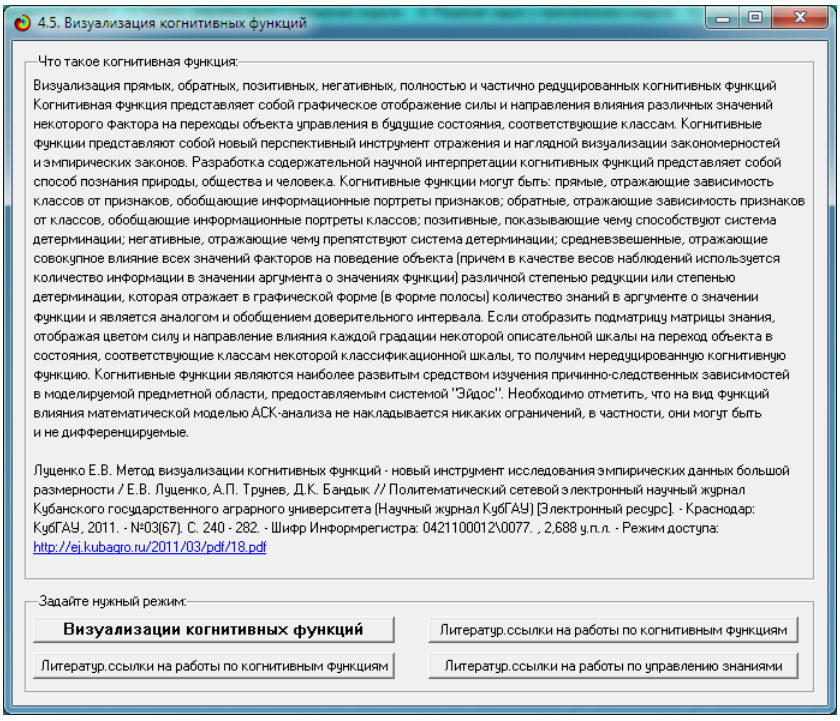


Рисунок 44. Help режима визуализации когнитивных функций

Когнитивная функция представляет собой графическое отображение силы и направления влияния различных значений некоторого фактора (признаков) на переходы объекта управления в будущие состояния,

[35] Подборка публикаций проф.Е.В.Луценко & С° по когнитивным функциям: http://lc.kubagro.ru/aidos/Works_on_cognitive_functions.htm.

соответствующие классам. Классы являются градациями классификационных шкал.

Когнитивные функции представляют собой новый перспективный инструмент отражения и наглядной визуализации эмпирических закономерностей и эмпирических законов. Разработка содержательной научной интерпретации когнитивных функций представляет собой способ познания природы, общества и человека [10][36].

Когнитивные функции могут быть: прямые, отражающие зависимость классов от признаков, обобщающие информационные портреты признаков; обратные, отражающие зависимость признаков от классов, обобщающие информационные портреты классов; позитивные, показывающие чему способствуют система детерминации (обозначены белой линией); негативные, отражающие чему препятствуют система детерминации (обозначены черной линией); средневзвешенные, отражающие совокупное влияние всех значений факторов на поведение объекта (причем в качестве весов наблюдений используется количество информации в значении аргумента о значениях функции) различной степенью редукции или степенью детерминации, которая отражает в графической форме (в форме полосы разной толщины) количество знаний в аргументе о значении функции и является аналогом и обобщением доверительного интервала.

Если отобразить подматрицу матрицы знания, отображая цветом силу и направление влияния каждой градации некоторой описательной шкалы на переход объекта в состояния, соответствующие классам некоторой классификационной шкалы, то получим нередуцированную когнитивную функцию.

Когнитивные функции являются наиболее развитым средством изучения причинно-следственных зависимостей в моделируемой предметной области, предоставляемым системой "Эйдос".

Необходимо отметить, что ***на вид функций влияния математической моделью АСК-анализа не накладывается никаких ограничений***, в частности, они могут быть и не дифференцируемые.

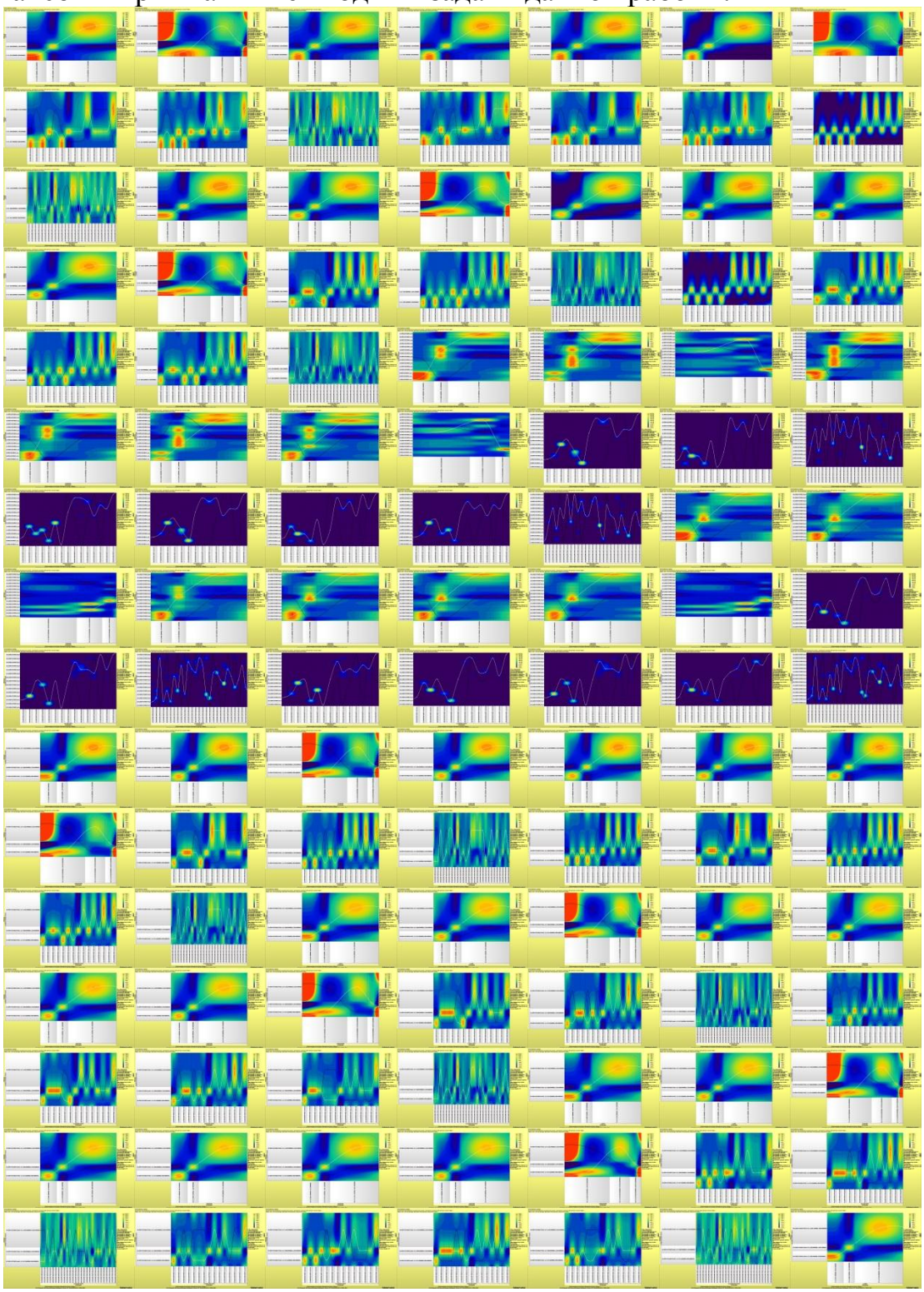
На рисунках 36 приведены когнитивные функции, наглядно отражающие силу и направление влияния значений (т.е. степени выраженности) различных характеристик ссудополучателей на риск невозврат полученных ими ссуд (класс).

Из когнитивных функций, приведенных на рисунке 36, хорошо видно, что ***зависимости между характеристиками финансового рынка***

[36] Работы проф.Е.В.Луценко & С° по выявлению, представлению и использованию знаний, логике и методологии научного познания: http://lc.kubagro.ru/aidos/Works_on_identification_presentation_and_use_of_knowledge.htm

и курсами акций компании Гугл и их динамикой имеют ярко выраженный и вполне очевидный и предсказуемый характер.

Но есть и несколько интересных неожиданных моментов, требующих специальной содержательной интерпретации. Эта содержательная интерпретация является делом специалистов по финансовым рынкам и не входит в задачи данной работы.



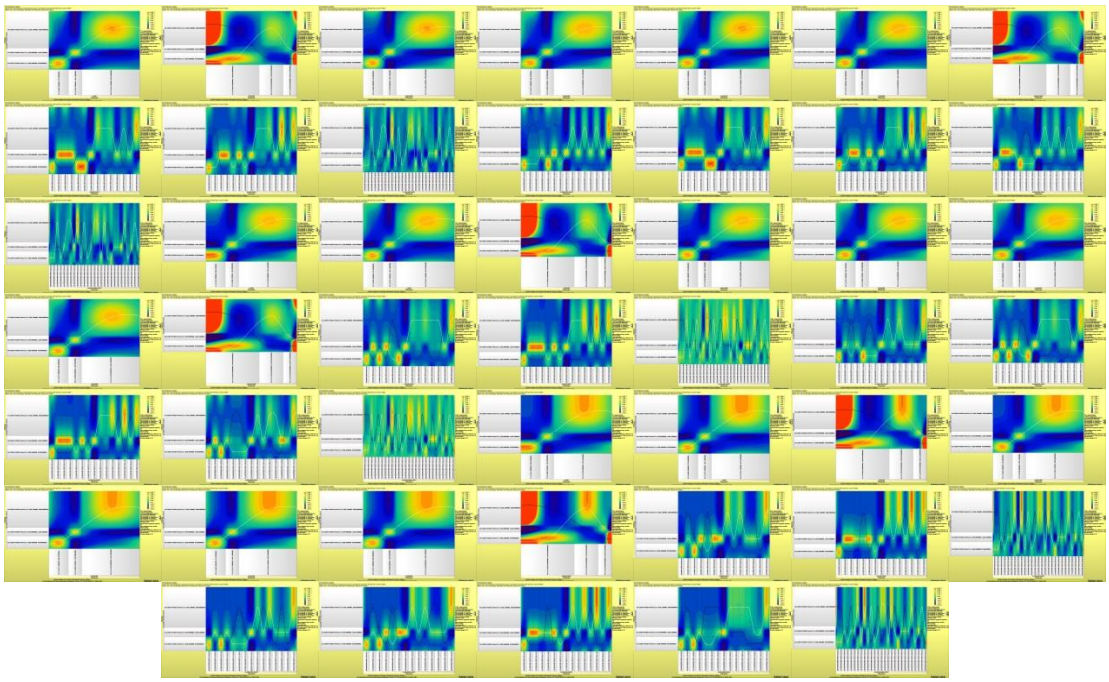


Рисунок 45. Примеры некоторых когнитивных функций в СК-модели INF5, отражающих силу и направление влияния значений характеристик финансового рынка на курсы акций компании Гугл[37]

13.3.5.4.8. Сила и направление влияния значений факторов на принадлежность к классам

На рисунках 12, 13, 14, 15 приведены некоторые статистические и системно-когнитивные модели, отражающие моделируемую предметную область.

Строки матриц моделей соответствуют значениям факторов, т.е. значениям характеристик ссудополучателей (градации описательных шкал).

Колонки матриц моделей соответствуют различным классам, отражающим риск невозврата ссуды (градации классификационных шкал).

Числовые значения в ячейках матриц моделей, находящихся на пересечении строк и колонок, отражают направление (знак) и силу влияния конкретной характеристики, соответствующей сроке, на конкретное значение класса – риска невозврата ссуды для ссудополучателем с такой характеристикой.

Если какая-то характеристика слабо влияет на класс риск невозврата ссуды, то в соответствующей строке матрицы модели будут малые по модулю значения разных знаков, если же влияние сильное – то и значения будут большие по модулю разных знаков.

[37] Не смотря на малый размер рисунков в работе они вполне читабельны при просмотре текста работы в увеличенном масштабе, например при масштабе 200% или 500%.

Если какая-либо характеристика способствует определенному риску невозврат ссуды, то в соответствующей этому результату ячейке матрицы модели будут положительные значения, если же понижает – то и значения будут отрицательные.

Из этого следует, что суммарную силу влияния той или иной характеристики ссудополучателя на класс (т.е. ценность данного значения характеристики для решения задачи прогнозирования риска невозврата ссуды и других задач) можно количественно оценивать ***степенью вариабельности значений*** в строке матрицы модели, соответствующей этой характеристике.

Существует много мер вариабельности значений: это и среднее модулей отклонения от среднего, и дисперсия, и среднеквадратичное отклонение и другие. В АСК-анализе и системе «Эйдос» для этой цели принято использовать среднеквадратичное отклонение. Численно оно равно стандартному отклонению и вычисляется по той же формуле, но мы предпочитаем не использовать термин «стандартное отклонение», т.к. он предполагает нормальность распределения исследуемых последовательностей чисел, а значит и проверку соответствующих статистических гипотез.

Самая правая колонка в матрицах моделей на рисунках 12-15 содержит количественную оценку вариабельности значений строки модели (среднеквадратичное отклонение), которая и представляет собой ценность характеристики, соответствующего строке, для решения задачи прогнозирования риска невозврата ссуды и других задач, рассмотренных в данной работе.

Если рассортировать матрицу модели по этой самой правой колонке в порядке убывания, а потом просуммировать значения в ней нарастающим итогом, то получим логистическую Парето-кривую, отражающую зависимость ценности модели от числа наиболее ценных признаков в ней (рисунок 37, таблица 8).

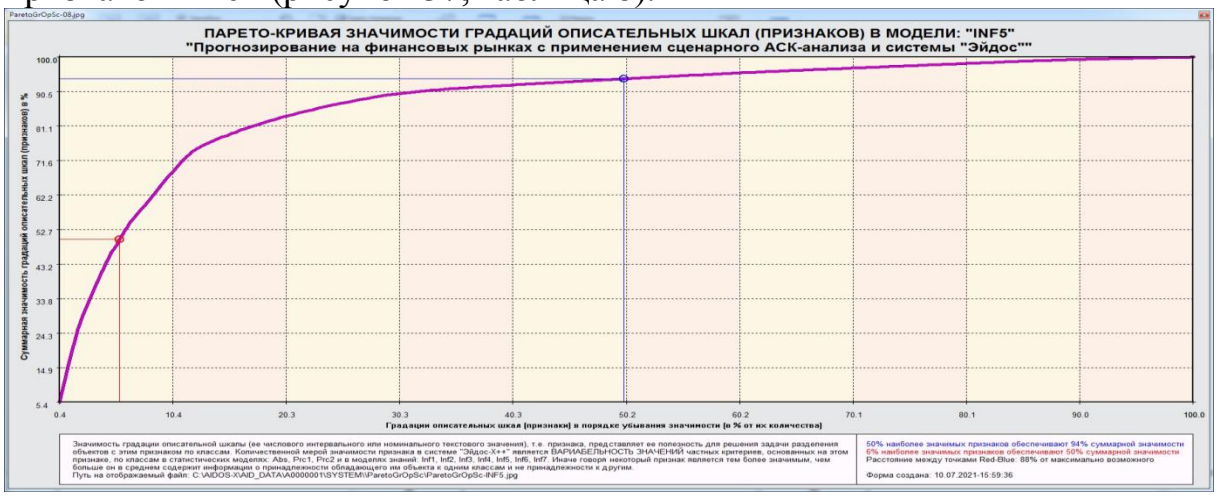


Рисунок 46. Парето-кривая значимости градаций описательных шкал

Таблица 19 – Парето-таблица значимости градаций описательных шкал, т.е. сила влияния значений характеристик финансового рынка на курсы акций компании Гугл и их динамику в СК-модели INF5

№	№%	Код значе-ния фактора	Наименование	Код фактора	Значи-мость в %	Значи-мость нараста-ющим итогом
1	0,420	27	HIGH-PAST3-HIGH-PAST3-01,02,01	9	**5,397**	5,397
2	0,840	98	ADJHIGH-PAST3-ADJHIGH-PAST3-13,14,13	13	5,397	10,794
3	1,261	127	ADJOPEN-PAST3-ADJOPEN-PAST3-19,20,19	15	5,397	16,191
4	1,681	30	HIGH-PAST3-HIGH-PAST3-02,02,01	9	4,770	20,962
5	2,101	101	ADJHIGH-PAST3-ADJHIGH-PAST3-14,14,13	13	4,770	25,732
6	2,521	29	HIGH-PAST3-HIGH-PAST3-02,01,01	9	3,476	29,208
7	2,941	100	ADJHIGH-PAST3-ADJHIGH-PAST3-14,13,13	13	3,476	32,685
8	3,361	130	ADJOPEN-PAST3-ADJOPEN-PAST3-20,19,20	15	2,887	35,572
9	3,782	41	LOW-PAST3-LOW-PAST3-04,05,04	10	2,887	38,459
10	4,202	112	ADJLOW-PAST3-ADJLOW-PAST3-16,17,16	14	2,887	41,345
11	4,622	86	ADJCLOSE-PAST3-ADJCLOSE-PAST3-11,10,11	12	2,691	44,036
12	5,042	83	ADJCLOSE-PAST3-ADJCLOSE-PAST3-10,11,10	12	2,378	46,414
13	5,462	40	LOW-PAST3-LOW-PAST3-04,04,05	10	2,206	48,620
14	5,882	111	ADJLOW-PAST3-ADJLOW-PAST3-16,16,17	14	2,206	50,826
15	6,303	129	ADJOPEN-PAST3-ADJOPEN-PAST3-20,19,19	15	2,014	52,840
16	6,723	137	ADJOPEN-PAST3-ADJOPEN-PAST3-21,20,21	15	1,969	54,809
17	7,143	82	ADJCLOSE-PAST3-ADJCLOSE-PAST3-10,10,11	12	1,594	56,403
18	7,563	43	LOW-PAST3-LOW-PAST3-05,04,04	10	1,581	57,984
19	7,983	114	ADJLOW-PAST3-ADJLOW-PAST3-17,16,16	14	1,581	59,565
20	8,403	131	ADJOPEN-PAST3-ADJOPEN-PAST3-20,20,19	15	1,581	61,146
21	8,824	26	HIGH-PAST3-HIGH-PAST3-01,01,02	9	1,580	62,725
22	9,244	97	ADJHIGH-PAST3-ADJHIGH-PAST3-13,13,14	13	1,580	64,305
23	9,664	126	ADJOPEN-PAST3-ADJOPEN-PAST3-19,19,20	15	1,580	65,885
24	10,084	85	ADJCLOSE-PAST3-ADJCLOSE-PAST3-11,10,10	12	1,580	67,464
25	10,504	33	HIGH-PAST3-HIGH-PAST3-02,03,02	9	1,479	68,943
26	10,924	104	ADJHIGH-PAST3-ADJHIGH-PAST3-14,15,14	13	1,479	70,423
27	11,345	84	ADJCLOSE-PAST3-ADJCLOSE-PAST3-10,11,11	12	1,438	71,861
28	11,765	42	LOW-PAST3-LOW-PAST3-04,05,05	10	1,181	73,042
29	12,185	113	ADJLOW-PAST3-ADJLOW-PAST3-16,17,17	14	1,181	74,223
30	12,605	48	LOW-PAST3-LOW-PAST3-05,06,05	10	0,765	74,988
31	13,025	119	ADJLOW-PAST3-ADJLOW-PAST3-17,18,17	14	0,765	75,753
32	13,445	49	LOW-PAST3-LOW-PAST3-05,06,06	10	0,555	76,308
33	13,866	120	ADJLOW-PAST3-ADJLOW-PAST3-17,18,18	14	0,555	76,863
34	14,286	134	ADJOPEN-PAST3-ADJOPEN-PAST3-20,21,20	15	0,554	77,417
35	14,706	35	HIGH-PAST3-HIGH-PAST3-03,02,02	9	0,547	77,964
36	15,126	106	ADJHIGH-PAST3-ADJHIGH-PAST3-15,14,14	13	0,547	78,511
37	15,546	133	ADJOPEN-PAST3-ADJOPEN-PAST3-20,20,21	15	0,491	79,002
38	15,966	91	ADJCLOSE-PAST3-ADJCLOSE-PAST3-11,12,12	12	0,487	79,489
39	16,387	90	ADJCLOSE-PAST3-ADJCLOSE-PAST3-11,12,11	12	0,465	79,954
40	16,807	47	LOW-PAST3-LOW-PAST3-05,05,06	10	0,441	80,394
41	17,227	118	ADJLOW-PAST3-ADJLOW-PAST3-17,17,18	14	0,441	80,835
42	17,647	69	VOLUME-PAST3-VOLUME-PAST3-08,09,07	11	0,422	81,257
43	18,067	155	ADJVOLUME-PAST3-ADJVOLUME-PAST3-23,24,22	16	0,422	81,679
44	18,487	32	HIGH-PAST3-HIGH-PAST3-02,02,03	9	0,416	82,095
45	18,908	103	ADJHIGH-PAST3-ADJHIGH-PAST3-14,14,15	13	0,416	82,511
46	19,328	93	ADJCLOSE-PAST3-ADJCLOSE-PAST3-12,11,12	12	0,406	82,917
47	19,748	92	ADJCLOSE-PAST3-ADJCLOSE-PAST3-12,11,11	12	0,391	83,308
48	20,168	36	HIGH-PAST3-HIGH-PAST3-03,02,03	9	0,340	83,648
49	20,588	107	ADJHIGH-PAST3-ADJHIGH-PAST3-15,14,15	13	0,340	83,988
50	21,008	37	HIGH-PAST3-HIGH-PAST3-03,03,02	9	0,338	84,326
51	21,429	108	ADJHIGH-PAST3-ADJHIGH-PAST3-15,15,14	13	0,338	84,664
52	21,849	89	ADJCLOSE-PAST3-ADJCLOSE-PAST3-11,11,12	12	0,316	84,980
53	22,269	136	ADJOPEN-PAST3-ADJOPEN-PAST3-21,20,20	15	0,315	85,295
54	22,689	74	VOLUME-PAST3-VOLUME-PAST3-09,07,09	11	0,313	85,608
55	23,109	160	ADJVOLUME-PAST3-ADJVOLUME-PAST3-24,22,24	16	0,313	85,921
56	23,529	77	VOLUME-PAST3-VOLUME-PAST3-09,08,09	11	0,294	86,215
57	23,950	163	ADJVOLUME-PAST3-ADJVOLUME-PAST3-24,23,24	16	0,294	86,509
58	24,370	135	ADJOPEN-PAST3-ADJOPEN-PAST3-20,21,21	15	0,284	86,793
59	24,790	34	HIGH-PAST3-HIGH-PAST3-02,03,03	9	0,274	87,067
60	25,210	105	ADJHIGH-PAST3-ADJHIGH-PAST3-14,15,15	13	0,274	87,342
61	25,630	138	ADJOPEN-PAST3-ADJOPEN-PAST3-21,21,20	15	0,271	87,613
62	26,050	50	LOW-PAST3-LOW-PAST3-06,05,05	10	0,259	87,872
63	26,471	121	ADJLOW-PAST3-ADJLOW-PAST3-18,17,17	14	0,259	88,130
64	26,891	94	ADJCLOSE-PAST3-ADJCLOSE-PAST3-12,12,11	12	0,247	88,377
65	27,311	61	VOLUME-PAST3-VOLUME-PAST3-07,09,08	11	0,238	88,615
66	27,731	147	ADJVOLUME-PAST3-ADJVOLUME-PAST3-22,24,23	16	0,238	88,853
67	28,151	52	LOW-PAST3-LOW-PAST3-06,06,05	10	0,222	89,075
68	28,571	123	ADJLOW-PAST3-ADJLOW-PAST3-18,18,17	14	0,222	89,296
69	28,992	70	VOLUME-PAST3-VOLUME-PAST3-08,09,08	11	0,204	89,500
70	29,412	156	ADJVOLUME-PAST3-ADJVOLUME-PAST3-23,24,23	16	0,204	89,704
71	29,832	72	VOLUME-PAST3-VOLUME-PAST3-09,07,07	11	0,152	89,856
72	30,252	158	ADJVOLUME-PAST3-ADJVOLUME-PAST3-24,22,22	16	0,152	90,009
73	30,672	64	VOLUME-PAST3-VOLUME-PAST3-08,07,08	11	0,141	90,150
74	31,092	150	ADJVOLUME-PAST3-ADJVOLUME-PAST3-23,22,23	16	0,141	90,291
75	31,513	57	VOLUME-PAST3-VOLUME-PAST3-07,08,07	11	0,137	90,428
76	31,933	143	ADJVOLUME-PAST3-ADJVOLUME-PAST3-22,23,22	16	0,137	90,565
77	32,353	55	VOLUME-PAST3-VOLUME-PAST3-07,07,08	11	0,125	90,690

78	32,773	141	ADJVOLUME-PAST3-ADJVOLUME-PAST3-22,22,23	16	0,125	90,814
79	33,193	59	VOLUME-PAST3-VOLUME-PAST3-07,08,09	11	0,107	90,921
80	33,613	145	ADJVOLUME-PAST3-ADJVOLUME-PAST3-22,23,24	16	0,107	91,028
81	34,034	75	VOLUME-PAST3-VOLUME-PAST3-09,08,07	11	0,107	91,135
82	34,454	161	ADJVOLUME-PAST3-ADJVOLUME-PAST3-24,23,22	16	0,107	91,242
83	34,874	76	VOLUME-PAST3-VOLUME-PAST3-09,08,08	11	0,092	91,334
84	35,294	162	ADJVOLUME-PAST3-ADJVOLUME-PAST3-24,23,23	16	0,092	91,425
85	35,714	67	VOLUME-PAST3-VOLUME-PAST3-08,08,08	11	0,086	91,511
86	36,134	153	ADJVOLUME-PAST3-ADJVOLUME-PAST3-23,23,23	16	0,086	91,596
87	36,555	5	LOW-2/3-{972.2500000, 1181.1200000}	2	0,081	91,677
88	36,975	17	ADJLOW-2/3-{972.2500000, 1181.1200000}	6	0,081	91,758
89	37,395	11	ADJCLOSE-2/3-{978.8900000, 1189.1300000}	4	0,081	91,839
90	37,815	201	ADJCLOSE-PAST3-POINT3-ADJCLOSE-PAST3-Point3-2/3-{978.8900000, 1189.1300000}	28	0,081	91,920
91	38,235	66	VOLUME-PAST3-VOLUME-PAST3-08,08,07	11	0,080	91,999
92	38,655	152	ADJVOLUME-PAST3-ADJVOLUME-PAST3-23,23,22	16	0,080	92,079
93	39,076	46	LOW-PAST3-LOW-PAST3-05,05,05	10	0,079	92,158
94	39,496	117	ADJLOW-PAST3-ADJLOW-PAST3-17,17,17	14	0,079	92,237
95	39,916	58	VOLUME-PAST3-VOLUME-PAST3-07,08,08	11	0,078	92,315
96	40,336	144	ADJVOLUME-PAST3-ADJVOLUME-PAST3-22,23,23	16	0,078	92,393
97	40,756	63	VOLUME-PAST3-VOLUME-PAST3-08,07,07	11	0,078	92,472
98	41,176	149	ADJVOLUME-PAST3-ADJVOLUME-PAST3-23,22,22	16	0,078	92,550
99	41,597	2	HIGH-2/3-{986.9100000, 1199.0000000}	1	0,078	92,628
100	42,017	14	ADJHIGH-2/3-{986.9100000, 1199.0000000}	5	0,078	92,705
101	42,437	174	HIGH-PAST3-POINT3-HIGH-PAST3-Point3-2/3-{986.9100000, 1199.0000000}	19	0,078	92,783
102	42,857	198	ADJCLOSE-PAST3-POINT2-ADJCLOSE-PAST3-Point2-2/3-{978.8900000, 1189.1300000}	27	0,076	92,860
103	43,277	171	HIGH-PAST3-POINT2-HIGH-PAST3-Point2-2/3-{986.9100000, 1199.0000000}	18	0,076	92,935
104	43,697	31	HIGH-PAST3-HIGH-PAST3-02,02,02	9	0,076	93,011
105	44,118	102	ADJHIGH-PAST3-ADJHIGH-PAST3-14,14,14	13	0,076	93,087
106	44,538	20	ADJOPEN-2/3-{980.0000000, 1190.9600000}	7	0,076	93,163
107	44,958	132	ADJOPEN-PAST3-ADJOPEN-PAST3-20,20,20	15	0,073	93,236
108	45,378	88	ADJCLOSE-PAST3-ADJCLOSE-PAST3-11,11,11	12	0,073	93,309
109	45,798	168	HIGH-PAST3-POINT1-HIGH-PAST3-Point1-2/3-{986.9100000, 1199.0000000}	17	0,072	93,382
110	46,218	195	ADJCLOSE-PAST3-POINT1-ADJCLOSE-PAST3-Point1-2/3-{978.8900000, 1189.1300000}	26	0,071	93,453
111	46,639	197	ADJCLOSE-PAST3-POINT2-ADJCLOSE-PAST3-Point2-1/3-{668.2600000, 978.8900000}	27	0,071	93,524
112	47,059	173	HIGH-PAST3-POINT3-HIGH-PAST3-Point3-1/3-{672.3000000, 986.9100000}	19	0,070	93,594
113	47,479	200	ADJCLOSE-PAST3-POINT3-ADJCLOSE-PAST3-Point3-1/3-{668.2600000, 978.8900000}	28	0,070	93,664
114	47,899	1	HIGH-1/3-{672.3000000, 986.9100000}	1	0,070	93,735
115	48,319	13	ADJHIGH-1/3-{672.3000000, 986.9100000}	5	0,070	93,805
116	48,739	10	ADJCLOSE-1/3-{668.2600000, 978.8900000}	4	0,070	93,875
117	49,160	19	ADJOPEN-1/3-{671.0000000, 980.0000000}	7	0,070	93,945
118	49,580	4	LOW-1/3-{663.2840000, 972.2500000}	2	0,070	94,015
119	50,000	16	ADJLOW-1/3-{663.2840000, 972.2500000}	6	0,070	94,084
120	50,420	170	HIGH-PAST3-POINT2-HIGH-PAST3-Point2-1/3-{672.3000000, 986.9100000}	18	0,069	94,153
121	50,840	39	LOW-PAST3-LOW-PAST3-04,04,04	10	0,068	94,221
122	51,261	110	ADJLOW-PAST3-ADJLOW-PAST3-16,16,16	14	0,068	94,290
123	51,681	25	HIGH-PAST3-HIGH-PAST3-01,01,01	9	0,068	94,358
124	52,101	96	ADJHIGH-PAST3-ADJHIGH-PAST3-13,13,13	13	0,068	94,426
125	52,521	194	ADJCLOSE-PAST3-POINT1-ADJCLOSE-PAST3-Point1-1/3-{668.2600000, 978.8900000}	26	0,068	94,494
126	52,941	125	ADJOPEN-PAST3-ADJOPEN-PAST3-19,19,19	15	0,068	94,561
127	53,361	167	HIGH-PAST3-POINT1-HIGH-PAST3-Point1-1/3-{672.3000000, 986.9100000}	17	0,068	94,629
128	53,782	81	ADJCLOSE-PAST3-ADJCLOSE-PAST3-10,10,10	12	0,067	94,696
129	54,202	6	LOW-3/3-{1181.1200000, 2078.5400000}	2	0,067	94,763
130	54,622	18	ADJLOW-3/3-{1181.1200000, 2078.5400000}	6	0,067	94,829
131	55,042	12	ADJCLOSE-3/3-{1189.1300000, 2098.0000000}	4	0,066	94,896
132	55,462	202	ADJCLOSE-PAST3-POINT3-ADJCLOSE-PAST3-Point3-3/3-{1189.1300000, 2098.0000000}	28	0,066	94,962
133	55,882	53	LOW-PAST3-LOW-PAST3-06,06,06	10	0,066	95,028
134	56,303	124	ADJLOW-PAST3-ADJLOW-PAST3-18,18,18	14	0,066	95,094
135	56,723	203	ADJHIGH-PAST3-POINT1-ADJHIGH-PAST3-Point1-1/3-{672.3000000, 986.9100000}	29	0,065	95,159
136	57,143	3	HIGH-3/3-{1199.0000000, 2123.5469000}	1	0,065	95,224
137	57,563	15	ADJHIGH-3/3-{1199.0000000, 2123.5469000}	5	0,065	95,289
138	57,983	175	HIGH-PAST3-POINT3-HIGH-PAST3-Point3-3/3-{1199.0000000, 2123.5469000}	19	0,065	95,354
139	58,403	176	LOW-PAST3-POINT1-LOW-PAST3-Point1-1/3-{663.2840000, 972.2500000}	20	0,064	95,418
140	58,824	95	ADJCLOSE-PAST3-ADJCLOSE-PAST3-12,12,12	12	0,064	95,482
141	59,244	172	HIGH-PAST3-POINT2-HIGH-PAST3-Point2-3/3-{1199.0000000, 2123.5469000}	18	0,064	95,546
142	59,664	38	HIGH-PAST3-HIGH-PAST3-03,03,03	9	0,064	95,610
143	60,084	109	ADJHIGH-PAST3-ADJHIGH-PAST3-15,15,15	13	0,064	95,674
144	60,504	21	ADJOPEN-3/3-{1190.9600000, 2105.9100000}	7	0,064	95,738
145	60,924	139	ADJOPEN-PAST3-ADJOPEN-PAST3-21,21,21	15	0,064	95,802
146	61,345	199	ADJCLOSE-PAST3-POINT2-ADJCLOSE-PAST3-Point2-3/3-{1189.1300000, 2098.0000000}	27	0,064	95,865
147	61,765	169	HIGH-PAST3-POINT1-HIGH-PAST3-Point1-3/3-{1199.0000000, 2123.5469000}	17	0,059	95,924
148	62,185	196	ADJCLOSE-PAST3-POINT1-ADJCLOSE-PAST3-Point1-3/3-{1189.1300000, 2098.0000000}	26	0,059	95,983
149	62,605	56	VOLUME-PAST3-VOLUME-PAST3-07,07,09	11	0,058	96,041
150	63,025	142	ADJVOLUME-PAST3-ADJVOLUME-PAST3-22,22,24	16	0,058	96,099
151	63,445	65	VOLUME-PAST3-VOLUME-PAST3-08,07,09	11	0,056	96,155
152	63,866	151	ADJVOLUME-PAST3-ADJVOLUME-PAST3-23,22,24	16	0,056	96,211
153	64,286	28	HIGH-PAST3-HIGH-PAST3-01,02,02	9	0,055	96,266
154	64,706	44	LOW-PAST3-LOW-PAST3-05,04,05	10	0,055	96,321
155	65,126	99	ADJHIGH-PAST3-ADJHIGH-PAST3-13,14,14	13	0,055	96,376
156	65,546	115	ADJLOW-PAST3-ADJLOW-PAST3-17,16,17	14	0,055	96,431
157	65,966	179	LOW-PAST3-POINT2-LOW-PAST3-Point2-1/3-{663.2840000, 972.2500000}	21	0,054	96,484
158	66,387	180	LOW-PAST3-POINT2-LOW-PAST3-Point2-2/3-{972.2500000, 1181.1200000}	21	0,054	96,538
159	66,807	181	LOW-PAST3-POINT2-LOW-PAST3-Point2-3/3-{1181.1200000, 2078.5400000}	21	0,054	96,592
160	67,227	182	LOW-PAST3-POINT3-LOW-PAST3-Point3-1/3-{663.2840000, 972.2500000}	22	0,054	96,645
161	67,647	183	LOW-PAST3-POINT3-LOW-PAST3-Point3-2/3-{972.2500000, 1181.1200000}	22	0,054	96,699
162	68,067	184	LOW-PAST3-POINT3-LOW-PAST3-Point3-3/3-{1181.1200000, 2078.5400000}	22	0,054	96,752
163	68,487	185	VOLUME-PAST3-POINT1-VOLUME-PAST3-Point1-1/3-{346753.0000000, 1267649.0000000}	23	0,054	96,806
164	68,908	186	VOLUME-PAST3-POINT1-VOLUME-PAST3-Point1-2/3-{1267649.0000000, 1679384.0000000}	23	0,054	96,860
165	69,328	187	VOLUME-PAST3-POINT1-VOLUME-PAST3-Point1-3/3-{1679384.0000000, 6207027.0000000}	23	0,054	96,913
166	69,748	188	VOLUME-PAST3-POINT2-VOLUME-PAST3-Point2-1/3-{346753.0000000, 1267649.0000000}	24	0,054	96,967

167	70,168	189	VOLUME-PAST3-POINT2-VOLUME-PAST3-Point2-2/3-{1267649.0000000, 1679384.0000000}	24	0,054	97,021
168	70,588	190	VOLUME-PAST3-POINT2-VOLUME-PAST3-Point2-3/3-{1679384.0000000, 6207027.0000000}	24	0,054	97,074
169	71,008	191	VOLUME-PAST3-POINT3-VOLUME-PAST3-Point3-1/3-{346753.0000000, 1267649.0000000}	25	0,054	97,128
170	71,429	192	VOLUME-PAST3-POINT3-VOLUME-PAST3-Point3-2/3-{1267649.0000000, 1679384.0000000}	25	0,054	97,182
171	71,849	193	VOLUME-PAST3-POINT3-VOLUME-PAST3-Point3-3/3-{1679384.0000000, 6207027.0000000}	25	0,054	97,235
172	72,269	206	ADJHIGH-PAST3-POINT2-ADJHIGH-PAST3-Point2-1/3-{672.3000000, 986.9100000}	30	0,054	97,289
173	72,689	207	ADJHIGH-PAST3-POINT2-ADJHIGH-PAST3-Point2-2/3-{986.9100000, 1199.0000000}	30	0,054	97,342
174	73,109	208	ADJHIGH-PAST3-POINT2-ADJHIGH-PAST3-Point2-3/3-{1199.0000000, 2123.5469000}	30	0,054	97,396
175	73,529	209	ADJHIGH-PAST3-POINT3-ADJHIGH-PAST3-Point3-1/3-{672.3000000, 986.9100000}	31	0,054	97,450
176	73,950	210	ADJHIGH-PAST3-POINT3-ADJHIGH-PAST3-Point3-2/3-{986.9100000, 1199.0000000}	31	0,054	97,503
177	74,370	211	ADJHIGH-PAST3-POINT3-ADJHIGH-PAST3-Point3-3/3-{1199.0000000, 2123.5469000}	31	0,054	97,557
178	74,790	212	ADJLOW-PAST3-POINT1-ADJLOW-PAST3-Point1-1/3-{663.2840000, 972.2500000}	32	0,054	97,611
179	75,210	213	ADJLOW-PAST3-POINT1-ADJLOW-PAST3-Point1-2/3-{972.2500000, 1181.1200000}	32	0,054	97,664
180	75,630	214	ADJLOW-PAST3-POINT1-ADJLOW-PAST3-Point1-3/3-{1181.1200000, 2078.5400000}	32	0,054	97,718
181	76,050	215	ADJLOW-PAST3-POINT2-ADJLOW-PAST3-Point2-1/3-{663.2840000, 972.2500000}	33	0,054	97,771
182	76,471	216	ADJLOW-PAST3-POINT2-ADJLOW-PAST3-Point2-2/3-{972.2500000, 1181.1200000}	33	0,054	97,825
183	76,891	217	ADJLOW-PAST3-POINT2-ADJLOW-PAST3-Point2-3/3-{1181.1200000, 2078.5400000}	33	0,054	97,879
184	77,311	218	ADJLOW-PAST3-POINT3-ADJLOW-PAST3-Point3-1/3-{663.2840000, 972.2500000}	34	0,054	97,932
185	77,731	219	ADJLOW-PAST3-POINT3-ADJLOW-PAST3-Point3-2/3-{972.2500000, 1181.1200000}	34	0,054	97,986
186	78,151	220	ADJLOW-PAST3-POINT3-ADJLOW-PAST3-Point3-3/3-{1181.1200000, 2078.5400000}	34	0,054	98,040
187	78,571	221	ADJOPEN-PAST3-POINT1-ADJOPEN-PAST3-Point1-1/3-{671.0000000, 980.0000000}	35	0,054	98,093
188	78,992	222	ADJOPEN-PAST3-POINT1-ADJOPEN-PAST3-Point1-2/3-{980.0000000, 1190.9600000}	35	0,054	98,147
189	79,412	223	ADJOPEN-PAST3-POINT1-ADJOPEN-PAST3-Point1-3/3-{1190.9600000, 2105.9100000}	35	0,054	98,201
190	79,832	224	ADJOPEN-PAST3-POINT2-ADJOPEN-PAST3-Point2-1/3-{671.0000000, 980.0000000}	36	0,054	98,254
191	80,252	225	ADJOPEN-PAST3-POINT2-ADJOPEN-PAST3-Point2-2/3-{980.0000000, 1190.9600000}	36	0,054	98,308
192	80,672	226	ADJOPEN-PAST3-POINT2-ADJOPEN-PAST3-Point2-3/3-{1190.9600000, 2105.9100000}	36	0,054	98,361
193	81,092	227	ADJOPEN-PAST3-POINT3-ADJOPEN-PAST3-Point3-1/3-{671.0000000, 980.0000000}	37	0,054	98,415
194	81,513	228	ADJOPEN-PAST3-POINT3-ADJOPEN-PAST3-Point3-2/3-{980.0000000, 1190.9600000}	37	0,054	98,469
195	81,933	229	ADJOPEN-PAST3-POINT3-ADJOPEN-PAST3-Point3-3/3-{1190.9600000, 2105.9100000}	37	0,054	98,522
196	82,353	51	LOW-PAST3-LOW-PAST3-06,05,06	10	0,052	98,575
197	82,773	122	ADJLOW-PAST3-ADJLOW-PAST3-18,17,18	14	0,052	98,627
198	83,193	60	VOLUME-PAST3-VOLUME-PAST3-07,09,07	11	0,052	98,679
199	83,613	146	ADJVOLUME-PAST3-ADJVOLUME-PAST3-22,24,22	16	0,052	98,730
200	84,034	68	VOLUME-PAST3-VOLUME-PAST3-08,08,09	11	0,049	98,779
201	84,454	154	ADJVOLUME-PAST3-ADJVOLUME-PAST3-23,23,24	16	0,049	98,828
202	84,874	80	VOLUME-PAST3-VOLUME-PAST3-09,09,09	11	0,047	98,875
203	85,294	166	ADJVOLUME-PAST3-ADJVOLUME-PAST3-24,24,24	16	0,047	98,922
204	85,714	79	VOLUME-PAST3-VOLUME-PAST3-09,09,08	11	0,047	98,968
205	86,134	165	ADJVOLUME-PAST3-ADJVOLUME-PAST3-24,24,23	16	0,047	99,015
206	86,555	62	VOLUME-PAST3-VOLUME-PAST3-07,09,09	11	0,043	99,057
207	86,975	148	ADJVOLUME-PAST3-ADJVOLUME-PAST3-22,24,24	16	0,043	99,100
208	87,395	54	VOLUME-PAST3-VOLUME-PAST3-07,07,07	11	0,040	99,140
209	87,815	140	ADJVOLUME-PAST3-ADJVOLUME-PAST3-22,22,22	16	0,040	99,181
210	88,235	128	ADJOPEN-PAST3-ADJOPEN-PAST3-19,20,20	15	0,039	99,220
211	88,655	177	LOW-PAST3-POINT1-LOW-PAST3-Point1-2/3-{972.2500000, 1181.1200000}	20	0,037	99,257
212	89,076	178	LOW-PAST3-POINT1-LOW-PAST3-Point1-3/3-{1181.1200000, 2078.5400000}	20	0,036	99,292
213	89,496	204	ADJHIGH-PAST3-POINT1-ADJHIGH-PAST3-Point1-2/3-{986.9100000, 1199.0000000}	29	0,036	99,328
214	89,916	8	VOLUME-2/3-{1267649.0000000, 1679384.0000000}	3	0,035	99,363
215	90,336	23	ADJVOLUME-2/3-{1267649.0000000, 1679384.0000000}	8	0,035	99,398
216	90,756	237	ADJVOLUME-PAST3-POINT3-ADJVOLUME-PAST3-Point3-2/3-{1267649.0000000, 1679384.0000000}	40	0,035	99,434
217	91,176	205	ADJHIGH-PAST3-POINT1-ADJHIGH-PAST3-Point1-3/3-{1199.0000000, 2123.5469000}	29	0,035	99,468
218	91,597	238	ADJVOLUME-PAST3-POINT3-ADJVOLUME-PAST3-Point3-3/3-{1679384.0000000, 6207027.0000000}	40	0,031	99,499
219	92,017	45	LOW-PAST3-LOW-PAST3-05,05,04	10	0,031	99,530
220	92,437	87	ADJCLOSE-PAST3-ADJCLOSE-PAST3-11,11,10	12	0,031	99,560
221	92,857	116	ADJLOW-PAST3-ADJLOW-PAST3-17,17,16	14	0,031	99,591
222	93,277	230	ADJVOLUME-PAST3-POINT1-ADJVOLUME-PAST3-Point1-1/3-{346753.0000000, 1267649.0000000}	38	0,031	99,622
223	93,697	9	VOLUME-3/3-{1679384.0000000, 6207027.0000000}	3	0,031	99,652
224	94,118	24	ADJVOLUME-3/3-{1679384.0000000, 6207027.0000000}	8	0,031	99,683
225	94,538	232	ADJVOLUME-PAST3-POINT1-ADJVOLUME-PAST3-Point1-3/3-{1679384.0000000, 6207027.0000000}	38	0,030	99,713
226	94,958	71	VOLUME-PAST3-VOLUME-PAST3-08,09,09	11	0,029	99,743
227	95,378	157	ADJVOLUME-PAST3-ADJVOLUME-PAST3-23,24,24	16	0,029	99,772
228	95,798	234	ADJVOLUME-PAST3-POINT2-ADJVOLUME-PAST3-Point2-2/3-{1267649.0000000, 1679384.0000000}	39	0,028	99,800
229	96,218	235	ADJVOLUME-PAST3-POINT2-ADJVOLUME-PAST3-Point2-3/3-{1679384.0000000, 6207027.0000000}	39	0,026	99,825
230	96,639	7	VOLUME-1/3-{346753.0000000, 1267649.0000000}	3	0,025	99,851
231	97,059	22	ADJVOLUME-1/3-{346753.0000000, 1267649.0000000}	8	0,025	99,876
232	97,479	236	ADJVOLUME-PAST3-POINT3-ADJVOLUME-PAST3-Point3-1/3-{346753.0000000, 1267649.0000000}	40	0,025	99,902
233	97,899	231	ADJVOLUME-PAST3-POINT1-ADJVOLUME-PAST3-Point1-2/3-{1267649.0000000, 1679384.0000000}	38	0,024	99,925
234	98,319	233	ADJVOLUME-PAST3-POINT2-ADJVOLUME-PAST3-Point2-1/3-{346753.0000000, 1267649.0000000}	39	0,022	99,948
235	98,739	78	VOLUME-PAST3-VOLUME-PAST3-09,09,07	11	0,017	99,965
236	99,160	164	ADJVOLUME-PAST3-ADJVOLUME-PAST3-24,24,22	16	0,017	99,982
237	99,580	73	VOLUME-PAST3-VOLUME-PAST3-09,07,08	11	0,009	99,991
238	100,000	159	ADJVOLUME-PAST3-ADJVOLUME-PAST3-24,22,23	16	**0,009**	100,000

Из таблицы 8 видно, что сила влияния на курсы акций наиболее сильного значения фактора в **605** раз превосходит силу влияния наиболее слабого значения фактора, что, конечно, очень существенно.

Данные, приведенные на рисунке 37 и в таблице 8, находится в XLS-файлах, созданных в режиме 3.7.5. Информация об этом содержится в экранной форме на рисунке 38:

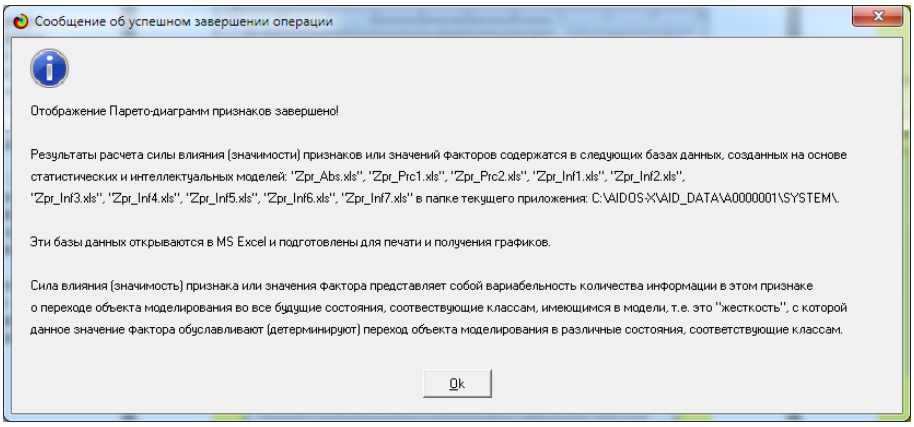


Рисунок 47. Информация о XLS-файлах

Из рисунка 37 и таблицы 8 видно, что 50% наиболее ценных для решения задачи прогнозирования значений характеристик финансового рынка обеспечивают 94% суммарного влияния всех характеристик на курсы акций компании Гугл и их динамику, а 50% суммарной ценности обеспечиваются всего 6% наиболее существенных значений характеристик финансового рынка.

Обращаем внимание на то, что наиболее ценной для решения задачи прогнозирования курсов акций компании Гугл и их динамики является HIGH-PAST3-HIGH-PAST3-01,02,01, а наименее ценным – ADJVOLUME-PAST3-ADJVOLUME-PAST3-24,22,23.

Из таблицы 8 видно, что наиболее сильное влияние на курсы акций компании Гугл и их динамику оказывают следующие 10 значений характеристик финансового рынка:

- HIGH-PAST3-HIGH-PAST3-01,02,01
- ADJHIGH-PAST3-ADJHIGH-PAST3-13,14,13
- ADJOPEN-PAST3-ADJOPEN-PAST3-19,20,19
- HIGH-PAST3-HIGH-PAST3-02,02,01
- ADJHIGH-PAST3-ADJHIGH-PAST3-14,14,13
- HIGH-PAST3-HIGH-PAST3-02,01,01
- ADJHIGH-PAST3-ADJHIGH-PAST3-14,13,13
- ADJOPEN-PAST3-ADJOPEN-PAST3-20,19,20
- LOW-PAST3-LOW-PAST3-04,05,04

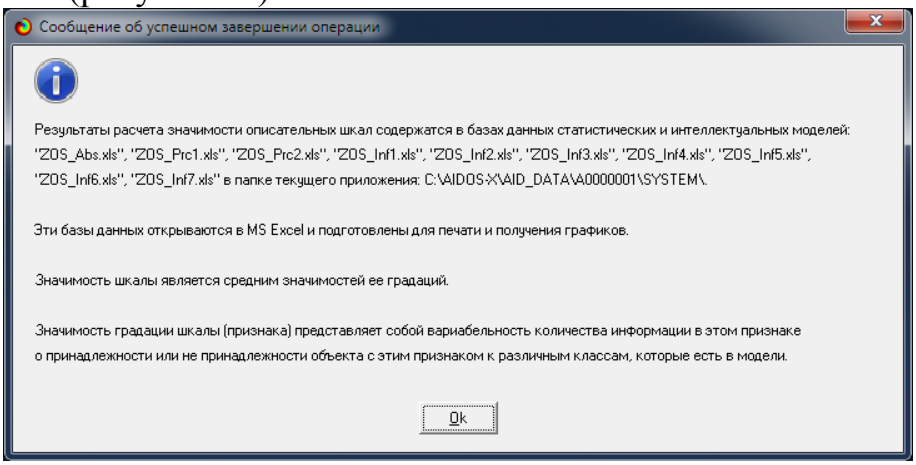
а наиболее слабое – следующие 10:

- ADJVOLUME-PAST3-POINT2-ADJVOLUME-PAST3-Point2-3/3-{1679384.0000000, 6207027.0000000}
- VOLUME-1/3-{346753.0000000, 1267649.0000000}
- ADJVOLUME-1/3-{346753.0000000, 1267649.0000000}
- ADJVOLUME-PAST3-POINT3-ADJVOLUME-PAST3-Point3-1/3-{346753.0000000, 1267649.0000000}
- ADJVOLUME-PAST3-POINT1-ADJVOLUME-PAST3-Point1-2/3-{1267649.0000000, 1679384.0000000}
- ADJVOLUME-PAST3-POINT2-ADJVOLUME-PAST3-Point2-1/3-{346753.0000000, 1267649.0000000}

- VOLUME-PAST3-VOLUME-PAST3-09,09,07
- ADJVOLUME-PAST3-ADJVOLUME-PAST3-24,24,22
- VOLUME-PAST3-VOLUME-PAST3-09,07,08
- ADJVOLUME-PAST3-ADJVOLUME-PAST3-24,22,23

При этом сила влияния наиболее и наименее значимых значений факторов на классы отличается, как мы уже упоминали, в **605** раз, что очень существенно.

Ценность же не значений характеристик финансового рынка, а характеристик в целом (всей описательной шкалы или фактора), для решения этих задач можно количественно оценивать как *среднее* от ценности значений этого параметра (таблица 9). Это можно сделать в режиме 3.7.4 (рисунок 39):



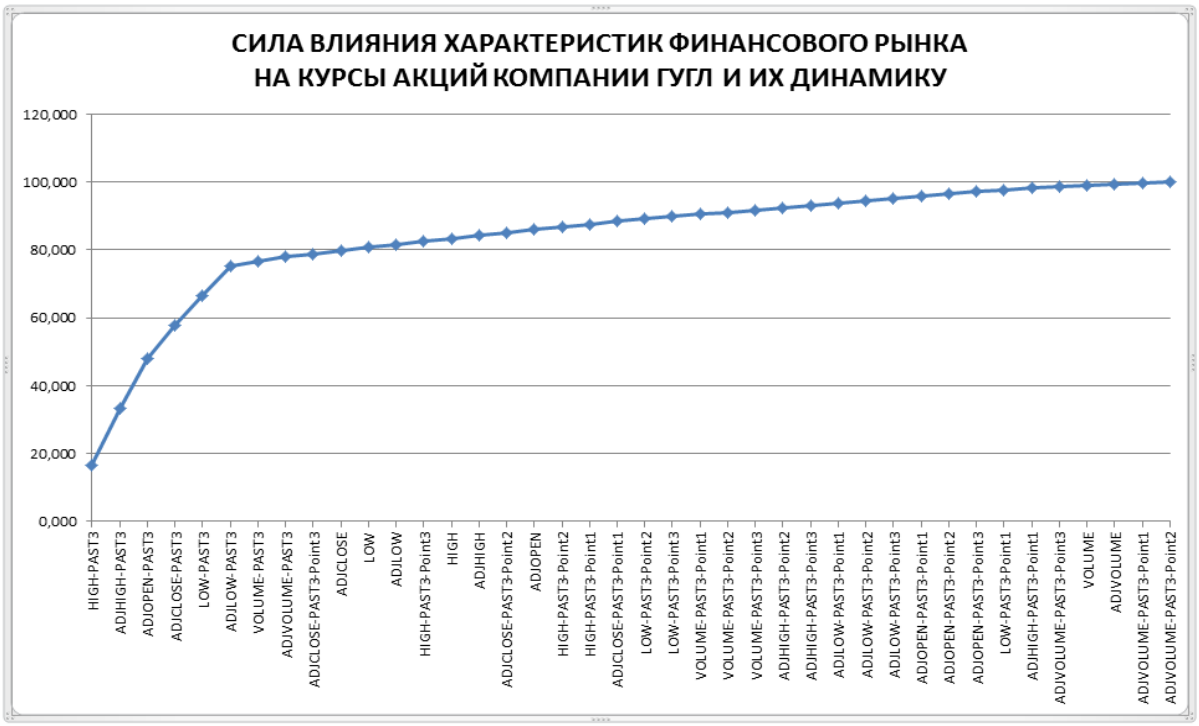


Рисунок 48. Информация о XLS-файлах и сила влияния характеристик финансового рынка на прогнозирование курсов акций компании Гугл и их динамику

Таблица 20 – Парето-таблица значимости описательных шкал, т.е. сила влияния характеристик финансового рынка на курсы акций компании Гугл и их динамику в СК-модели INF5

№	№%	Код	Наименование	Значимость, %	Значимость нараст. итогом, %
1	2,500	9	HIGH-PAST3	16,681	16,681
2	5,000	13	ADJHIGH-PAST3	16,681	33,363
3	7,500	15	ADJOPEN-PAST3	14,502	47,865
4	10,000	12	ADJCLOSE-PAST3	10,083	57,948
5	12,500	10	LOW-PAST3	8,614	66,562
6	15,000	14	ADJLOW-PAST3	8,614	75,176
7	17,500	11	VOLUME-PAST3	1,420	76,596
8	20,000	16	ADJVOLUME-PAST3	1,420	78,016
9	22,500	28	ADJCLOSE-PAST3-Point3	0,897	78,913
10	25,000	4	ADJCLOSE	0,896	79,809
11	27,500	2	LOW	0,896	80,705
12	30,000	6	ADJLOW	0,896	81,600
13	32,500	19	HIGH-PAST3-Point3	0,879	82,479
14	35,000	1	HIGH	0,878	83,357
15	37,500	5	ADJHIGH	0,878	84,235
16	40,000	27	ADJCLOSE-PAST3-Point2	0,868	85,103
17	42,500	7	ADJOPEN	0,863	85,966
18	45,000	18	HIGH-PAST3-Point2	0,860	86,826
19	47,500	17	HIGH-PAST3-Point1	0,819	87,646
20	50,000	26	ADJCLOSE-PAST3-Point1	0,816	88,462
21	52,500	21	LOW-PAST3-Point2	0,663	89,125
22	55,000	22	LOW-PAST3-Point3	0,663	89,788
23	57,500	23	VOLUME-PAST3-Point1	0,663	90,452
24	60,000	24	VOLUME-PAST3-Point2	0,663	91,115
25	62,500	25	VOLUME-PAST3-Point3	0,663	91,778
26	65,000	30	ADJHIGH-PAST3-Point2	0,663	92,442
27	67,500	31	ADJHIGH-PAST3-Point3	0,663	93,105
28	70,000	32	ADJLOW-PAST3-Point1	0,663	93,768
29	72,500	33	ADJLOW-PAST3-Point2	0,663	94,432
30	75,000	34	ADJLOW-PAST3-Point3	0,663	95,095
31	77,500	35	ADJOPEN-PAST3-Point1	0,663	95,759
32	80,000	36	ADJOPEN-PAST3-Point2	0,663	96,422
33	82,500	37	ADJOPEN-PAST3-Point3	0,663	97,085
34	85,000	20	LOW-PAST3-Point1	0,565	97,651
35	87,500	29	ADJHIGH-PAST3-Point1	0,557	98,208
36	90,000	40	ADJVOLUME-PAST3-Point3	0,377	98,585
37	92,500	3	VOLUME	0,376	98,961
38	95,000	8	ADJVOLUME	0,376	99,337
39	97,500	38	ADJVOLUME-PAST3-Point1	0,350	99,687
40	100,000	39	ADJVOLUME-PAST3-Point2	0,313	100,000

Из таблицы 9 видно, что всего 10% характеристик финансового рынка, т.е. характеристики, вместе оказывают около 58% суммарного влияния на прогнозирование курсов акций компании Гугл и их динамики, а оставшиеся 90% характеристик дают вместе лишь около 42% влияния.

При этом сила влияния наиболее и наименее значимых факторов отличается более чем в **53** раза, что очень существенно.

13.3.5.4.9. Степень детерминированности классов значениями обуславливающих их факторов

Степень детерминированности (обусловленности) класса в системе «Эйдос» количественно оценивается ***степенью вариабельности значений факторов*** (градаций описательных шкал) в колонке матрицы модели, соответствующей данному классу (рисунки 12-15).

В данной работе у нас классами являются курсы акций компании Гугл и их динамика (сценарии), а также значения точек сценариев, а значениями градаций описательных шкал – значения характеристик финансового рынка.

На рисунке 40 мы видим Парето-кривую степени детерминированности классов значениями характеристик финасового рынка нарастающим итогом.

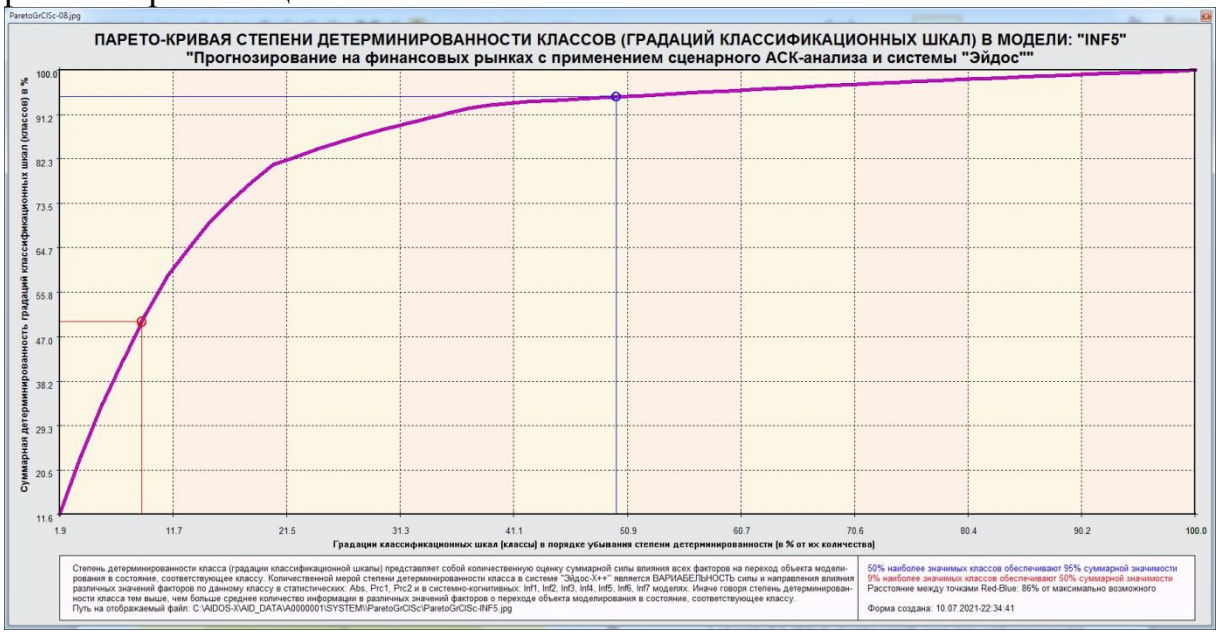


Рисунок 49. Парето-кривая степени детерминированности классов

Эта информация есть и в табличной форме (рисунок 41, таблица 10):

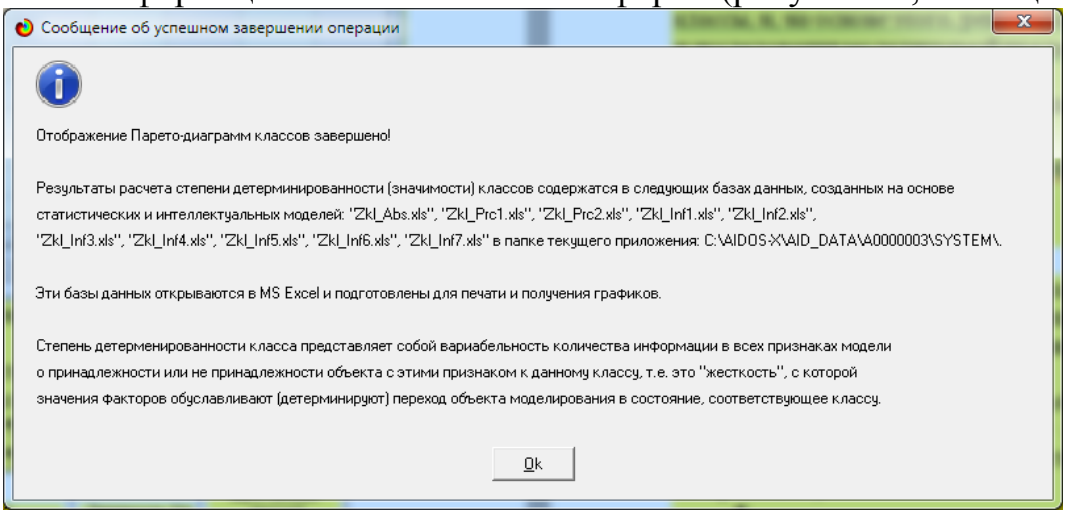


Рисунок 50. Информация о XLS-файлах

Таблица 21 – Парето-таблица степеней детерминированности (обусловленности) классов значениями характеристик финансового рынка в СК-модели INF5

№	№%	Код	Наименование	Значимость, %	Значимсть нараст. итогом, %
1	1,852	26	CLOSE-FUTURE3-CLOSE-FUTURE3-5,4,4	11,645	11,645
2	3,704	13	OPEN-FUTURE3-OPEN-FUTURE3-2,2,1	11,645	23,290
3	5,556	8	OPEN-FUTURE3-OPEN-FUTURE3-1,1,2	10,298	33,588
4	7,407	10	OPEN-FUTURE3-OPEN-FUTURE3-1,2,2	8,892	42,480
5	9,259	11	OPEN-FUTURE3-OPEN-FUTURE3-2,1,1	8,725	51,205
6	11,111	23	CLOSE-FUTURE3-CLOSE-FUTURE3-4,4,5	7,546	58,751
7	12,963	25	CLOSE-FUTURE3-CLOSE-FUTURE3-4,5,5	5,513	64,264
8	14,815	28	CLOSE-FUTURE3-CLOSE-FUTURE3-5,5,4	5,412	69,676
9	16,667	12	OPEN-FUTURE3-OPEN-FUTURE3-2,1,2	4,270	73,946
10	18,519	19	OPEN-FUTURE3-OPEN-FUTURE3-3,2,3	3,782	77,728
11	20,370	34	CLOSE-FUTURE3-CLOSE-FUTURE3-6,5,6	3,454	81,182

12	22,222	17	OPEN-FUTURE3-OPEN-FUTURE3-2,3,3	1,527	82,710
13	24,074	32	CLOSE-FUTURE3-CLOSE-FUTURE3-5,6,6	1,500	84,210
14	25,926	31	CLOSE-FUTURE3-CLOSE-FUTURE3-5,6,5	1,391	85,602
15	27,778	33	CLOSE-FUTURE3-CLOSE-FUTURE3-6,5,5	1,306	86,907
16	29,630	30	CLOSE-FUTURE3-CLOSE-FUTURE3-5,5,6	1,139	88,046
17	31,481	18	OPEN-FUTURE3-OPEN-FUTURE3-3,2,2	1,119	89,165
18	33,333	20	OPEN-FUTURE3-OPEN-FUTURE3-3,3,2	1,078	90,243
19	35,185	15	OPEN-FUTURE3-OPEN-FUTURE3-2,2,3	1,049	91,293
20	37,037	35	CLOSE-FUTURE3-CLOSE-FUTURE3-6,6,5	1,030	92,323
21	38,889	9	OPEN-FUTURE3-OPEN-FUTURE3-1,2,1	0,646	92,969
22	40,741	24	CLOSE-FUTURE3-CLOSE-FUTURE3-4,5,4	0,479	93,448
23	42,593	27	CLOSE-FUTURE3-CLOSE-FUTURE3-5,4,5	0,318	93,766
24	44,444	1	OPEN-1/3-{671.0, 980.0}	0,236	94,002
25	46,296	37	OPEN-FUTURE3-POINT1-OPEN-FUTURE3-Point1-1/3-{671.0, 980.0}	0,236	94,238
26	48,148	4	CLOSE-1/3-{668.3, 978.9}	0,234	94,472
27	50,000	49	CLOSE-FUTURE3-POINT2-CLOSE-FUTURE3-Point2-1/3-{668.3, 978.9}	0,234	94,706
28	51,852	46	CLOSE-FUTURE3-POINT1-CLOSE-FUTURE3-Point1-1/3-{668.3, 978.9}	0,234	94,940
29	53,704	22	CLOSE-FUTURE3-CLOSE-FUTURE3-4,4,4	0,233	95,173
30	55,556	52	CLOSE-FUTURE3-POINT3-CLOSE-FUTURE3-Point3-1/3-{668.3, 978.9}	0,232	95,405
31	57,407	43	OPEN-FUTURE3-POINT3-OPEN-FUTURE3-Point3-1/3-{671.0, 980.0}	0,231	95,636
32	59,259	40	OPEN-FUTURE3-POINT2-OPEN-FUTURE3-Point2-1/3-{671.0, 980.0}	0,230	95,866
33	61,111	7	OPEN-FUTURE3-OPEN-FUTURE3-1,1,1	0,227	96,093
34	62,963	38	OPEN-FUTURE3-POINT1-OPEN-FUTURE3-Point1-2/3-{980.0, 1191.0}	0,219	96,312
35	64,815	5	CLOSE-2/3-{978.9, 1189.1}	0,218	96,529
36	66,667	41	OPEN-FUTURE3-POINT2-OPEN-FUTURE3-Point2-2/3-{980.0, 1191.0}	0,211	96,740
37	68,519	16	OPEN-FUTURE3-OPEN-FUTURE3-2,3,2	0,210	96,951
38	70,370	2	OPEN-2/3-{980.0, 1191.0}	0,207	97,158
39	72,222	47	CLOSE-FUTURE3-POINT1-CLOSE-FUTURE3-Point1-2/3-{978.9, 1189.1}	0,207	97,365
40	74,074	44	OPEN-FUTURE3-POINT3-OPEN-FUTURE3-Point3-2/3-{980.0, 1191.0}	0,204	97,569
41	75,926	50	CLOSE-FUTURE3-POINT2-CLOSE-FUTURE3-Point2-2/3-{978.9, 1189.1}	0,201	97,769
42	77,778	29	CLOSE-FUTURE3-CLOSE-FUTURE3-5,5,5	0,197	97,967
43	79,630	14	OPEN-FUTURE3-OPEN-FUTURE3-2,2,2	0,195	98,162
44	81,481	53	CLOSE-FUTURE3-POINT3-CLOSE-FUTURE3-Point3-2/3-{978.9, 1189.1}	0,193	98,355
45	83,333	39	OPEN-FUTURE3-POINT1-OPEN-FUTURE3-Point1-3/3-{1191.0, 2105.9}	0,179	98,534
46	85,185	6	CLOSE-3/3-{1189.1, 2098.0}	0,179	98,713
47	87,037	3	OPEN-3/3-{1191.0, 2105.9}	0,170	98,883
48	88,889	21	OPEN-FUTURE3-OPEN-FUTURE3-3,3,3	0,168	99,051
49	90,741	36	CLOSE-FUTURE3-CLOSE-FUTURE3-6,6,6	0,163	99,214
50	92,593	42	OPEN-FUTURE3-POINT2-OPEN-FUTURE3-Point2-3/3-{1191.0, 2105.9}	0,162	99,376
51	94,444	48	CLOSE-FUTURE3-POINT1-CLOSE-FUTURE3-Point1-3/3-{1189.1, 2098.0}	0,159	99,534
52	96,296	54	CLOSE-FUTURE3-POINT3-CLOSE-FUTURE3-Point3-3/3-{1189.1, 2098.0}	0,157	99,691
53	98,148	51	CLOSE-FUTURE3-POINT2-CLOSE-FUTURE3-Point2-3/3-{1189.1, 2098.0}	0,156	99,847
54	100,000	45	OPEN-FUTURE3-POINT3-OPEN-FUTURE3-Point3-3/3-{1191.0, 2105.9}	0,153	100,000

Из рисунка 40 и таблицы 10 мы видим, что:

– всего 9% классов обуславливают 50% суммарной детерминированности всех классов;

– 50% наиболее сильно детерминированных классов обеспечивают 95% суммарной детерминированности.

Значения характеристик финансового рынка наиболее сильно детерминируют (обуславливают) следующие классы:

- CLOSE-FUTURE3-CLOSE-FUTURE3-5,4,4
- OPEN-FUTURE3-OPEN-FUTURE3-2,2,1
- OPEN-FUTURE3-OPEN-FUTURE3-1,1,2
- OPEN-FUTURE3-OPEN-FUTURE3-1,2,2
- OPEN-FUTURE3-OPEN-FUTURE3-2,1,1

а наименее сильно:

- OPEN-FUTURE3-POINT2-OPEN-FUTURE3-Point2-3/3-{1191.0, 2105.9}
- CLOSE-FUTURE3-POINT1-CLOSE-FUTURE3-Point1-3/3-{1189.1, 2098.0}
- CLOSE-FUTURE3-POINT3-CLOSE-FUTURE3-Point3-3/3-{1189.1, 2098.0}
- CLOSE-FUTURE3-POINT2-CLOSE-FUTURE3-Point2-3/3-{1189.1, 2098.0}
- OPEN-FUTURE3-POINT3-OPEN-FUTURE3-Point3-3/3-{1191.0, 2105.9}

При этом степень детерминированности наиболее и наименее детерминированных классов отличается в **76** раз, что очень существенно.

Чем выше степень детерминированности класса характеристиками рынка, тем легче определить этот класс по этим характеристикам.

Степень детерминированности классификационной шкалы является средним от степеней детерминированности ее градаций (классов).

В системе «Эйдос» эта информация приводится в табличных файлах, имена которых приведены на рисунке 42 и в таблице 11:

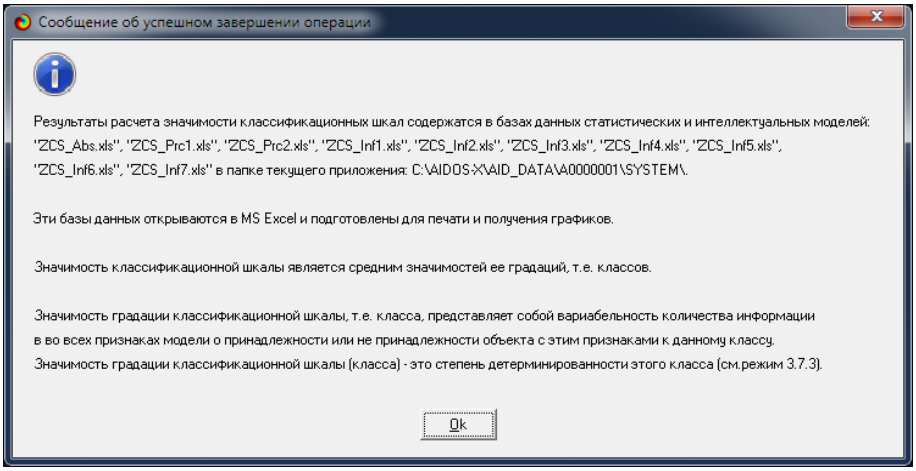


Рисунок 51. Информация о XLS-файлах

Таблица 22 – Парето-таблица степеней детерминированности (обусловленности) классификационных шкал значениями характеристик финансового рынка в СК-модели INF5

№	№%	Код	Наименование	Значимость, %	Значимость нараст. итогом, %
1	10,000	3	OPEN-FUTURE3	45,100	45,100
2	20,000	4	CLOSE-FUTURE3	34,624	79,724
3	30,000	5	OPEN-FUTURE3-Point1	2,654	82,378
4	40,000	2	CLOSE	2,643	85,021
5	50,000	1	OPEN	2,568	87,590
6	60,000	6	OPEN-FUTURE3-Point2	2,527	90,116
7	70,000	8	CLOSE-FUTURE3-Point1	2,510	92,627
8	80,000	9	CLOSE-FUTURE3-Point2	2,473	95,100
9	90,000	7	OPEN-FUTURE3-Point3	2,462	97,562
10	100,000	10	CLOSE-FUTURE3-Point3	2,438	100,000

Из таблицы 11 мы видим, что практически 80% суммарной детерминированности всех классификационных шкал обеспечивают 2 шкалы из 10, это сценарии курсов открытия и закрытия:

OPEN-FUTURE3

CLOSE-FUTURE3

Остальные 8 классификационных шкал суммарно обеспечивают лишь 20% суммарной детерминированности.

Слабее всего детерминированы значения 3-й точки сценариев открытия и закрытия.

Наиболее терминированная шкала: OPEN-FUTURE3 обусловлена характеристиками финансового рынка примерно в **18** раз сильнее, чем наименее детерминированная: CLOSE-FUTURE3-Point3.

13.3.6. Выводы

Как показывает анализ результатов численного эксперимента предложенное и реализованное в системе «Эйдос» решение поставленных задач является вполне эффективным, что позволяет обоснованно утверждать, что цель работы достигнута, поставленная проблема решена.

В результате проделанной работы, с помощью системы «Эйдос» были созданы 3 статистические и 7 системно-когнитивных моделей, в которых непосредственно на основе эмпирических данных сформированы обобщенные образы классов по курсам акций компании Гугл и их динамике, изучено влияние характеристик финансового рынка на эти классы, и, на основе этого, решены задачи идентификации и прогнозирования, классификации и исследования моделируемой предметной области путем исследования ее модели.

Со всеми моделями, созданными в данной статье, можно ознакомиться установив облачное Эйдос-приложение №**295** в режиме 1.3 системы «Эйдос». Саму систему можно бесптано скачать с сайта ее автора и разработчика проф.Е.В.Луценко по ссылке: http://lc.kubagro.ru/aidos/_Aidos-X.htm.

Дополнительную информацию по рассматриваемым вопросам можно получить из работ [16-22].

13.4. Выводы

В работе рассматривается теорема А.Н.Колмогорова, являющаяся обобщением теоремы В.И. Арнольда (1957) и представляющая собой важный шаг на пути к математическому решению 13-й проблемы Гильберта.

По своей сути замечательная теорема А.Н. Колмогорова является теоретической основой всей математической теории разложения функций в ряды, т.е. так называемой теории рядов. В математике разработано много различных конкретных вариантов разложений функций в ряды.

Однако, к сожалению определение вида базисных функций hij и весовых коэффициентов gj для данной конкретной функции F представляет собой математическую проблему, для которой пока не найдено общего математически строго решения.

При этом для частных случаев, т.е. конкретных видов базисных функций, таких решений найдено довольно много.

В данной работе предлагается рассматривать математическую модель АСК-анализа как вариант общего и универсального, но не строгого в математическом смысле, а практического решения проблемы разработки базисных функций и весовых коэффициентов для разложения в ряд по ним произвольной функции состояния идентифицируемого объекта.

В этом контексте функция F интерпретируется как конкретный образ состояния идентифицируемого объекта, функция hij – обобщенный образ j-

го класса, а функция gj – мера сходства образа объекта с образом класса. Приводятся численные примеры технического, фундаментального и технофундаментального сценарного АСК-анализа.

Таким образом, сценарный метод АСК-анализа обеспечивает синтез технического и фундаментального подходов путем применения теории информации для обобщения теории рядов.

В этих численных примерах на основе анализа исходных данных выявляются ранее наблюдавшиеся прошлые и будущие сценарии развития событий и на основе их обобщения формируются обобщенные образы сценариев развития событий, которые рассматриваются в виде базисных функций классов и детерминирующих их значений факторов. При прогнозировании текущая ситуация сравнивается с этими обобщенными образами и разлагается в ряд по ним (прямое преобразование, объектный анализ). Средневзвешенный прогноз формируется путем обратного преобразования образов классов с их весами, т.е. как их взвешенная суперпозиция. При этом в качестве базисных функций используются обобщенные образы прогнозируемых сценариев того что будет и того что не будет с их весами, в качестве которых используется достоверность прогноза.

Предлагаемый метод сценарного автоматизированного системно-когнитивного анализа и реализующий его программный инструментарий, в качестве которого в настоящее время выступает интеллектуальная система «Эйдос», разработаны в универсальной постановке, не зависящей от предметной области. Это означает, что они могут быть применены в любом направлении науки и практической деятельности, в которых накоплена информация о реальных сценариях развития событий.

Необходимо также отметить, что интеллектуальное облачное Эйдос-приложение, использованное в данной работе для численных примеров, размещено в Эйдос-облаке под номером 205 и доступно для загрузки и исследования в диспетчере приложения (режим 1.3) системы «Эйдос». Сама система «Эйдос» представляет собой открытое программное обеспечение и находится в полном открытом бесплатном доступе на сайте автора по адресу: http://lc.kubagro.ru/aidos/_Aidos-X.htm./

Из-за ограничений на объем статьи численный пример, наглядно демонстрирующий сценарный АСК-анализ в системе «Эйдос» будет представлен в следующей статье. Данная статья, объединенная с последующей, в которой описан численный пример, размещена в РесечГейт по адресу: https://www.researchgate.net/publication/343365649

В заключение автор выражает огромную благодарность доктору технических наук, доктору экономических наук, кандидату физико-математических наук, профессору Александру Ивановичу Орлову за внимательное ознакомление с первым вариантом статьи и ряд ценных замечаний, учет которых способствовал повышению качества статьи.

ГЛАВА 14. СПЕКТРАЛЬНЫЙ АВТОМАТИЗИРОВАННЫЙ СИСТЕМНО-КОГНИТИВНЫЙ АНАЛИЗ КОНКРЕТНЫХ И ОБОБЩЕННЫХ ИЗОБРАЖЕНИЙ

14.1. Введение

Автоматизированный системно-когнитивный анализ (АСК-анализ) изображений обеспечивает автоматическое выявление признаков конкретных изображений из цветов пикселей и контуров изображений, синтез обобщенных образов изображений (классов), выявление наиболее характерных и нехарактерных для классов признаков изображений, определение ценности признаков изображений для их различения, удаление из модели малоценных признаков (абстрагирование), решение задач количественного сравнения конкретных изображений с обобщенными образами классов и обобщенных образов классов друг с другом, а также задачи исследования моделируемой предметной области путем исследования ее модели [1-10].

В работе рассматриваются новые возможности АСК-анализа и реализующей его интеллектуальной системы «Эйдос», обеспечивающие выявление признаков изображений путем их спектрального анализа, формирования обобщенных спектров классов, решение задач сравнения изображений конкретных объектов с классами и классов друг с другом по их спектрам.

Впервые стало возможным формировать обобщенные спектры классов с весами цветов по степени их характерности и нехарактерности для классов, причем это не интенсивность цвета в спектре, а количество информации в цвете о принадлежности объекта с этим цветом к данному классу.

По сути, речь идет об обобщении спектрального анализа путем применении интеллектуальных когнитивных технологий и теории информации в спектральном анализе.

Во-первых, все говорят о том, что в спектральных линиях содержится информация о том, какой элемент или вещество входят в состав объекта, но никто не удосужился посчитать какое же это конкретно количество этой информации, а затем использовать его для определения состава объекта методы распознавания образов, основанные на использовании этой информации.

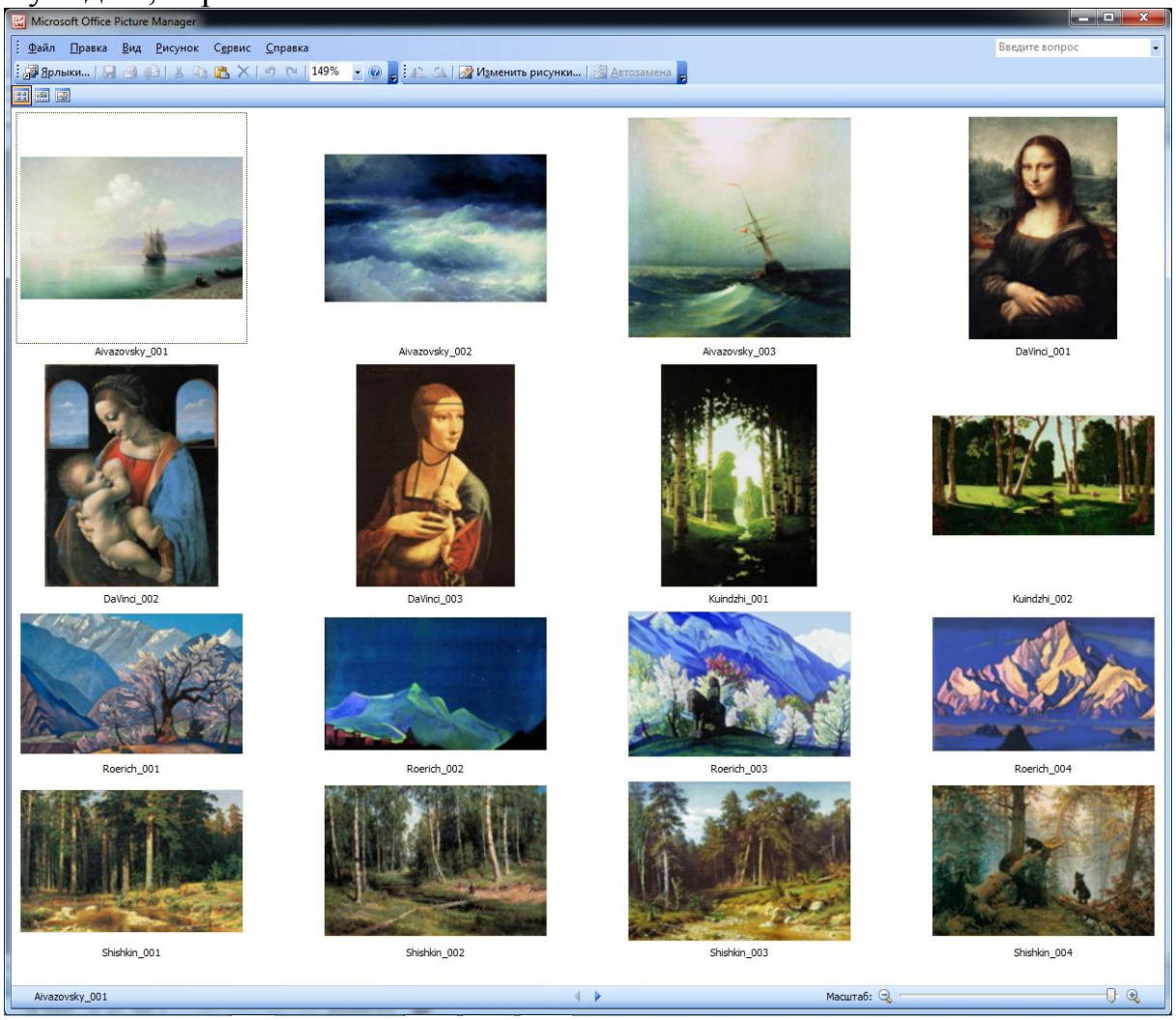
Во-вторых, спектральный анализ традиционно используется для определения элементарного и молекулярного состава объекта, а мы предлагаем использовать его не только для этого, но и для идентификации любых изображений.

14.2. Постановка задачи

Пусть у нас есть некоторое количество изображений, сгруппированных по определенному принципу в классы. Необходимо получить спектральные образы каждого конкретного изображения и обобщенные спектры классов и решить задачу идентификации конкретных изображений с классами по их спектрам и задачу сравнения классов по их обобщенным спектрам, а также другие задачи, решаемые в АСК-анализе и системе «Эйдос».

14.3. Исходные данные

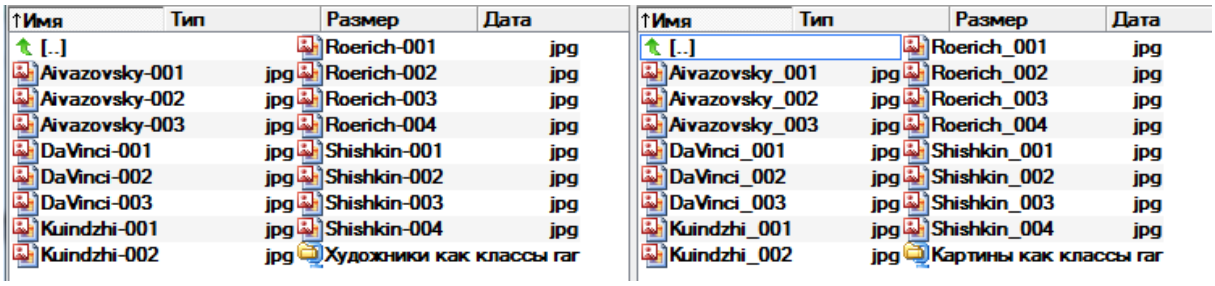
В качестве исходных данных для численного примера выбрано 16 картин выдающихся художников Леонардо да Винчи, Айвазовского, Куинджи, Рериха и Шишкина:



Изображения картин должны быть записаны в папку: c:\Aidos-X\AID_DATA\Inp_data\ в виде файлов формата jpg или bmp. Желательно, чтобы размер изображений не превосходил 700 × 700 pix. Если изображения будут большего размера это не только увеличит время обработки, но и может привести к их неправильному отображению на

графических форах. Режимы пакетного переименования, изменения формата изображений и изменения их размеров есть во многих программах, например в ACDSee.

Часть имени файла изображения до черточки «-» воспринимается системой как имя класса, к которому относится данное изображение (на рисунке слева). Если мы хотим, чтобы как классы рассматривались сами изображения, то в их именах не должно быть черточки (на рисунке справа она заменена нижней чертой).



Если мы хотим, чтобы оба варианта осуществлялись одновременно, то необходимо продублировать файлы с одними и теми же изображениями с именами как слева на рисунке, и как справа, в одну папку: c:\Aidos-X\AID_DATA\Inp_data\.

14.4. Формализация предметной области

Формализация предметной области включает разработку классификационных и описательных шкал и градаций и кодирование исходных данных с их использованием, в результате чего формируется обучающая выборка, представляющая собой нормализованную базу исходных данных.

В случае обработки изображений по их спектрам формализация предметной области осуществляется в режиме 4.7. АСК-анализ изображений по пикселям, спектрам и контурам в системе "Эйдос"'.

Благодаря данному режиму система "Эйдос" может:

1. Измерять спектры графических объектов (т.е. очень точно определять цвета, присутствующие в изображении).

2. Формировать обобщенные спектры классов. При этом рассчитывается количество информации в каждом цвете обобщенного спектра класса о принадлежности конкретного объекта с этим цветом в спектре к данному классу.

3. Сравнивать конкретные объекты с классами по их спектрам. При этом рассчитывается суммарное количество информации в цветах спектра конкретного объекта о его принадлежности к обобщенному образу класса.

4. Сравнивать классы друг с другом по их спектрам.

В качестве спектра изображения в системе рассматривается доля пикселей разных цветов в общем числе пикселей изображения.

Данный режим обеспечивает:

- ввод изображений в систему по пикселям (для этого выполнить первые два режима подготовки данных);

- измерение спектров изображений с заданным числом цветовых диапазонов (цветовых интервалов) (выполнить 4-й режим подготовки данных);

- рассмотрение характеристик спектра конкретных изображений как их признаков при формировании моделей (наряду с пикселями);

- вывод исходных изображения с их спектрами на экран и запись в виде файлов в папку: ..\AID_DATA\InpSpectrPix\.

- формирование обобщенных спектров изображений, относящихся к различным группам, классам (обобщенные спектры классов);

- количественное сравнение конкретных изображений по их спектрам с обобщенными спектрами классов, т.е. решение задачу идентификации (классификации, диагностики, распознавания, прогнозирования);

- количественное сравнение обобщенных спектров классов друг с другом и решение задач кластерно-конструктивного анализа;

- другие стандартные возможности работы системы "Эйдос" с созданными моделями, отражающими спектры изображений.

Исходные изображения должны быть в формате jpg или bmp и находиться непосредственно в папке: ../Aid_data/Inp_data/, если ставится формализации предметной области и синтеза модели, ../Aid_data/Inp_rasp/, если ставится цель формирования распознаваемой выборки.

Для режимов спектрального анализа изображений не важно, как они масштабированы и повернуты, но желательно, чтобы они были без фона. Пакетные on-line сервисы, обеспечивающие "оконтуриване и удаление фона" изображений можно найти в Internet по запросу, который в кавычках.

Порядок работы в системе "Эйдос" для создания и верификации моделей описан в режиме 6.4.

1. Исходные изображения должны быть в папке: ../AID_DATA/INP_DATA/ без поддиректорий. Часть имени файла до тире: "-" , если оно есть, используется как имя класса, для формирования которого используется данное изображение. Если тире нет, то как имя класса используется имя файла изображения целиком.

2. Для создания модели нужно в режиме 2.3.2.5 или "Подготовка данных" сбросить БД "Image.dbf" и ввести в нее исходные изображения, затем создать базу "Inp_data".

3. После ввода изображений в систему (режим подготовки данных) необходимо создать модель в 3-м режиме АСК-анализа изображений по пикселям (режим 2.3.2.3 с параметрами по умолчанию).

4. Посмотреть на классификационные шкалы и градации в режиме 2.1.5. Посмотреть на описательные шкалы и градации в режиме 2.2.

6. Посмотреть на обучающую выборку в режиме 2.3.1.

7. Посмотреть файл исходных данных Inp_data.xls или Inp_rasp.xls в папке: ../AID_DATA/INP_DATA/.

8. Запустить режим синтеза и верификации моделей с параметрами по умолчанию (режим 3.5).

9. Посмотреть сформированные модели в режиме 5.5.

10. Посмотреть достоверность моделей в режиме 4.1.3.6.

11. Посмотреть частотные распределения уровней сходства при истинно и ложно положительных и отрицательных решениях (режим 4.1.3.11).

12. Сделать текущей наиболее достоверную модель по L2-критерию (в режиме 5.6).

13. Провести распознавание в наиболее достоверной модели в режиме 4.1.2.

14. Посмотреть результаты распознавания в режимах 4.1.3.

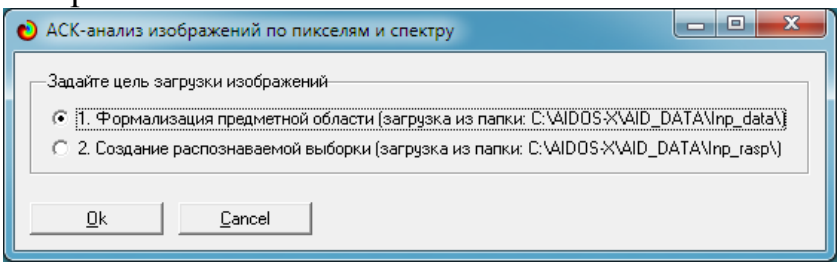
15. Провести анализ наиболее достоверной модели в 4-й подсистеме, в которой, в частности, можно сравнить классы по их обобщенным спектрам.

При распознавании изображений по их спектрам в ранее созданной модели необходимо в режиме 2.3.2.5 или "Подготовка данных" сбросить БД "Image.dbf" и ввести в нее изображения из папки: ../Aid_data/Inp_rasp/, затем создать базу "Inp_rasp", ввести ее в систему в режиме 2.3.2.3 и провести распознавание в режиме 4.1.2. Результаты распознавания будут в различных выходных формах режима 4.1.3.

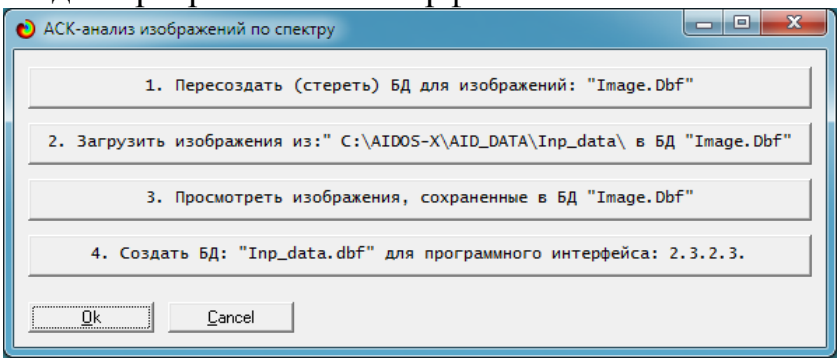
Желательно, чтобы изображения были *не более* 640 на 480 пикселей, а лучше около 400 на 300 pix или еще меньше, например 150 или 100 пикселей.

Экранная форма помощи по данному режиму приведена ниже:

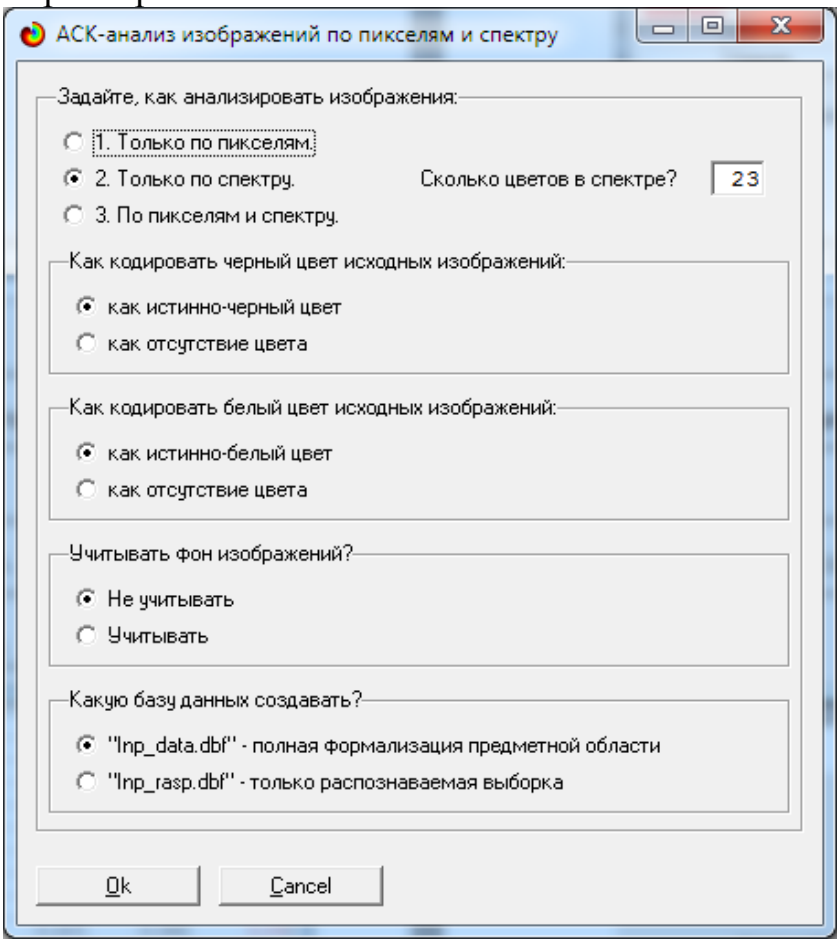
Входим в режим 2.3.2.5:

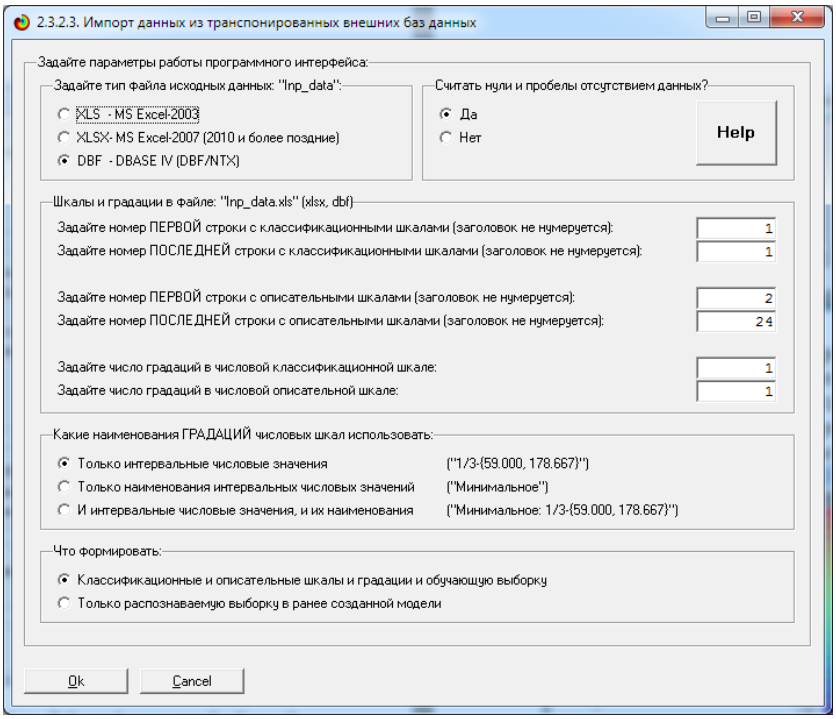


Если кликнуть по кнопке «Подготовка данных» и последовательно выполнить режимы, представленные на рисунке, то создается база данных «Inp_data.dbf» для программного интерфейса 2.3.2.3.



Непосредственно перед созданием этой базы данных запрашиваются следующие параметры:

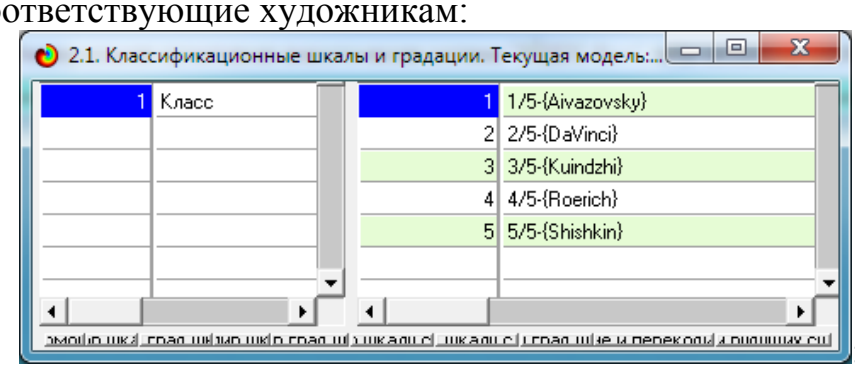




В результате формируется файл Inp_data.dbf с информацией об изображениях для универсального программного интерфейса 2.3.2.3, обеспечивающего ввода данных из внешних источников данных. Этот программный интерфейс создает новое приложение, название которого можно поменять в диспетчере приложений 1.3, включающее классификационные и описательные шкалы и градации и обучающую выборку.

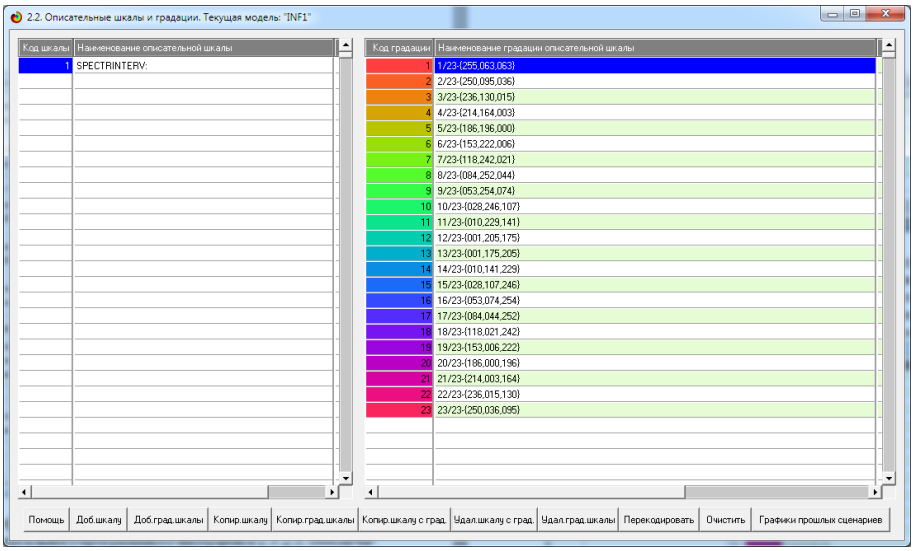
14.4.1. Классификационные и описательные шкалы и градации

Ниже представлена классификационная шкала и ее градации, т.е. классы, соответствующие художникам:



:

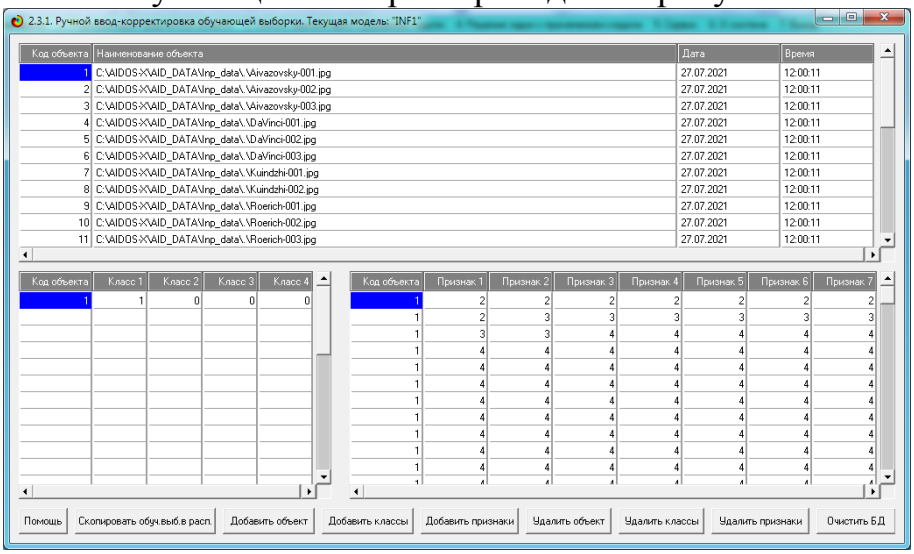
Описательная шкала представляет собой спектр, включающий 35 цветовых диапазонов. Количество цветовых диапазонов (в данном случае 35) задано перед подготовкой данных. Каждому цветовому диапазону соответствует сове сочетание яркостей лучей Red, Green, Blue, которые получены в системе «Эйдос» расчетным путем для формирования спектра:



14.4.2. Обучающая выборка

Обучающая выборка представляет собой описание каждой картины в виде онтологии, по сути, конкретного определения путем подведения под более общую категорию (класс) и указания специфических признаков (значений пикселей)[38], т.е. путем указания для каждого ее пикселя кода цветового диапазона в соответствии с описательными шкалами и градациями и указания принадлежности к обобщающему классу в соответствии классификационными шкалами и градациями.

Фрагмент обучающей выборки приведен на рисунке ниже:



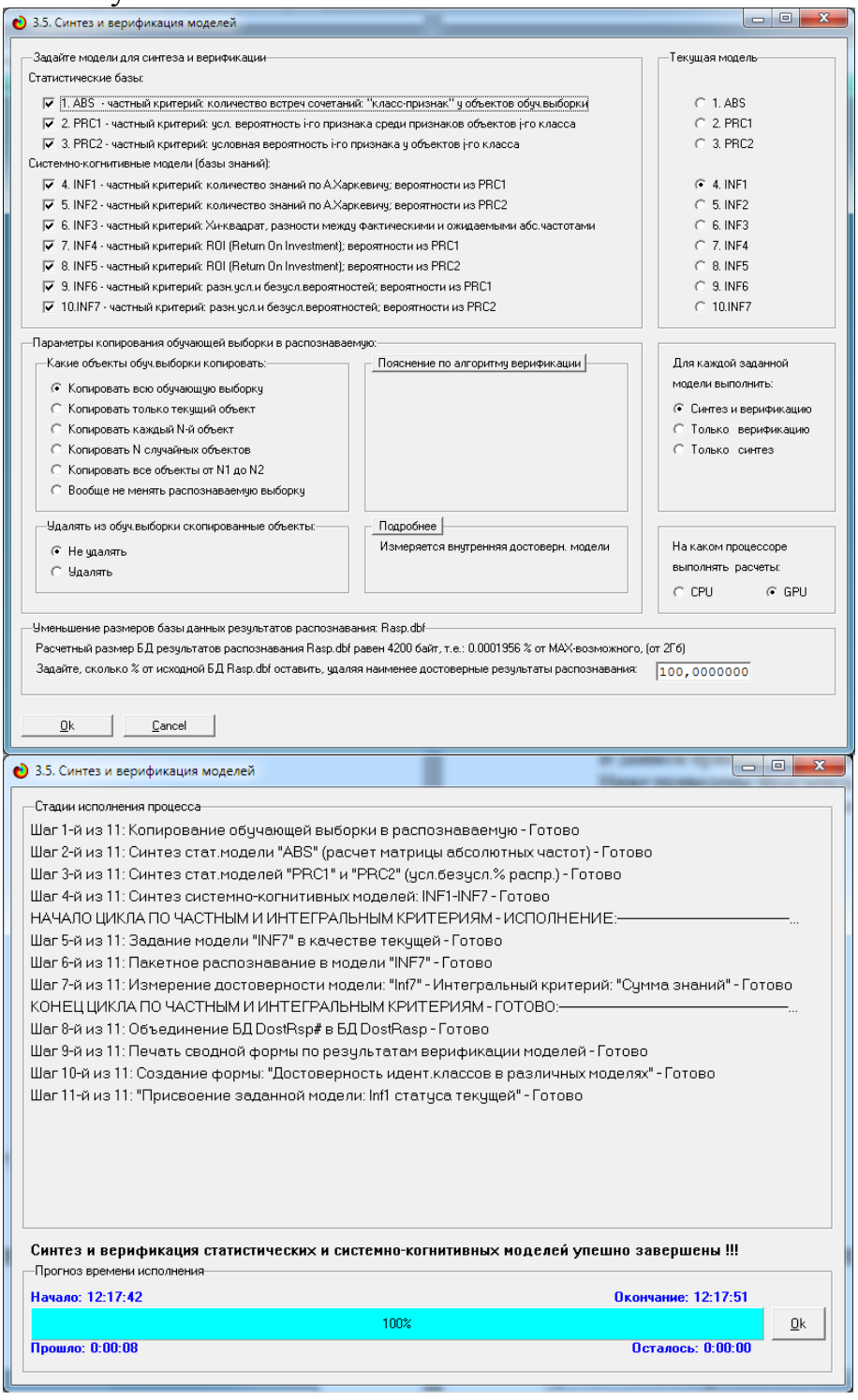
В обучающей выборке для каждого объекта обучающей выборки (верхнее окно) приведен код класса, к которому относится этот объект (нижнее левое окно) и для каждого пикселя изображения приведен код его цветового диапазона (нижнее правое окно).

---

38 сочетания экстенсионального и интенсионального описаний

14.5. Синтез и верификация модели

Синтез и верификация модели осуществляется в режиме 3.5 с параметрами по умолчанию:



В данном примере это занимает 8 секунд на среднем компьютере.

Ниже приведены фрагменты некоторых из созданных 3 статистических и 7 системно-когнитивных моделей:

5.5. Модель: "1. ABS - частный критерий: количество встреч сочетаний: "Класс-признак" у объектов обуч.выборки"

Код признака	Наименование описательной шкалы и градации	1. КЛАСС 1/5 {AIVAZOVSKY}	2. КЛАСС 2/5 {DAVINCI}	3. КЛАСС 3/5 {KUINDZHI}	4. КЛАСС 4/5 {ROERICH}	5. КЛАСС 5/5 {SHISHKIN}	Сумма	Среднее	Средн. квадр. откл.
1	SPECTRINTERV:-1/23-{255,063,063}		30	4	40	10	84	16.80	17.36
2	SPECTRINTERV:-2/23-{250,095,036}	38	178	40	92	200	548	109.60	76.04
3	SPECTRINTERV:-3/23-{236,130,015}	36	465	88	100	476	1165	233.00	218.17
4	SPECTRINTERV:-4/23-{214,164,003}	185	474	280	61		1000	200.00	187.76
5	SPECTRINTERV:-5/23-{186,196,000}	13	61	224	41	294	633	126.60	124.54
6	SPECTRINTERV:-6/23-{153,222,006}	132	86	452	58	500	1228	245.60	212.66
7	SPECTRINTERV:-7/23-{118,242,021}	45	1	44	15	9	114	22.80	20.43
8	SPECTRINTERV:-8/23-{084,252,044}	128	110		50	337	625	125.00	128.87
9	SPECTRINTERV:-9/23-{053,254,074}	144	2	20	35	34	235	47.00	55.85
10	SPECTRINTERV:-10/23-{028,246,107}	214	5	23	55	53	350	70.00	83.19
11	SPECTRINTERV:-11/23-{010,229,141}	200	83	28	111	291	713	142.60	103.68
12	SPECTRINTERV:-12/23-{001,205,175}	198	6	19	129	42	394	78.80	82.10
13	SPECTRINTERV:-13/23-{001,175,205}	191	46	14	189	123	563	112.60	81.01
14	SPECTRINTERV:-14/23-{010,141,229}	418	14	29	451	92	1004	200.80	215.65
15	SPECTRINTERV:-15/23-{028,107,246}	29	26	5	306	18	384	76.80	128.46
16	SPECTRINTERV:-16/23-{053,074,254}	47		3	195	1	246	49.20	83.88
17	SPECTRINTERV:-17/23-{084,044,252}	41	21	6	139	9	216	43.20	55.29
18	SPECTRINTERV:-18/23-{118,021,242}				9		9	1.80	4.02
19	SPECTRINTERV:-19/23-{153,006,222}	2	6		62	1	71	14.20	26.82
20	SPECTRINTERV:-20/23-{186,000,196}	31		5	26		62	12.40	14.94
21	SPECTRINTERV:-21/23-{214,003,164}	6	47	1	75	7	136	27.20	32.48
22	SPECTRINTERV:-22/23-{236,015,130}				33	1	34	6.80	14.65
23	SPECTRINTERV:-23/23-{250,036,095}	23	80	4	104	34	245	49.00	41.57
	Сумма числа признаков	2121	1741	1289	2376	2532	10059		
	Среднее	92	76	56	103	110		87.47	
	Среднеквадратичное отклонение	105	132	112	103	159			121.57

5.5. Модель: "2. PRC1 - частный критерий: усл. вероятность i-го признака среди признаков объектов j-го класса"

Код признака	Наименование описательной шкалы и градации	1. КЛАСС 1/5 {AIVAZOV...	2. КЛАСС 2/5 {DAVINCI}	3. КЛАСС 3/5 {KUINDZHI}	4. КЛАСС 4/5 {ROERICH}	5. КЛАСС 5/5 {SHISHKIN}	Безусл. вероятн.	Среднее	Средн. квадр. откл.
1	SPECTRINTERV:-1/23-{255,063,063}		1.723	0.310	1.684	0.395	4.112	0.822	0.818
2	SPECTRINTERV:-2/23-{250,095,036}	1.792	10.224	3.103	3.872	7.899	26.890	5.378	3.541
3	SPECTRINTERV:-3/23-{236,130,015}	1.697	26.709	6.827	4.209	18.799	58.241	11.648	10.672
4	SPECTRINTERV:-4/23-{214,164,003}	8.722	27.226	21.722	2.567		60.238	12.048	11.938
5	SPECTRINTERV:-5/23-{186,196,000}	0.613	3.504	17.378	1.726	11.611	34.831	6.966	7.242
6	SPECTRINTERV:-6/23-{153,222,006}	6.223	4.940	35.066	2.441	19.747	68.417	13.683	13.716
7	SPECTRINTERV:-7/23-{118,242,021}	2.122	0.057	3.413	0.631	0.355	6.579	1.316	1.416
8	SPECTRINTERV:-8/23-{084,252,044}	6.035	6.318		2.104	13.310	27.767	5.553	5.092
9	SPECTRINTERV:-9/23-{053,254,074}	6.789	0.115	1.552	1.473	1.343	11.272	2.254	2.602
10	SPECTRINTERV:-10/23-{028,246,107}	10.090	0.287	1.784	2.315	2.093	16.569	3.314	3.870
11	SPECTRINTERV:-11/23-{010,229,141}	9.430	4.767	2.172	4.672	11.493	32.534	6.507	3.827
12	SPECTRINTERV:-12/23-{001,205,175}	9.335	0.345	1.474	5.429	1.659	18.242	3.648	3.712
13	SPECTRINTERV:-13/23-{001,175,205}	9.005	2.642	1.086	7.955	4.858	25.546	5.109	3.377
14	SPECTRINTERV:-14/23-{010,141,229}	19.708	0.804	2.250	18.981	3.633	45.377	9.075	9.431
15	SPECTRINTERV:-15/23-{028,107,246}	1.367	1.493	0.388	12.879	0.711	16.838	3.368	5.337
16	SPECTRINTERV:-16/23-{053,074,254}	2.216		0.233	8.207	0.039	10.695	2.139	3.516
17	SPECTRINTERV:-17/23-{084,044,252}	1.933	1.206	0.465	5.850	0.355	9.810	1.962	2.264
18	SPECTRINTERV:-18/23-{118,021,242}				0.379		0.379	0.076	0.169
19	SPECTRINTERV:-19/23-{153,006,222}	0.094	0.345		2.609	0.039	3.088	0.618	1.122
20	SPECTRINTERV:-20/23-{186,000,196}	1.462		0.388	1.094		2.944	0.589	0.662
21	SPECTRINTERV:-21/23-{214,003,164}	0.283	2.700	0.078	3.157	0.276	6.493	1.299	1.499
22	SPECTRINTERV:-22/23-{236,015,130}				1.389	0.039	1.428	0.286	0.617
23	SPECTRINTERV:-23/23-{250,036,095}	1.084	4.595	0.310	4.377	1.343	11.710	2.342	1.995
	Сумма	100.000	100.000	100.000	100.000	100.000	500.000		
	Среднее	4.348	4.348	4.348	4.348	4.348		4.348	
	Среднеквадратичное отклонение	4.930	7.588	8.676	4.318	6.265			6.442

5.5. Модель: "4. INF1 - частный критерий: количество знаний по А.Харкевичу; вероятности из PRC1"

Код признака	Наименование описательной шкалы и градации	1. КЛАСС 1/5 (AIVAZOVSKY)	2. КЛАСС 2/5 (DAVINCI)	3. КЛАСС 3/5 (KUINDZHI)	4. КЛАСС 4/5 (ROERICH)	5. КЛАСС 5/5 (SHISHKIN)	Сумма	Среднее	Средн. квадр. откл.
1	SPECTRINTERV:-1/23-{255,063,063}		0.183	-0.249	0.177	-0.189	-0.079	-0.016	0.201
2	SPECTRINTERV:-2/23-{250,095,036}	-0.280	0.159	-0.142	-0.086	0.094	-0.256	-0.051	0.178
3	SPECTRINTERV:-3/23-{236,130,015}	-0.484	0.211	-0.133	-0.255	0.122	-0.539	-0.108	0.282
4	SPECTRINTERV:-4/23-{214,164,003}	-0.033	0.254	0.197	-0.341		0.077	0.015	0.234
5	SPECTRINTERV:-5/23-{186,196,000}	-0.587	-0.148	0.256	-0.326	0.154	-0.650	-0.130	0.345
6	SPECTRINTERV:-6/23-{153,222,006}	-0.170	-0.228	0.266	-0.406	0.121	-0.416	-0.083	0.272
7	SPECTRINTERV:-7/23-{118,242,021}	0.158	-0.751	0.278	-0.147	-0.292	-0.755	-0.151	0.406
8	SPECTRINTERV:-8/23-{084,252,044}	-0.007	0.004		-0.273	0.192	-0.084	-0.017	0.166
9	SPECTRINTERV:-9/23-{053,254,074}	0.269	-0.759	-0.103	-0.116	-0.140	-0.849	-0.170	0.370
10	SPECTRINTERV:-10/23-{028,246,107}	0.268	-0.628	-0.168	-0.103	-0.128	-0.759	-0.152	0.319
11	SPECTRINTERV:-11/23-{010,229,141}	0.072	-0.100	-0.298	-0.105	0.122	-0.309	-0.062	0.166
12	SPECTRINTERV:-12/23-{001,205,175}	0.219	-0.612	-0.246	0.082	-0.216	-0.774	-0.155	0.323
13	SPECTRINTERV:-13/23-{001,175,205}	0.120	-0.189	-0.413	0.089	-0.036	-0.430	-0.086	0.220
14	SPECTRINTERV:-14/23-{010,141,229}	0.171	-0.635	-0.375	0.162	-0.255	-0.931	-0.186	0.350
15	SPECTRINTERV:-15/23-{028,107,246}	-0.259	-0.236	-0.576	0.306	-0.423	-1.188	-0.238	0.334
16	SPECTRINTERV:-16/23-{053,074,254}	-0.025		-0.593	0.305	-1.039	-1.352	-0.270	0.538
17	SPECTRINTERV:-17/23-{084,044,252}	-0.026	-0.145	-0.385	0.253	-0.453	-0.758	-0.152	0.285
18	SPECTRINTERV:-18/23-{118,021,242}				0.364		0.364	0.073	0.163
19	SPECTRINTERV:-19/23-{153,006,222}	-0.507	-0.181		0.329	-0.726	-1.085	-0.217	0.416
20	SPECTRINTERV:-20/23-{186,000,196}	0.218		-0.117	0.145		0.245	0.049	0.132
21	SPECTRINTERV:-21/23-{214,003,164}	-0.394	0.174	-0.720	0.214	-0.400	-1.126	-0.225	0.405
22	SPECTRINTERV:-22/23-{236,015,130}				0.356	-0.541	-0.185	-0.037	0.321
23	SPECTRINTERV:-23/23-{250,036,095}	-0.204	0.160	-0.519	0.148	-0.150	-0.565	-0.113	0.282
	Сумма	-1.482	-3.469	-4.042	0.770	-4.184	-12.405		
	Среднее	-0.064	-0.151	-0.176	0.033	-0.182		-0.108	
	Среднеквадратичное отклонение	0.255	0.319	0.284	0.247	0.308			0.279

5.5. Модель: "6. INF3 - частный критерий: Хи-квадрат, разности между фактическими и ожидаемыми абс.частотами"

Код признака	Наименование описательной шкалы и градации	1. КЛАСС 1/5 (AIVAZOVSKY)	2. КЛАСС 2/5 (DAVINCI)	3. КЛАСС 3/5 (KUINDZHI)	4. КЛАСС 4/5 (ROERICH)	5. КЛАСС 5/5 (SHISHKIN)	Сумма	Среднее	Средн. квадр. откл.
1	SPECTRINTERV:-1/23-{255,063,063}	-17.712	15.461	-6.764	20.159	-11.144	0.000	0.000	16.801
2	SPECTRINTERV:-2/23-{250,095,036}	-77.549	83.153	-30.223	-37.441	62.060	0.000	0.000	69.092
3	SPECTRINTERV:-3/23-{236,130,015}	-209.647	263.363	-61.288	-175.180	182.752	0.000	0.000	212.811
4	SPECTRINTERV:-4/23-{214,164,003}	-25.856	300.921	151.856	-175.206	-251.715	0.000	0.000	228.221
5	SPECTRINTERV:-5/23-{186,196,000}	-120.472	-48.559	142.885	-108.519	134.664	0.000	0.000	129.613
6	SPECTRINTERV:-6/23-{153,222,006}	-126.931	-126.541	294.639	-232.061	190.894	0.000	0.000	228.708
7	SPECTRINTERV:-7/23-{118,242,021}	20.962	-18.731	29.392	-11.928	-19.695	0.000	0.000	23.368
8	SPECTRINTERV:-8/23-{084,252,044}	-3.785	1.826	-80.090	-97.629	179.678	0.000	0.000	109.827
9	SPECTRINTERV:-9/23-{053,254,074}	94.449	-38.674	-10.114	-20.508	-25.153			53.786
10	SPECTRINTERV:-10/23-{028,246,107}	140.200	-55.578	-21.850	-27.672	-35.100	0.000	0.000	79.404
11	SPECTRINTERV:-11/23-{010,229,141}	49.660	-40.405	-63.367	-57.415	111.527			77.215
12	SPECTRINTERV:-12/23-{001,205,175}	114.923	-62.193	-31.489	35.935	-57.176	0.000	0.000	75.212
13	SPECTRINTERV:-13/23-{001,175,205}	72.288	-51.443	-58.145	56.016	-18.715	0.000	0.000	60.706
14	SPECTRINTERV:-14/23-{010,141,229}	206.301	-159.771	-99.657	213.849	-160.722			193.378
15	SPECTRINTERV:-15/23-{028,107,246}	-51.969	-40.462	-44.207	215.297	-78.659	0.000	0.000	121.277
16	SPECTRINTERV:-16/23-{053,074,254}	-4.871	-42.577	-28.523	136.893	-60.922	0.000	0.000	79.217
17	SPECTRINTERV:-17/23-{084,044,252}	-4.545	-16.385	-21.679	87.979	-45.370	0.000	0.000	51.376
18	SPECTRINTERV:-18/23-{118,021,242}	-1.898	-1.558	-1.153	6.874	-2.265	0.000		3.865
19	SPECTRINTERV:-19/23-{153,006,222}	-12.971	-6.289	-9.098	45.229	-16.872	0.000	0.000	25.597
20	SPECTRINTERV:-20/23-{186,000,196}	17.927	-10.731	-2.945	11.355	-15.606	0.000	0.000	14.298
21	SPECTRINTERV:-21/23-{214,003,164}	-22.676	23.461	-16.428	42.876	-27.233	0.000	0.000	31.283
22	SPECTRINTERV:-22/23-{236,015,130}	-7.169	-5.885	-4.357	24.969	-7.558	0.000		14.014
23	SPECTRINTERV:-23/23-{250,036,095}	-28.660	37.596	-27.395	46.129	-27.670	0.000	0.000	38.337
	Сумма	0.000	0.000	0.000	0.000	0.000	0.000		
	Среднее	0.000	0.000	0.000	0.000	0.000		0.000	
	Среднеквадратичное отклонение	90.123	102.231	86.747	111.352	105.699			97.905

В матрице абсолютных частот ABS показано число встреч пикселей разных спектральных диапазонов в разных классах. В матрице условных и безусловных процентных распределений PRC1 показаны доли пикселей разных спектральных диапазонов среди всех пикселей изображений разных классов. В матрице информативностей INF1 показано количество

информации (в битах) содержащееся в факте наличия в изображении пикселя определенного спектрального диапазона о том, что изображение с ним относится к тому или иному классу.

14.6. Выбор наиболее достоверной модели и придание ей статуса текущей

Оценка достоверности моделей производится сразу после их создания в режиме 3.5 по F-критерию Ван Ризбергена, а также L1- и L2-мерам, предложенным проф.Е.В.Луценко [40] и преодолевающим некоторые недостатки F-меры: ее четкость, моноклассовость и зависимость от объема выборки.

Классическая количественная мера достоверности моделей: F-мера Ван Ризбергена основана на подсчете суммарного количества верно и ошибочно классифицированных и не классифицированных объектов обучающей выборки. В ***мультиклассовых*** системах классификации объект может одновременно относится ко многим классам. Соответственно, при синтезе модели его описание используется для формирования обобщенных образов многих классов, к которым он относится.

При использовании модели для классификации определяется степень сходства-различия объекта со всеми классами, причем истинно-положительным решением может являться принадлежность объекта сразу к нескольким классам. В результате такой классификации получается, что объект не просто правильно или ошибочно относится или не относится к различным классам, как в классической F-мере, но правильно или ошибочно относится или не относится к ним ***в различной степени***.

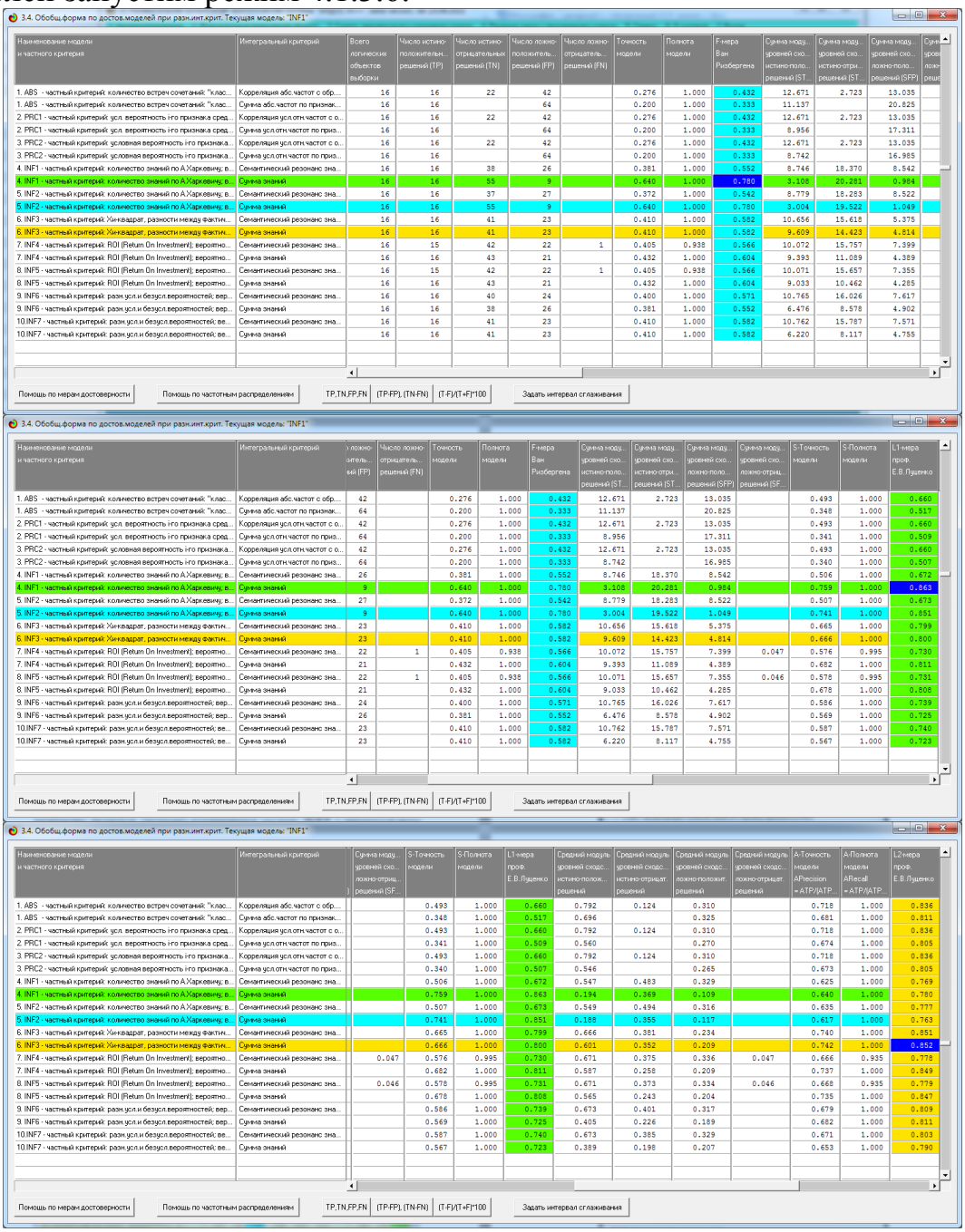
Однако классическая F-мера не учитывает того, что объект может, фактически, одновременно относится ко многим классам (мультиклассовость) и того, что в результате классификации может быть получена различная степень сходства-различия объекта с классами (нечеткость).

На численных примерах в работе [40] автором установлено, что при истинно-положительных и истинно-отрицательных решениях модуль сходства-различия объекта с классами значительно выше, чем при ложно-положительных и ложно-отрицательных решениях. Поэтому была предложена L1-мера достоверности моделей, учитывающая не просто сам факт истинно или ложно положительного или отрицательного решения, но и степень уверенности классификатора в этих решениях [40].

При классификации больших данных было обнаружено большое количество ложно-положительных решений с низким уровнем сходства, которые, однако, суммарно вносят большой вклад в снижение достоверности модели. Чтобы преодолеть эту проблему предложена L2-мера, в которой вместо сумм уровней сходства используется средние уровни сходства по различным вариантам классификации [40].

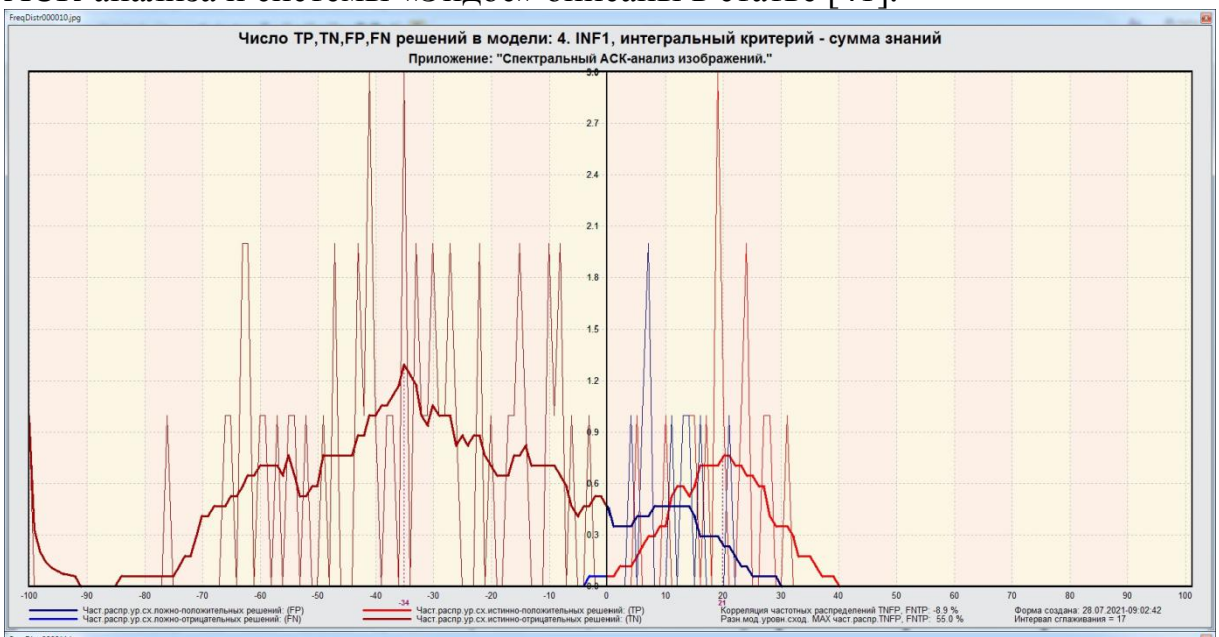
Таким образом, в работе [40] предлагаются меры достоверности моделей, названые L1-мера и L2-мера, смягчающие и преодолевающие недостатки F-меры, эти меры описаны математически и их применение продемонстрировано на простом численном примере. В интеллектуальной системе «Эйдос», которая является программным инструментарием автоматизированного системно-когнитивного анализа (АСК-анализ), реализованы все эти меры достоверности моделей: F, L1 и L2

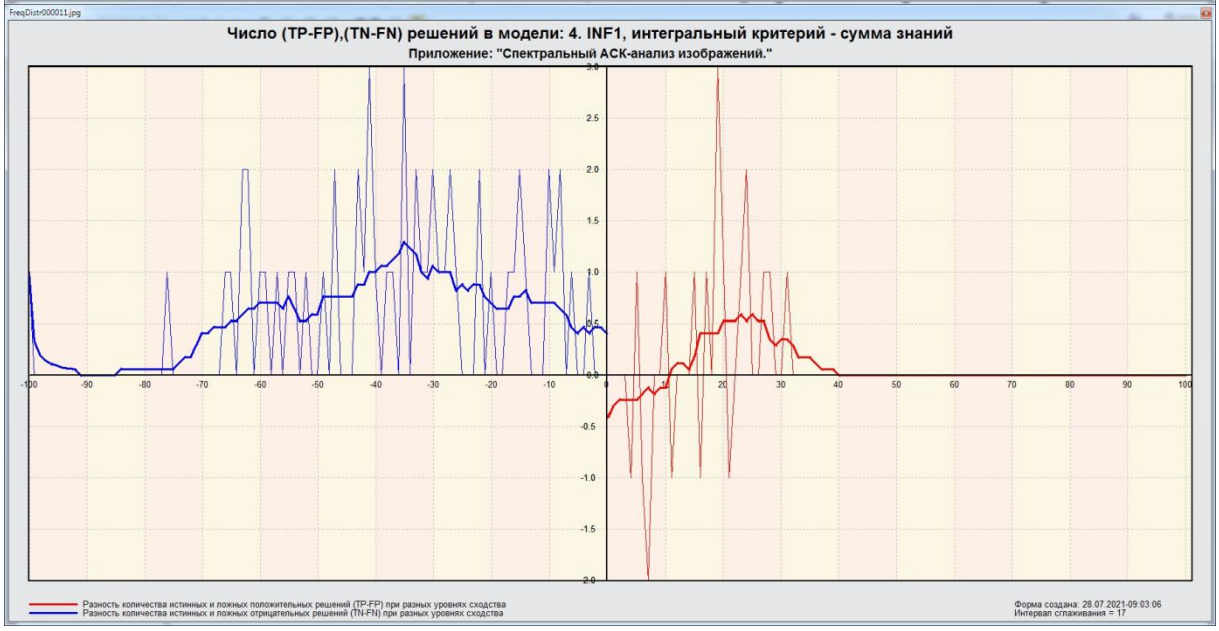
Для отображения формы с результатами оценки достоверности моделей запустим режим 4.1.3.6:

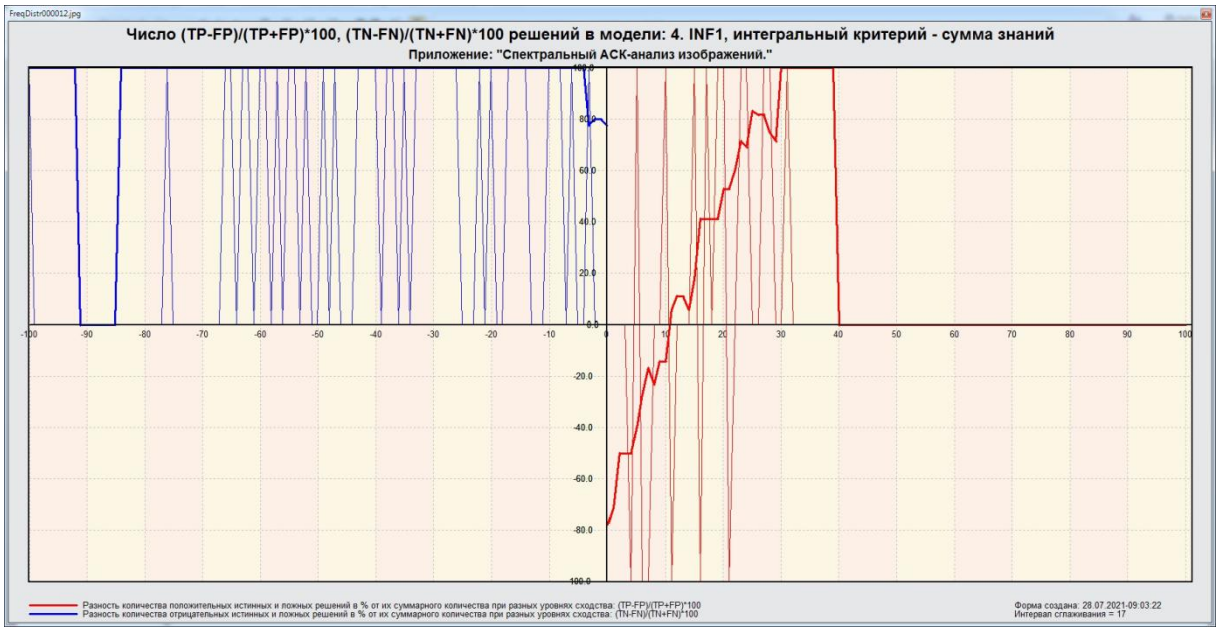


В этой экранной форме голубым цветом показана достоверность моделей по F-критерию Ван Ризбергена, зеленым по L1- мере, а желтым по L2-мере проф.Е.В.Луценко.

Из этих экранных форм видно, что наиболее достоверной по L1-критерию является системно-когнитивная модель INF1 с интегральным критерием «Сумма знаний». Значение L1-криетрия, равное 0,863, при теоретическом максимуме 1.000, является довольно хорошим результатом для моделируемой предметной области. Частные и интегральные критерии АСК-анализа и системы «Эйдос» описаны в статье [41].







Из приведенных выше экранных форм частотных распределений истинно- и ложно- положительных и отрицательных решений при различных уровнях сходства (режим 3.4) мы видим, что в данной модели при уровнях сходства:

–выше 22% наблюдаются только истинно-положительные решения (*это и есть критерий, позволяющий отделить гарантированно истинные решения от сомнительных и ложных при решении задач классификации и прогнозирования*);

–ниже 3% – только ложно-положительные решения;

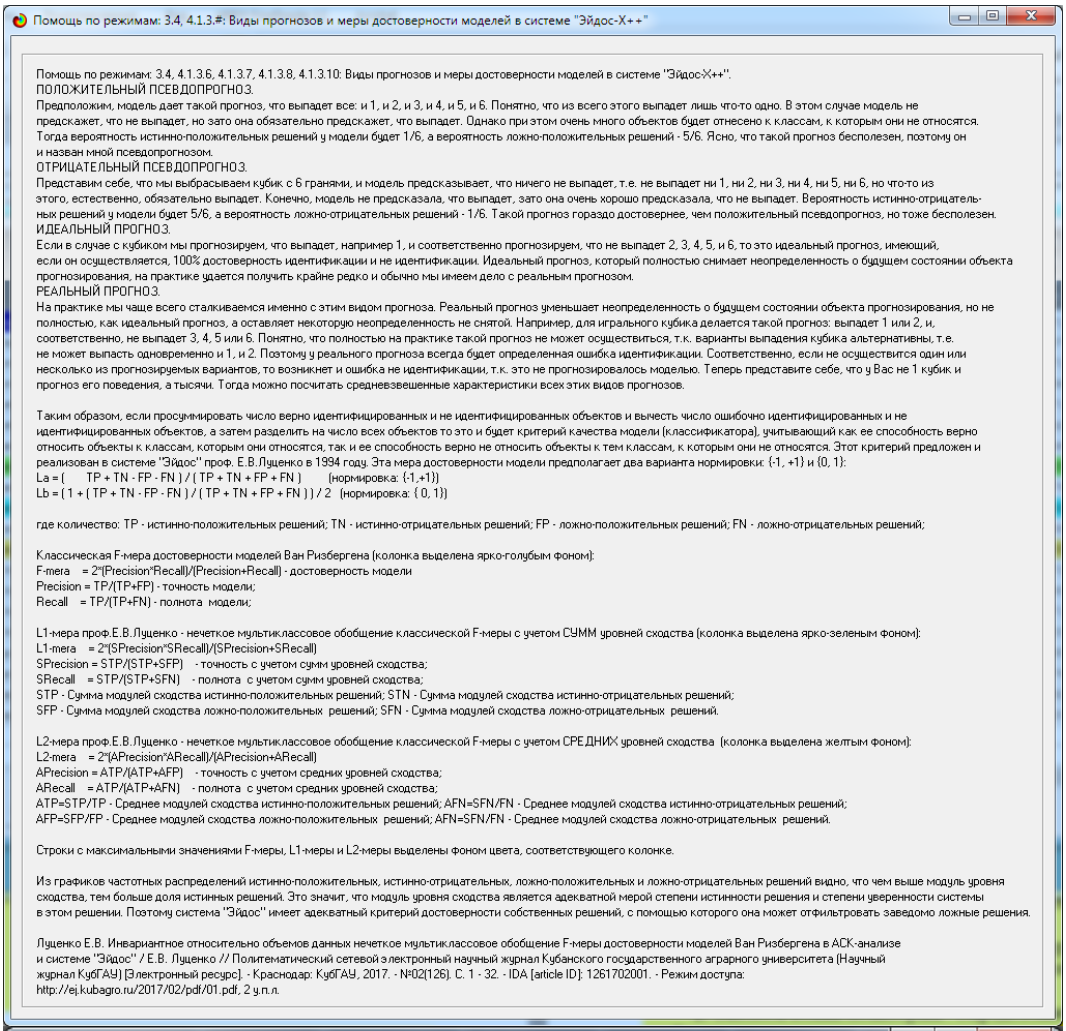
– в диапазоне от 3% до 22% встречаются и истинно-положительные, и ложно-положительные решения, причем *доля истинно-положительных решений растет с увеличением уровня сходства*[39].

Это означает, что уровень сходства конкретного объекта с обобщенным образом класса является адекватной мерой степени достоверности решения о принадлежности объекта к классу или степени его принадлежности к классу (в смысле нечеткой логики). Важно, что это внутренняя мера достоверности для системы «Эйдос», т.е. она не просто может принимать решения или прогнозировать, но адекватно оценивает степень истинности своих решений и прогнозов.

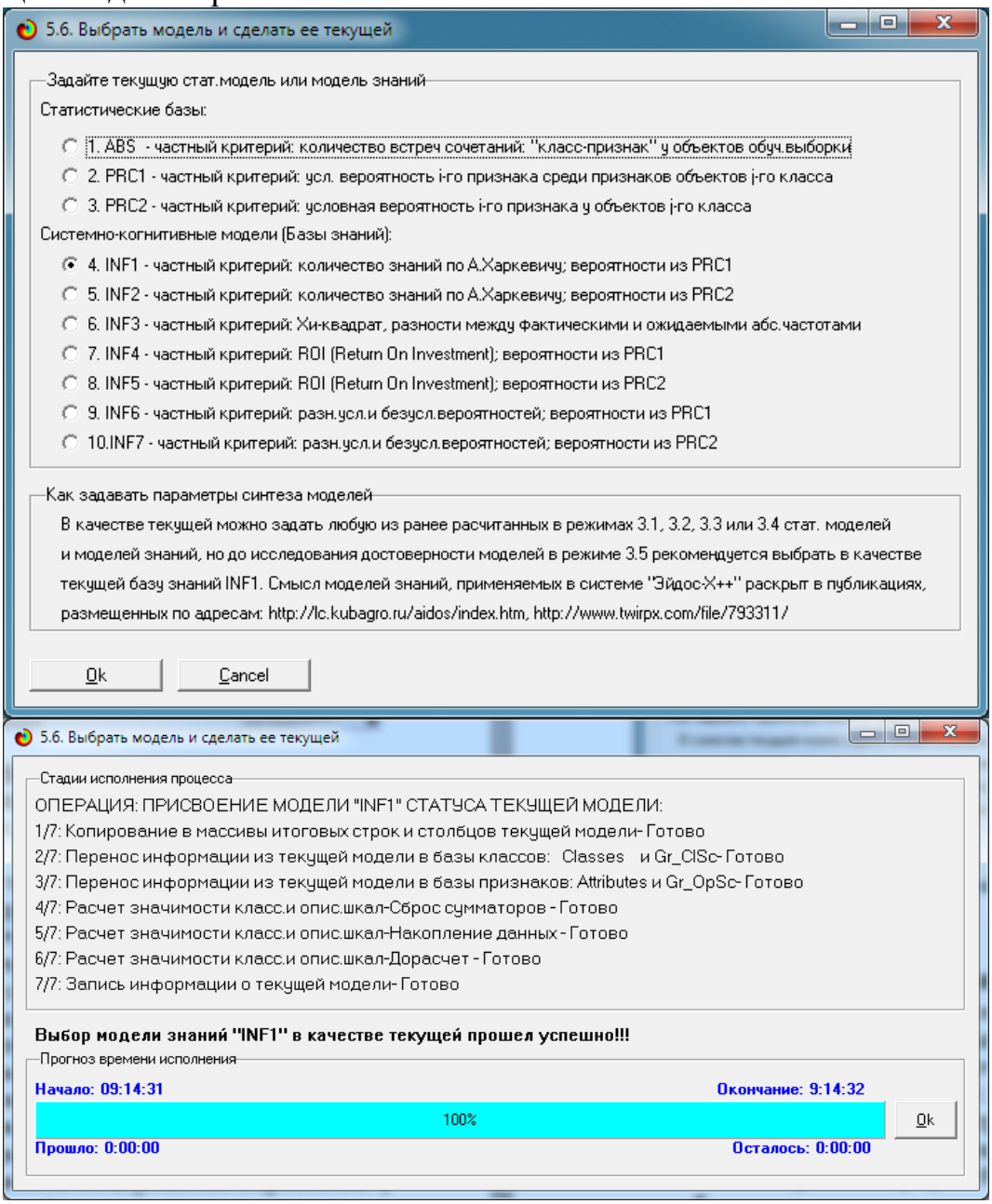
Ложно-отрицательных решений вообще не встречается, т.е. все отрицательные решения истинные.

Ниже приведены экранные формы помощи режима 3.4:

[39] Эта же закономерность, как правило, наблюдается и в интеллектуальных Эйдос-приложениях в других предметных областях.


Помощь по режимам: 3.4, 4.1.3.#: Виды прогнозов и меры достоверности моделей в системе "Эйдос-X++"
Помощь по режимам: 3.4, 4.1.3.6, 4.1.3.7, 4.1.3.8, 4.1.3.10: Виды прогнозов и меры достоверности моделей в системе "Эйдос-X++".
ПОЛОЖИТЕЛЬНЫЙ ПСЕВДОПРОГНОЗ.
Предположим, модель дает такой прогноз, что выпадет все: и 1, и 2, и 3, и 4, и 5, и 6. Понятно, что из всего этого выпадет лишь что-то одно. В этом случае модель не предскажет, что не выпадет, но зато она обязательно предскажет, что выпадет. Однако при этом очень много объектов будет отнесено к классам, к которым они не относятся. Тогда вероятность истинно-положительных решений у модели будет 1/6, а вероятность ложно-положительных решений - 5/6. Ясно, что такой прогноз бесполезен, поэтому он и назван мной псевдопрогнозом.
ОТРИЦАТЕЛЬНЫЙ ПСЕВДОПРОГНОЗ.
Представим себе, что мы выбрасываем кубик с 6 гранями, и модель предсказывает, что ничего не выпадет, т.е. не выпадет ни 1, ни 2, ни 3, ни 4, ни 5, ни 6, но что-то из этого, естественно, обязательно выпадет. Конечно, модель не предсказала, что выпадет, зато она очень хорошо предсказала, что не выпадет. Вероятность истинно-отрицательных решений у модели будет 5/6, а вероятность ложно-отрицательных решений - 1/6. Такой прогноз гораздо достовернее, чем положительный псевдопрогноз, но тоже бесполезен.
ИДЕАЛЬНЫЙ ПРОГНОЗ.
Если в случае с кубиком мы прогнозируем, что выпадет, например 1, и соответственно прогнозируем, что не выпадет 2, 3, 4, 5, и 6, то это идеальный прогноз, имеющий, если он осуществляется, 100% достоверность идентификации и не идентификации. Идеальный прогноз, который полностью снимает неопределенность о будущем состоянии объекта прогнозирования, на практике удается получить крайне редко и обычно мы имеем дело с реальным прогнозом.
РЕАЛЬНЫЙ ПРОГНОЗ.
На практике мы чаще всего сталкиваемся именно с этим видом прогноза. Реальный прогноз уменьшает неопределенность о будущем состоянии объекта прогнозирования, но не полностью, как идеальный прогноз, а оставляет некоторую неопределенность не снятой. Например, для игрального кубика делается такой прогноз: выпадет 1 или 2, и, соответственно, не выпадет 3, 4, 5 или 6. Понятно, что полностью на практике такой прогноз не может осуществиться, т.к. варианты выпадения кубика альтернативны, т.е. не может выпасть одновременно и 1, и 2. Поэтому у реального прогноза всегда будет определенная ошибка идентификации. Соответственно, если не осуществится один или несколько из прогнозируемых вариантов, то возникнет и ошибка не идентификации, т.к. это не прогнозировалось моделью. Теперь представите себе, что у Вас не 1 кубик и прогноз его поведения, а тысячи. Тогда можно посчитать средневзвешенные характеристики всех этих видов прогнозов.
Таким образом, если просуммировать число верно идентифицированных и не идентифицированных объектов и вычесть число ошибочно идентифицированных и не идентифицированных объектов, а затем разделить на число всех объектов то это и будет критерий качества модели (классификатора), учитывающий как ее способность верно относить объекты к классам, которым они относятся, так и ее способность верно не относить объекты к тем классам, к которым они не относятся. Этот критерий предложен и реализован в системе "Эйдос" проф. Е.В.Луценко в 1994 году. Эта мера достоверности модели предполагает два варианта нормировки: {-1, +1} и {0, 1}:
La = (TP + TN - FP - FN) / (TP + TN + FP + FN) (нормировка: {-1,+1})
Lb = (1 + (TP + TN - FP - FN) / (TP + TN + FP + FN)) / 2 (нормировка: { 0, 1})
где количество: TP - истинно-положительных решений; TN - истинно-отрицательных решений; FP - ложно-положительных решений; FN - ложно-отрицательных решений;
Классическая F-мера достоверности моделей Ван Ризбергена (колонка выделена ярко-голубым фоном):
F-mera = 2*(Precision*Recall)/(Precision+Recall) - достоверность модели
Precision = TP/(TP+FP) - точность модели;
Recall = TP/(TP+FN) - полнота модели;
L1-мера проф.Е.В.Луценко - нечеткое мультиклассовое обобщение классической F-меры с учетом СУММ уровней сходства (колонка выделена ярко-зеленым фоном):
L1-mera = 2*(SPrecision*SRecall)/(SPrecision+SRecall)
SPrecision = STP/(STP+SFP) - точность с учетом сумм уровней сходства;
SRecall = STP/(STP+SFN) - полнота с учетом сумм уровней сходства;
STP - Сумма модулей сходства истинно-положительных решений; STN - Сумма модулей сходства истинно-отрицательных решений;
SFP - Сумма модулей сходства ложно-положительных решений; SFN - Сумма модулей сходства ложно-отрицательных решений.
L2-мера проф.Е.В.Луценко - нечеткое мультиклассовое обобщение классической F-меры с учетом СРЕДНИХ уровней сходства (колонка выделена желтым фоном):
L2-mera = 2*(APrecision*ARecall)/(APrecision+ARecall)
APrecision = ATP/(ATP+AFP) - точность с учетом средних уровней сходства;
ARecall = ATP/(ATP+AFN) - полнота с учетом средних уровней сходства;
ATP=STP/TP - Среднее модулей сходства истинно-положительных решений; AFN=SFN/FN - Среднее модулей сходства истинно-отрицательных решений;
AFP=SFP/FP - Среднее модулей сходства ложно-положительных решений; AFN=SFN/FN - Среднее модулей сходства ложно-отрицательных решений.
Строки с максимальными значениями F-меры, L1-меры и L2-меры выделены фоном цвета, соответствующего колонке.
Из графиков частотных распределений истинно-положительных, истинно-отрицательных, ложно-положительных и ложно-отрицательных решений видно, что чем выше модуль уровня сходства, тем больше доля истинных решений. Это значит, что модуль уровня сходства является адекватной мерой степени истинности решения и степени уверенности системы в этом решении. Поэтому система "Эйдос" имеет адекватный критерий достоверности собственных решений, с помощью которого она может отфильтровать заведомо ложные решения.
Луценко Е.В. Инвариантное относительно объемов данных нечеткое мультиклассовое обобщение F-меры достоверности моделей Ван Ризбергена в АСК-анализе и системе "Эйдос" / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. - Краснодар: КубГАУ, 2017. - №02(126). С. 1 - 32. - IDA [article ID]: 1261702001. - Режим доступа: http://ej.kubagro.ru/2017/02/pdf/01.pdf, 2 у.п.л.
Помощь по режиму 4.1.3.11. (С) Система "ЭЙДОС-X++"
Режим: 4.1.3.11. РАСЧЕТ И ГРАФИЧЕСКАЯ ВИЗУАЛИЗАЦИЯ ЧАСТОТНЫХ РАСПРЕДЕЛЕНИЙ УРОВНЕЙ СХОДСТВА:
- TP,TN,FP,FN, интегральный критерий - резонанс знаний;
- TP,TN,FP,FN, интегральный критерий - сумма знаний;
- (TP-FP), (TN-FN), интегральный критерий - резонанс знаний;
- (TP-FP), (TN-FN), интегральный критерий - сумма знаний;
- (TP-FP)/(TP+FP)*100 и (TN-FN)/(TN+FN)*100, интегральный критерий - резонанс знаний;
- (TP-FP)/(TP+FP)*100 и (TN-FN)/(TN+FN)*100, интегральный критерий - сумма знаний;
где:
TP - истинно-положительное решение;
TN - истинно-отрицательное решение;
FP - ложно-положительное решение;
FN - ложно-отрицательное решение.
Отображается график того частотного распределения, на котором в экранной форме стоит курсор.
Луценко Е.В. Инвариантное относительно объемов данных нечеткое мультиклассовое обобщение F-меры достоверности моделей Ван Ризбергена в АСК-анализе и системе "Эйдос" / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. - Краснодар: КубГАУ, 2017. - №02(126). С. 1 - 32. -IDA [article ID]: 1261702001. -Режим доступа: http://ej.kubagro.ru/2017/02/pdf/01.pdf, 2 у.п.л.

Выберем наиболее достоверную модель и придадим ей статус текущей модели в режиме 5.6:

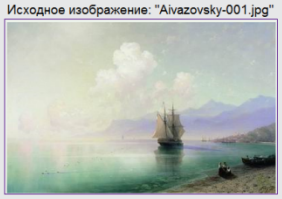


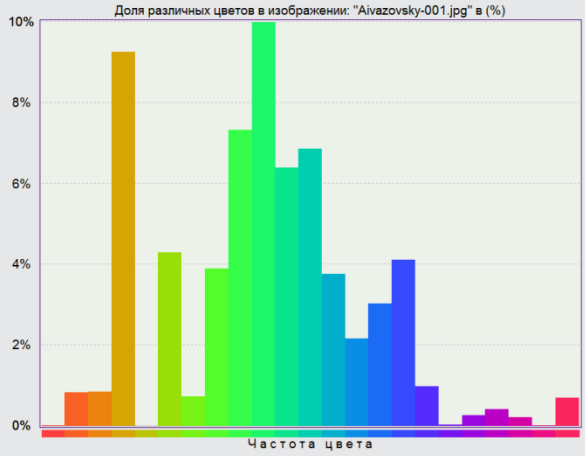
14.7. Спектры конкретных изображений

Для вывода спектров конкретных изображений (картин художников) создадим модель как описано выше на основе файлов, в именах которых вместо тире используется нижнее подчеркивание. Выполним режимы 4.7 (подготовка данных), 2.3.2.3 и 3.5. Затем войдем в режим «4.7. АСК-анализ изображений по пикселям, спектрам и контурам» и кликнем по кнопке: «Изображения и спектры объектов». В результате получим в папке: c:\Aidos-X\AID_DATA\InpSpectrPix\ следующие изображения, которые содержат всю необходимую для их интерпретации информацию:

ИСХОДНОЕ ИЗОБРАЖЕНИЕ: 1/16-"Aivazovsky-001.jpg" И ЕГО СПЕКТР
Приложение: "Спектральный АСК-анализ изображений"

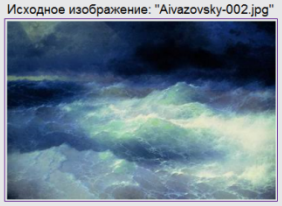
Исходное изображение: "Aivazovsky-001.jpg"

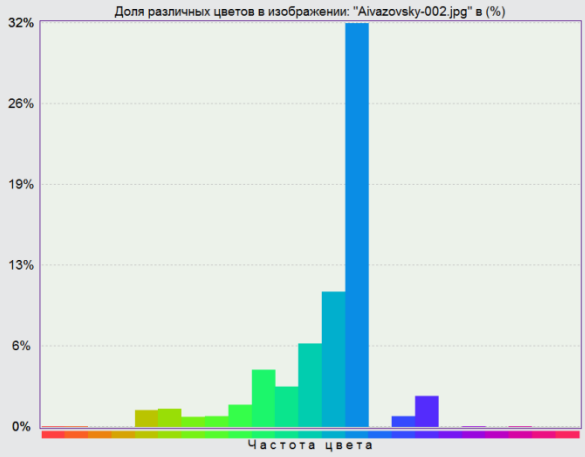




ИСХОДНОЕ ИЗОБРАЖЕНИЕ: 2/16-"Aivazovsky-002.jpg" И ЕГО СПЕКТР
Приложение: "Спектральный АСК-анализ изображений"

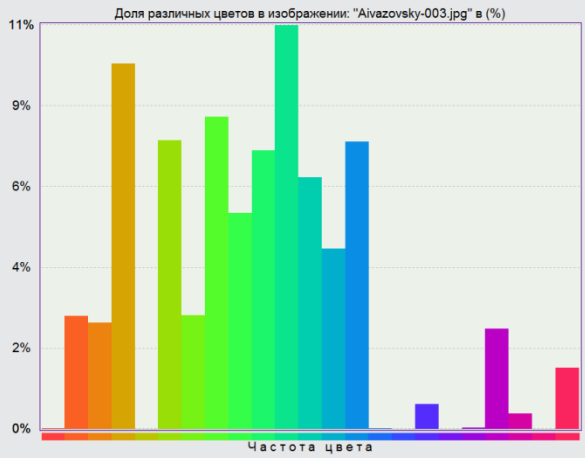
Исходное изображение: "Aivazovsky-002.jpg"





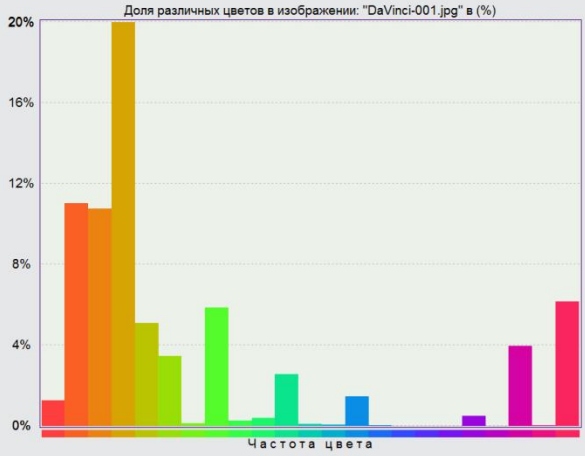
ИСХОДНОЕ ИЗОБРАЖЕНИЕ: 3/16-"Aivazovsky-003.jpg" И ЕГО СПЕКТР
Приложение: "Спектральный АСК-анализ изображений"

Исходное изображение: "Aivazovsky-003.jpg"



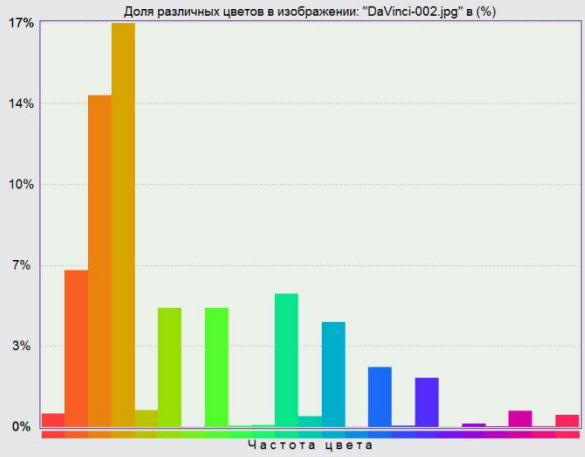
ИСХОДНОЕ ИЗОБРАЖЕНИЕ: 4/16-"DaVinci-001.jpg" И ЕГО СПЕКТР
Приложение: "Спектральный АСК-анализ изображений"

Исходное изображение: "DaVinci-001.jpg"



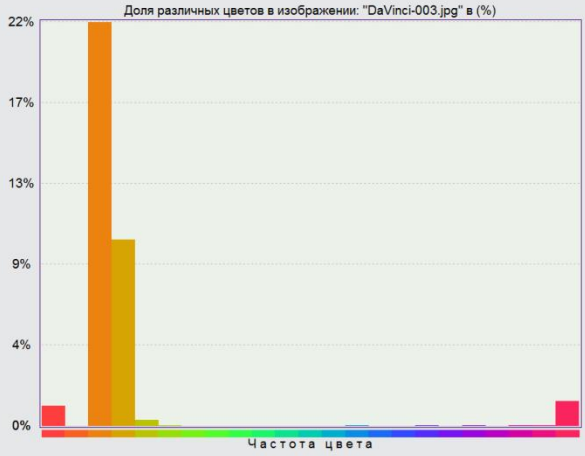
ИСХОДНОЕ ИЗОБРАЖЕНИЕ: 5/16-"DaVinci-002.jpg" И ЕГО СПЕКТР
Приложение: "Спектральный АСК-анализ изображений"

Исходное изображение: "DaVinci-002.jpg"



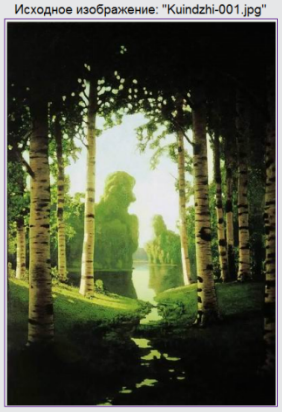
ИСХОДНОЕ ИЗОБРАЖЕНИЕ: 6/16-"DaVinci-003.jpg" И ЕГО СПЕКТР
Приложение: "Спектральный АСК-анализ изображений"

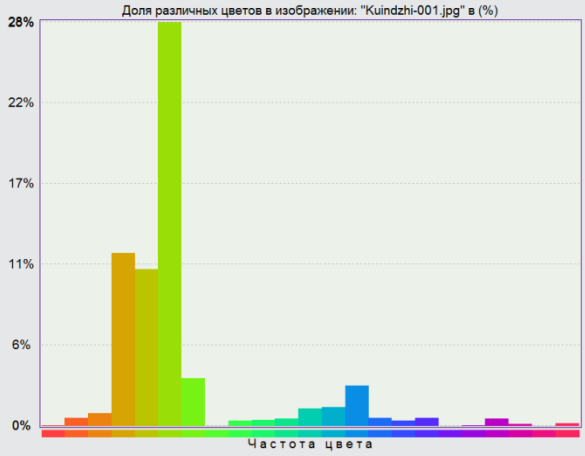
Исходное изображение: "DaVinci-003.jpg"



ИСХОДНОЕ ИЗОБРАЖЕНИЕ: 7/16-"Kuindzhi-001.jpg" И ЕГО СПЕКТР
Приложение: "Спектральный АСК-анализ изображений"

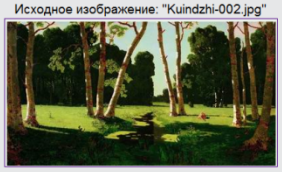
Исходное изображение: "Kuindzhi-001.jpg"

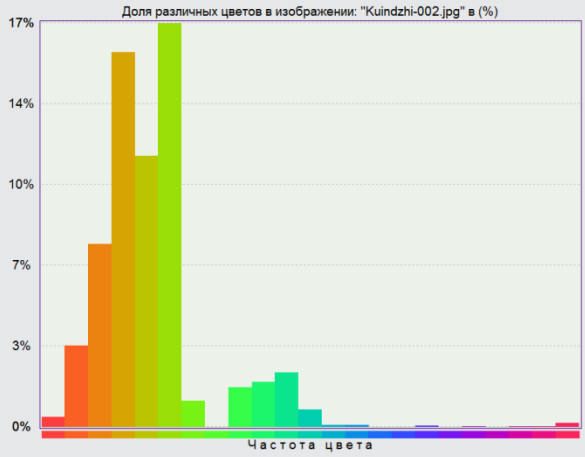




ИСХОДНОЕ ИЗОБРАЖЕНИЕ: 8/16-"Kuindzhi-002.jpg" И ЕГО СПЕКТР
Приложение: "Спектральный АСК-анализ изображений"

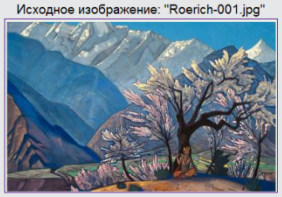
Исходное изображение: "Kuindzhi-002.jpg"

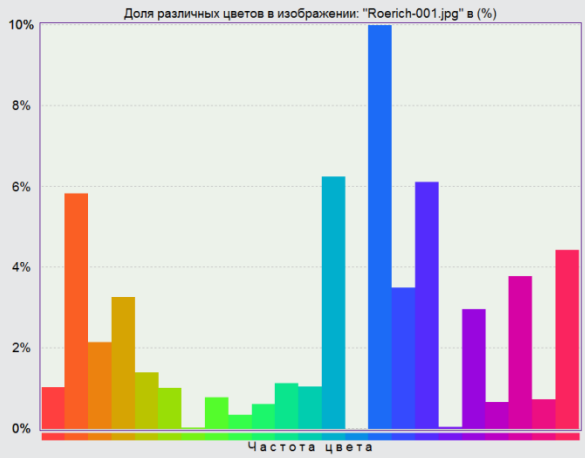




ИСХОДНОЕ ИЗОБРАЖЕНИЕ: 9/16-"Roerich-001.jpg" И ЕГО СПЕКТР
Приложение: "Спектральный АСК-анализ изображений"

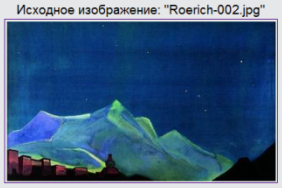
Исходное изображение: "Roerich-001.jpg"

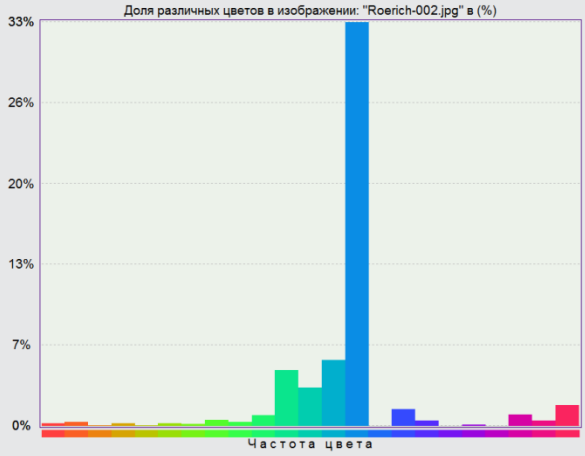




ИСХОДНОЕ ИЗОБРАЖЕНИЕ: 10/16-"Roerich-002.jpg" И ЕГО СПЕКТР
Приложение: "Спектральный АСК-анализ изображений"

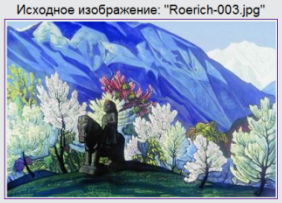
Исходное изображение: "Roerich-002.jpg"

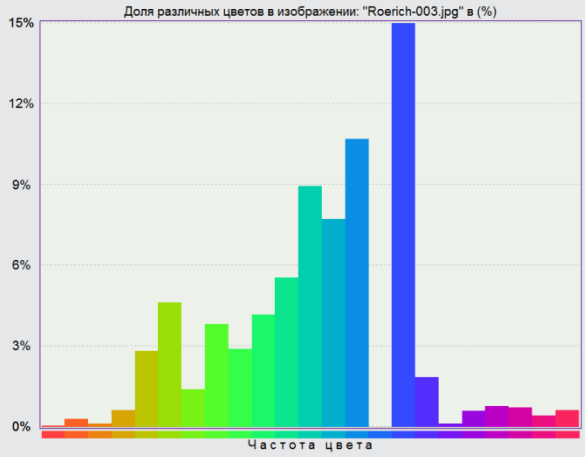




ИСХОДНОЕ ИЗОБРАЖЕНИЕ: 11/16-"Roerich-003.jpg" И ЕГО СПЕКТР
Приложение: "Спектральный АСК-анализ изображений"

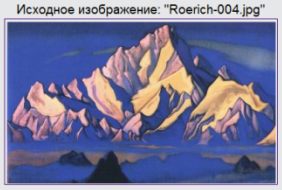
Исходное изображение: "Roerich-003.jpg"

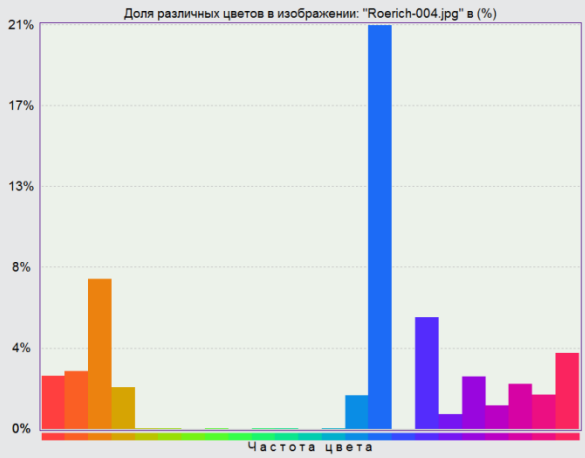




ИСХОДНОЕ ИЗОБРАЖЕНИЕ: 12/16-"Roerich-004.jpg" И ЕГО СПЕКТР
Приложение: "Спектральный АСК-анализ изображений"

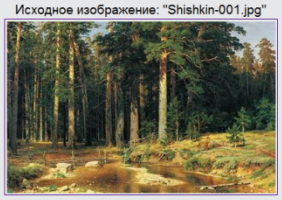
Исходное изображение: "Roerich-004.jpg"

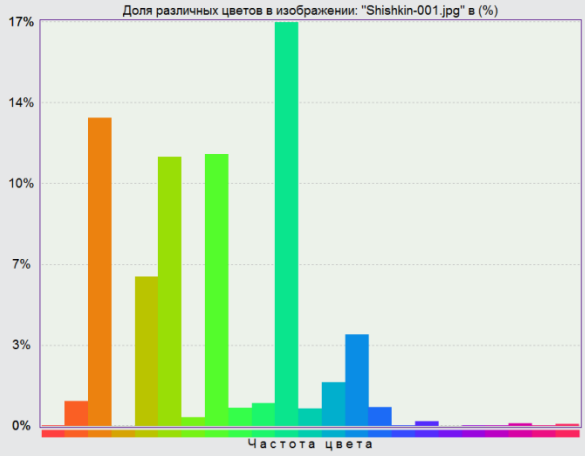




ИСХОДНОЕ ИЗОБРАЖЕНИЕ: 13/16-"Shishkin-001.jpg" И ЕГО СПЕКТР
Приложение: "Спектральный АСК-анализ изображений"

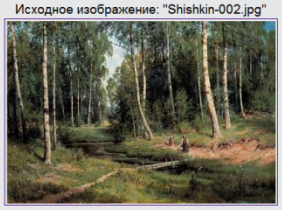
Исходное изображение: "Shishkin-001.jpg"

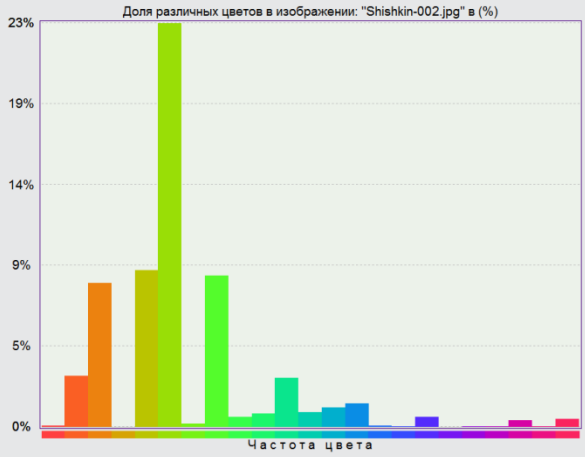




ИСХОДНОЕ ИЗОБРАЖЕНИЕ: 14/16-"Shishkin-002.jpg" И ЕГО СПЕКТР
Приложение: "Спектральный АСК-анализ изображений"

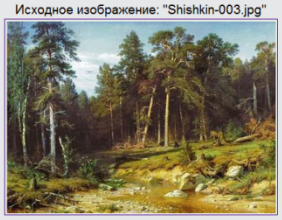
Исходное изображение: "Shishkin-002.jpg"

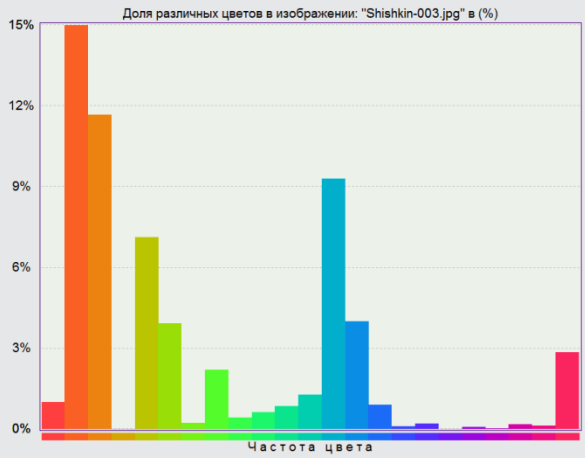


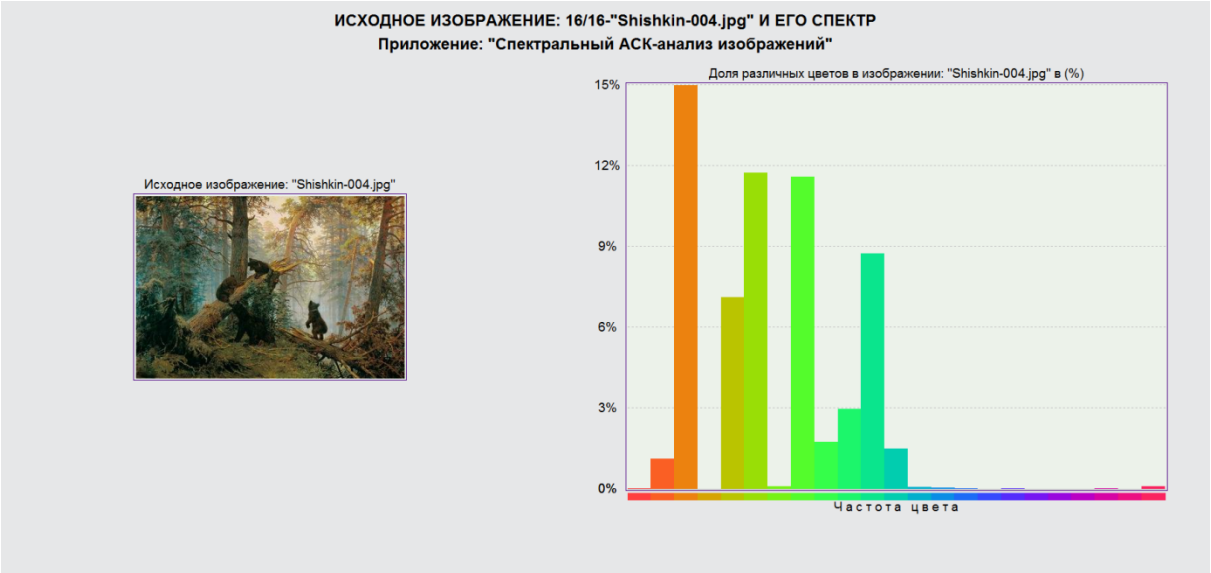


ИСХОДНОЕ ИЗОБРАЖЕНИЕ: 15/16-"Shishkin-003.jpg" И ЕГО СПЕКТР
Приложение: "Спектральный АСК-анализ изображений"

Исходное изображение: "Shishkin-003.jpg"

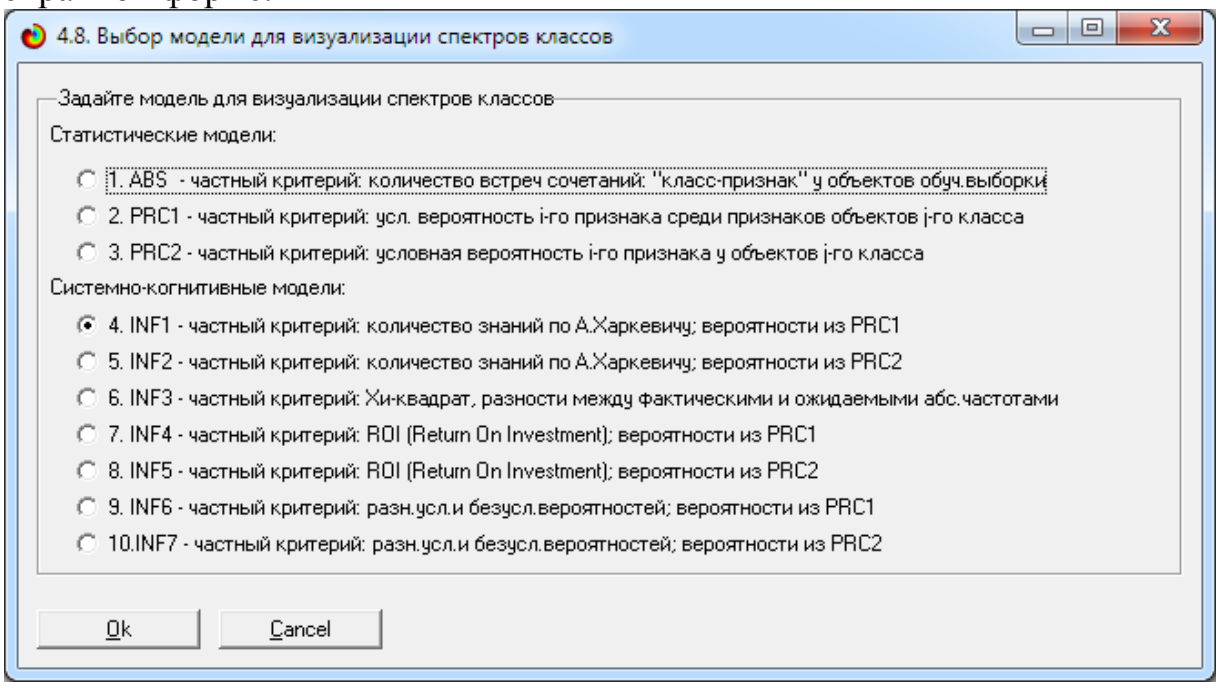






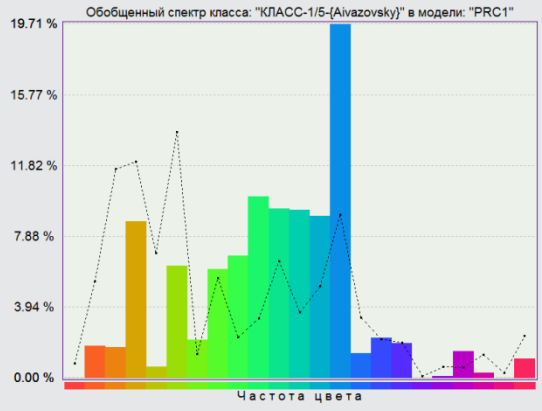
14.8. Спектры обобщенных изображений классов

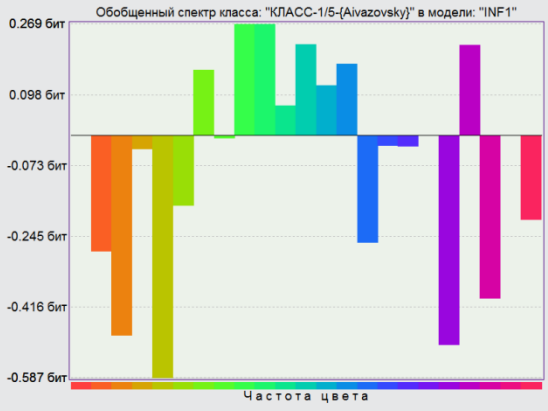
Войдем в режим «4.7. АСК-анализ изображений по пикселям, спектрам и контурам» и кликнем по кнопке: «Изображения и спектры классов». Выберем наиболее достоверную модель INF1 на появившейся экранной форме:



В результате получим в папке: c:\Aidos-X\AID_DATA\InpSpectrCls\ следующие изображения:

ОБОБЩЕННЫЕ СПЕКТРЫ КЛАССА: "КЛАСС-1/5-{Aivazovsky}" В МОДЕЛЯХ: "PRC1" И "INF1"
Приложение: "Спектральный АСК-анализ изображений."

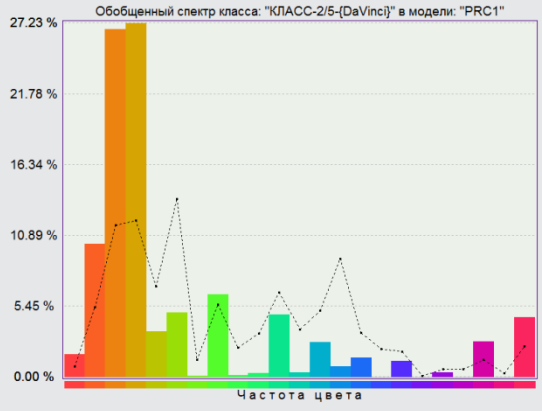


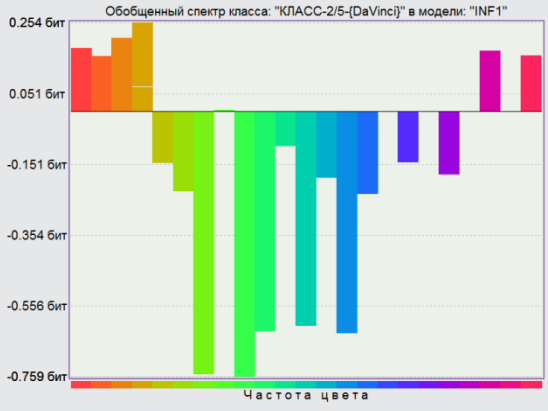


Гистограмма отражает условную вероятность встречи каждого цвета в изображениях данного класса.
Пунктирная линия отражает безусловную (среднюю) вероятность встречи цвета по всей выборке изображений.
2. PRC1 - частный критерий: условная вероятность i-го признака среди признаков объектов j-го класса

Гистограмма отражает степень характерности/нехарактерности каждого цвета для изображений данного класса, т.е. степень отличия его доли в изображении от среднего по выборке, а знак зависит от того, больше она или меньше.
4. INF1 - частный критерий: количество знаний по А.Харкевичу; вероятности из PRC1

ОБОБЩЕННЫЕ СПЕКТРЫ КЛАССА: "КЛАСС-2/5-{DaVinci}" В МОДЕЛЯХ: "PRC1" И "INF1"
Приложение: "Спектральный АСК-анализ изображений."

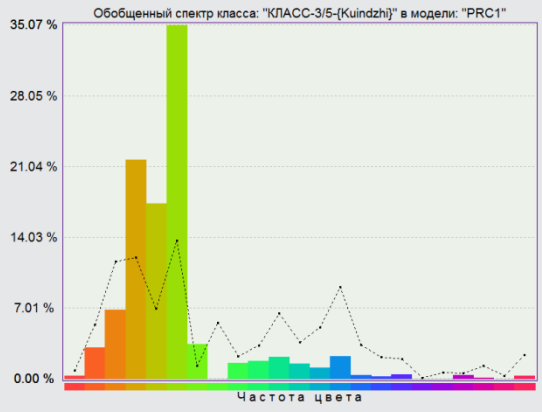


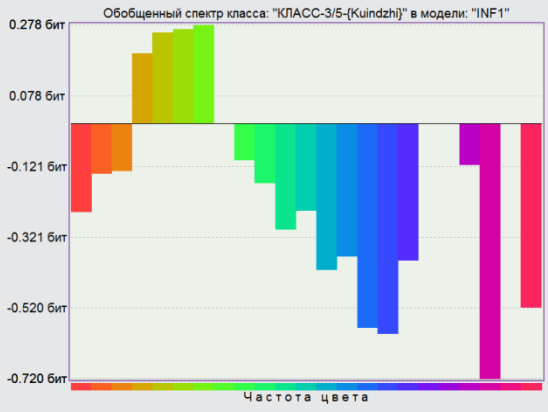


Гистограмма отражает условную вероятность встречи каждого цвета в изображениях данного класса.
Пунктирная линия отражает безусловную (среднюю) вероятность встречи цвета по всей выборке изображений.
2. PRC1 - частный критерий: условная вероятность i-го признака среди признаков объектов j-го класса

Гистограмма отражает степень характерности/нехарактерности каждого цвета для изображений данного класса, т.е. степень отличия его доли в изображении от среднего по выборке, а знак зависит от того, больше она или меньше.
4. INF1 - частный критерий: количество знаний по А.Харкевичу; вероятности из PRC1

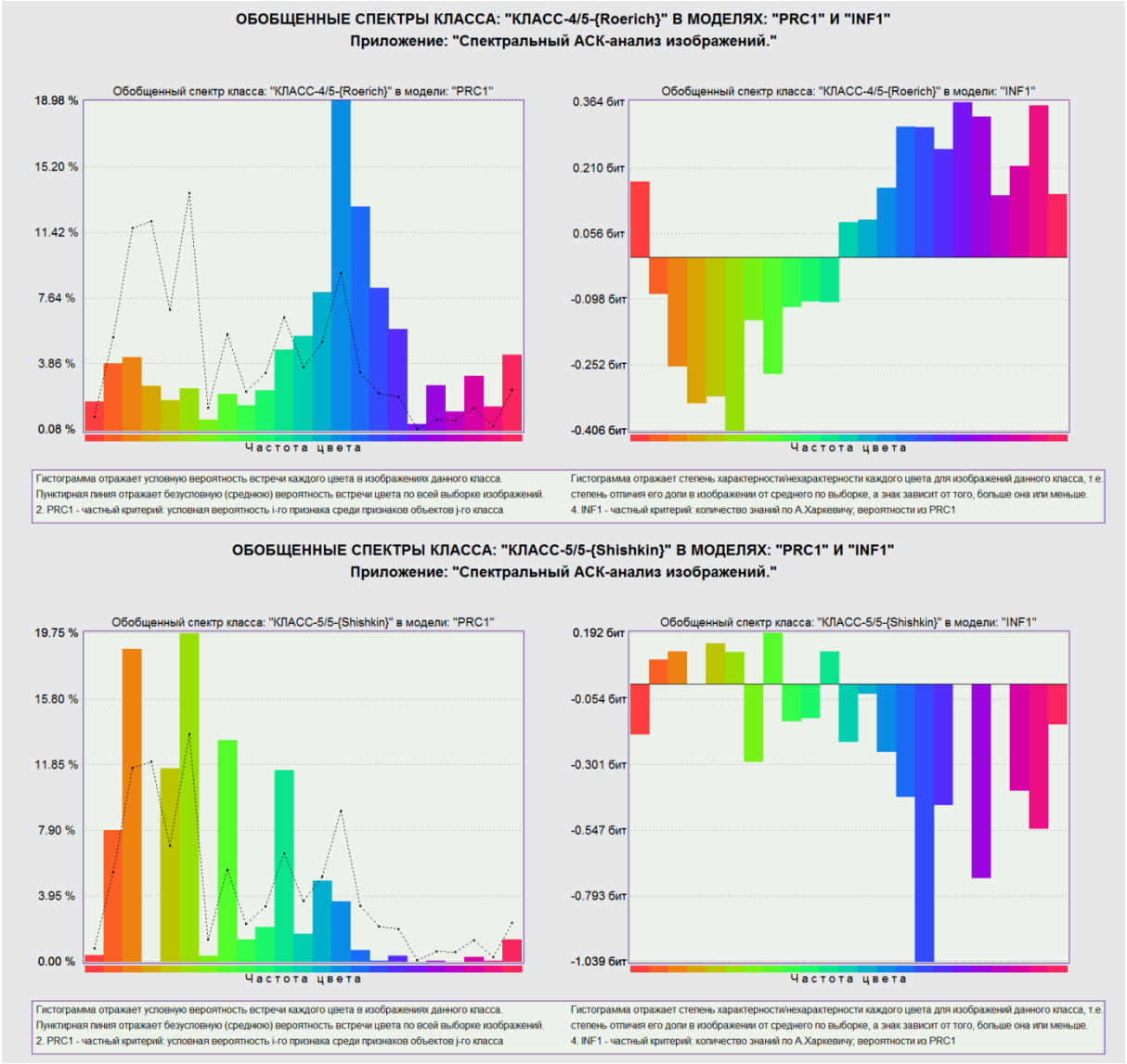
ОБОБЩЕННЫЕ СПЕКТРЫ КЛАССА: "КЛАСС-3/5-{Kuindzhi}" В МОДЕЛЯХ: "PRC1" И "INF1"
Приложение: "Спектральный АСК-анализ изображений."





Гистограмма отражает условную вероятность встречи каждого цвета в изображениях данного класса.
Пунктирная линия отражает безусловную (среднюю) вероятность встречи цвета по всей выборке изображений.
2. PRC1 - частный критерий: условная вероятность i-го признака среди признаков объектов j-го класса

Гистограмма отражает степень характерности/нехарактерности каждого цвета для изображений данного класса, т.е. степень отличия его доли в изображении от среднего по выборке, а знак зависит от того, больше она или меньше.
4. INF1 - частный критерий: количество знаний по А.Харкевичу; вероятности из PRC1



Слева на этих изображениях показан спектр класса в статистической модели PRC1, т.е. ***условная вероятность*** встретить пиксели каждого спектрального диапазона в изображениях данного класса (в процентах от суммарного числа пикселей в изображении). Кроме того слева пунктирной линией показана ***безусловная вероятность*** встречи пикселей каждого спектрального диапазона во всех изображениях обучающей выборки, т.е. по всем классам.

Справа показано ***количество информации*** в цвете каждого спектрального диапазона о принадлежности объекта с пикселями этого цвета к данному классу.

Если в данном классе условная вероятность встретить пиксели этого цвета выше, чем безусловная вероятность его встречи в среднем по всей выборке, то данный спектральный диапазон является характерным для данного класса, если ниже, чем по всей выборке – то нехарактерным, если же условная вероятность встречи пикселей данного спектрального диапазона близка к безусловной (средней по всей

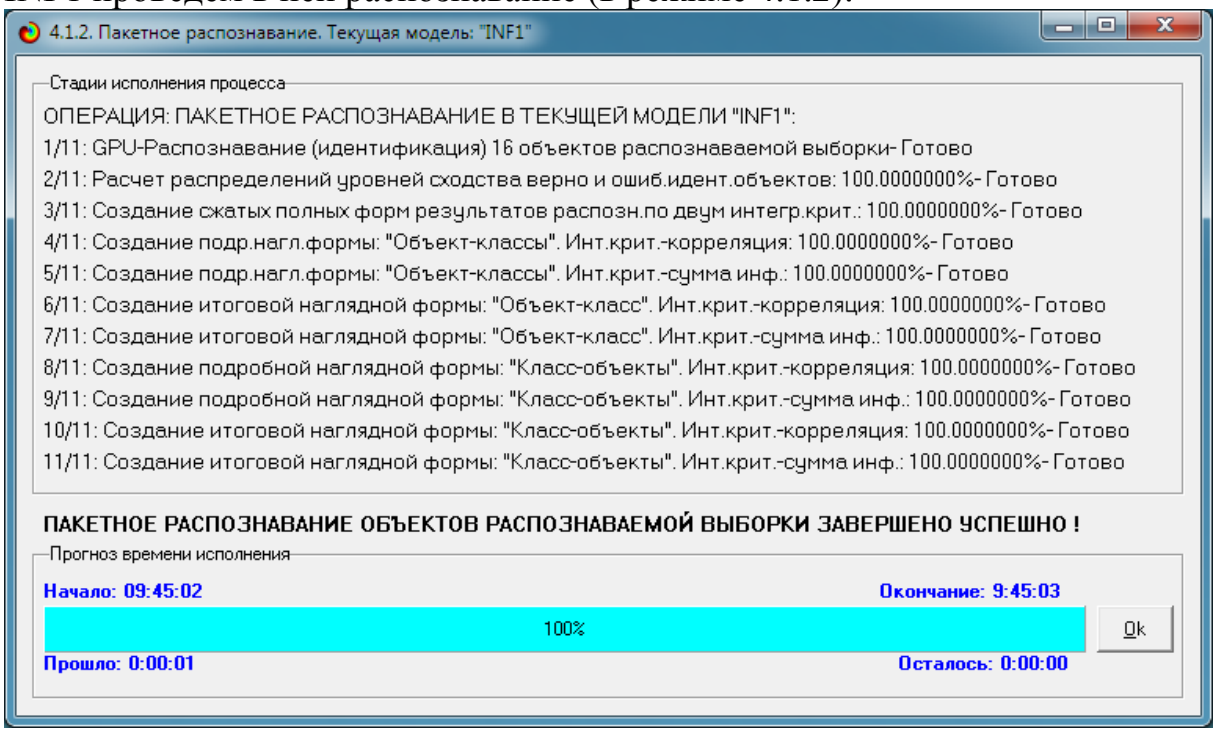
выборке), то обнаружение данного цвета у конкретного изображения практически ничего не говорит о принадлежности или непринадлежности данного изображения к этому классу.

Таким образом, ***обобщенный спектр класса не просто отражает условную вероятность встречи различных цветов у изображений данного класса, но содержит в себе информацию о результатах сравнения условных вероятностей встречи этих цветов в данном классе с безусловными вероятностями их встречи по всей выборке изображений, принадлежащих ко всем классам. При этом в настоящее время в системе «Эйдос» используется 7 системно-когнитивных моделей, отличающихся способами расчета и сравнения условных и безусловных вероятностей [41] и два инегральных критерия для определения степени сходства конкретного образа объекта с обобщенным образом класса.***

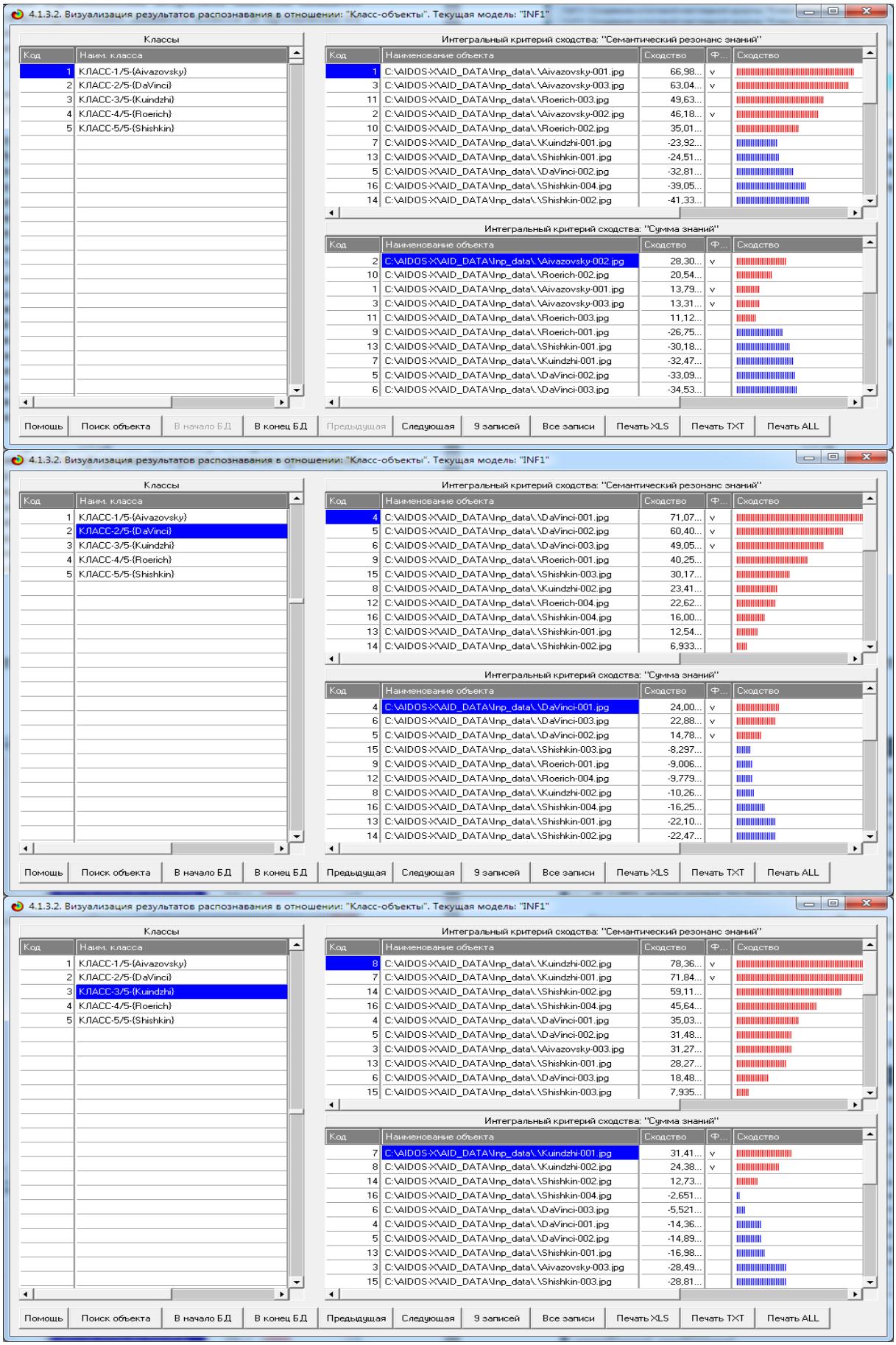
14.9. Решение задач в наиболее достоверной модели

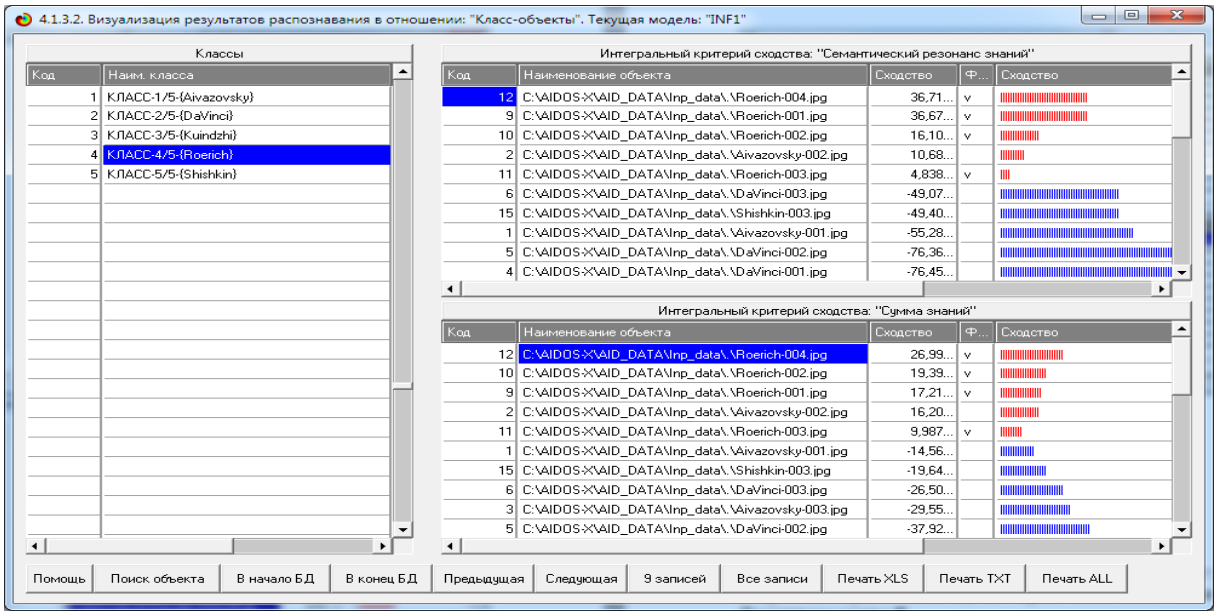
14.9.1. Решение задачи сравнения конкретных изображений с обобщенными образами классов

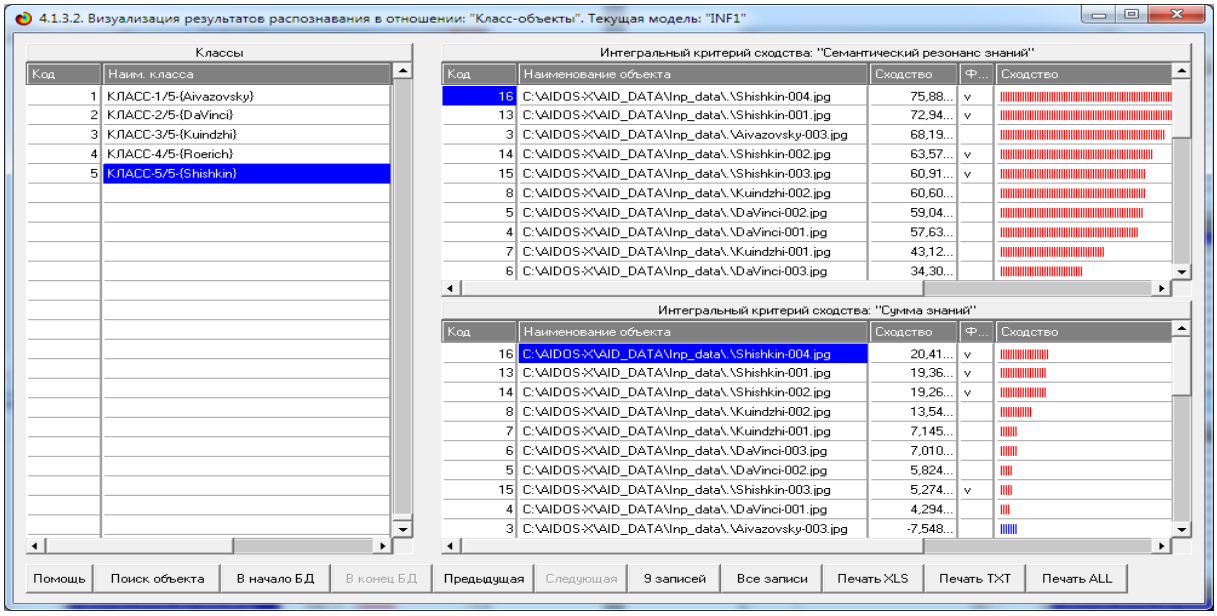
После того, как мы сделали текущей наиболее достоверную модель INF1 проведем в ней распознавание (в режиме 4.1.2):



В данном случае распознавание представляет собой сравнение спекторов конкретных картин художников с обобщенными спектрами образов классов, каждый из которых сформирован на основе всех картин соотвествующего художника. Результаты этого сравнения приведены ниже:


4.1.3.2. Визуализация результатов распознавания в отношении: "Класс-объекты". Текущая модель: "INF1"
Классы
Код
Наим. класса
1 КЛАСС-1/5-{Aivazovsky}
2 КЛАСС-2/5-{DaVinci}
3 КЛАСС-3/5-{Kuindzhi}
4 КЛАСС-4/5-{Roerich}
5 КЛАСС-5/5-{Shishkin}
Интегральный критерий сходства: "Семантический резонанс знаний"
Код
Наименование объекта
Сходство
Ф...
Сходство
1 C:\AIDOS-X\AID_DATA\Inp_data\.\Aivazovsky-001.jpg 66,98... v
3 C:\AIDOS-X\AID_DATA\Inp_data\.\Aivazovsky-003.jpg 63,04... v
11 C:\AIDOS-X\AID_DATA\Inp_data\.\Roerich-003.jpg 49,63...
2 C:\AIDOS-X\AID_DATA\Inp_data\.\Aivazovsky-002.jpg 46,18... v
10 C:\AIDOS-X\AID_DATA\Inp_data\.\Roerich-002.jpg 35,01...
7 C:\AIDOS-X\AID_DATA\Inp_data\.\Kuindzhi-001.jpg -23,92...
13 C:\AIDOS-X\AID_DATA\Inp_data\.\Shishkin-001.jpg -24,51...
5 C:\AIDOS-X\AID_DATA\Inp_data\.\DaVinci-002.jpg -32,81...
16 C:\AIDOS-X\AID_DATA\Inp_data\.\Shishkin-004.jpg -39,05...
14 C:\AIDOS-X\AID_DATA\Inp_data\.\Shishkin-002.jpg -41,33...
Интегральный критерий сходства: "Сумма знаний"
Код
Наименование объекта
Сходство
Ф...
Сходство
2 C:\AIDOS-X\AID_DATA\Inp_data\.\Aivazovsky-002.jpg 28,30... v
10 C:\AIDOS-X\AID_DATA\Inp_data\.\Roerich-002.jpg 20,54...
1 C:\AIDOS-X\AID_DATA\Inp_data\.\Aivazovsky-001.jpg 13,79... v
3 C:\AIDOS-X\AID_DATA\Inp_data\.\Aivazovsky-003.jpg 13,31... v
11 C:\AIDOS-X\AID_DATA\Inp_data\.\Roerich-003.jpg 11,12...
9 C:\AIDOS-X\AID_DATA\Inp_data\.\Roerich-001.jpg -26,75...
13 C:\AIDOS-X\AID_DATA\Inp_data\.\Shishkin-001.jpg -30,18...
7 C:\AIDOS-X\AID_DATA\Inp_data\.\Kuindzhi-001.jpg -32,47...
5 C:\AIDOS-X\AID_DATA\Inp_data\.\DaVinci-002.jpg -33,09...
6 C:\AIDOS-X\AID_DATA\Inp_data\.\DaVinci-003.jpg -34,53...
Помощь
Поиск объекта
В начало БД
В конец БД
Предыдущая
Следующая
9 записей
Все записи
Печать XLS
Печать TXT
Печать ALL
4.1.3.2. Визуализация результатов распознавания в отношении: "Класс-объекты". Текущая модель: "INF1"
Классы
Код
Наим. класса
1 КЛАСС-1/5-{Aivazovsky}
2 КЛАСС-2/5-{DaVinci}
3 КЛАСС-3/5-{Kuindzhi}
4 КЛАСС-4/5-{Roerich}
5 КЛАСС-5/5-{Shishkin}
Интегральный критерий сходства: "Семантический резонанс знаний"
Код
Наименование объекта
Сходство
Ф...
Сходство
4 C:\AIDOS-X\AID_DATA\Inp_data\.\DaVinci-001.jpg 71,07... v
5 C:\AIDOS-X\AID_DATA\Inp_data\.\DaVinci-002.jpg 60,40... v
6 C:\AIDOS-X\AID_DATA\Inp_data\.\DaVinci-003.jpg 49,05... v
9 C:\AIDOS-X\AID_DATA\Inp_data\.\Roerich-001.jpg 40,25...
15 C:\AIDOS-X\AID_DATA\Inp_data\.\Shishkin-003.jpg 30,17...
8 C:\AIDOS-X\AID_DATA\Inp_data\.\Kuindzhi-002.jpg 23,41...
12 C:\AIDOS-X\AID_DATA\Inp_data\.\Roerich-004.jpg 22,62...
16 C:\AIDOS-X\AID_DATA\Inp_data\.\Shishkin-004.jpg 16,00...
13 C:\AIDOS-X\AID_DATA\Inp_data\.\Shishkin-001.jpg 12,54...
14 C:\AIDOS-X\AID_DATA\Inp_data\.\Shishkin-002.jpg 6,933...
Интегральный критерий сходства: "Сумма знаний"
Код
Наименование объекта
Сходство
Ф...
Сходство
4 C:\AIDOS-X\AID_DATA\Inp_data\.\DaVinci-001.jpg 24,00... v
6 C:\AIDOS-X\AID_DATA\Inp_data\.\DaVinci-003.jpg 22,88... v
5 C:\AIDOS-X\AID_DATA\Inp_data\.\DaVinci-002.jpg 14,78... v
15 C:\AIDOS-X\AID_DATA\Inp_data\.\Shishkin-003.jpg -8,297...
9 C:\AIDOS-X\AID_DATA\Inp_data\.\Roerich-001.jpg -9,006...
12 C:\AIDOS-X\AID_DATA\Inp_data\.\Roerich-004.jpg -9,779...
8 C:\AIDOS-X\AID_DATA\Inp_data\.\Kuindzhi-002.jpg -10,26...
16 C:\AIDOS-X\AID_DATA\Inp_data\.\Shishkin-004.jpg -16,25...
13 C:\AIDOS-X\AID_DATA\Inp_data\.\Shishkin-001.jpg -22,10...
14 C:\AIDOS-X\AID_DATA\Inp_data\.\Shishkin-002.jpg -22,47...
Помощь
Поиск объекта
В начало БД
В конец БД
Предыдущая
Следующая
9 записей
Все записи
Печать XLS
Печать TXT
Печать ALL
4.1.3.2. Визуализация результатов распознавания в отношении: "Класс-объекты". Текущая модель: "INF1"
Классы
Код
Наим. класса
1 КЛАСС-1/5-{Aivazovsky}
2 КЛАСС-2/5-{DaVinci}
3 КЛАСС-3/5-{Kuindzhi}
4 КЛАСС-4/5-{Roerich}
5 КЛАСС-5/5-{Shishkin}
Интегральный критерий сходства: "Семантический резонанс знаний"
Код
Наименование объекта
Сходство
Ф...
Сходство
8 C:\AIDOS-X\AID_DATA\Inp_data\.\Kuindzhi-002.jpg 78,36... v
7 C:\AIDOS-X\AID_DATA\Inp_data\.\Kuindzhi-001.jpg 71,84... v
14 C:\AIDOS-X\AID_DATA\Inp_data\.\Shishkin-002.jpg 59,11...
16 C:\AIDOS-X\AID_DATA\Inp_data\.\Shishkin-004.jpg 45,64...
4 C:\AIDOS-X\AID_DATA\Inp_data\.\DaVinci-001.jpg 35,03...
5 C:\AIDOS-X\AID_DATA\Inp_data\.\DaVinci-002.jpg 31,48...
3 C:\AIDOS-X\AID_DATA\Inp_data\.\Aivazovsky-003.jpg 31,27...
13 C:\AIDOS-X\AID_DATA\Inp_data\.\Shishkin-001.jpg 28,27...
6 C:\AIDOS-X\AID_DATA\Inp_data\.\DaVinci-003.jpg 18,48...
15 C:\AIDOS-X\AID_DATA\Inp_data\.\Shishkin-003.jpg 7,935...
Интегральный критерий сходства: "Сумма знаний"
Код
Наименование объекта
Сходство
Ф...
Сходство
7 C:\AIDOS-X\AID_DATA\Inp_data\.\Kuindzhi-001.jpg 31,41... v
8 C:\AIDOS-X\AID_DATA\Inp_data\.\Kuindzhi-002.jpg 24,38... v
14 C:\AIDOS-X\AID_DATA\Inp_data\.\Shishkin-002.jpg 12,73...
16 C:\AIDOS-X\AID_DATA\Inp_data\.\Shishkin-004.jpg -2,651...
6 C:\AIDOS-X\AID_DATA\Inp_data\.\DaVinci-003.jpg -5,521...
4 C:\AIDOS-X\AID_DATA\Inp_data\.\DaVinci-001.jpg -14,36...
5 C:\AIDOS-X\AID_DATA\Inp_data\.\DaVinci-002.jpg -14,89...
13 C:\AIDOS-X\AID_DATA\Inp_data\.\Shishkin-001.jpg -16,98...
3 C:\AIDOS-X\AID_DATA\Inp_data\.\Aivazovsky-003.jpg -28,49...
15 C:\AIDOS-X\AID_DATA\Inp_data\.\Shishkin-003.jpg -28,81...
Помощь
Поиск объекта
В начало БД
В конец БД
Предыдущая
Следующая
9 записей
Все записи
Печать XLS
Печать TXT
Печать ALL



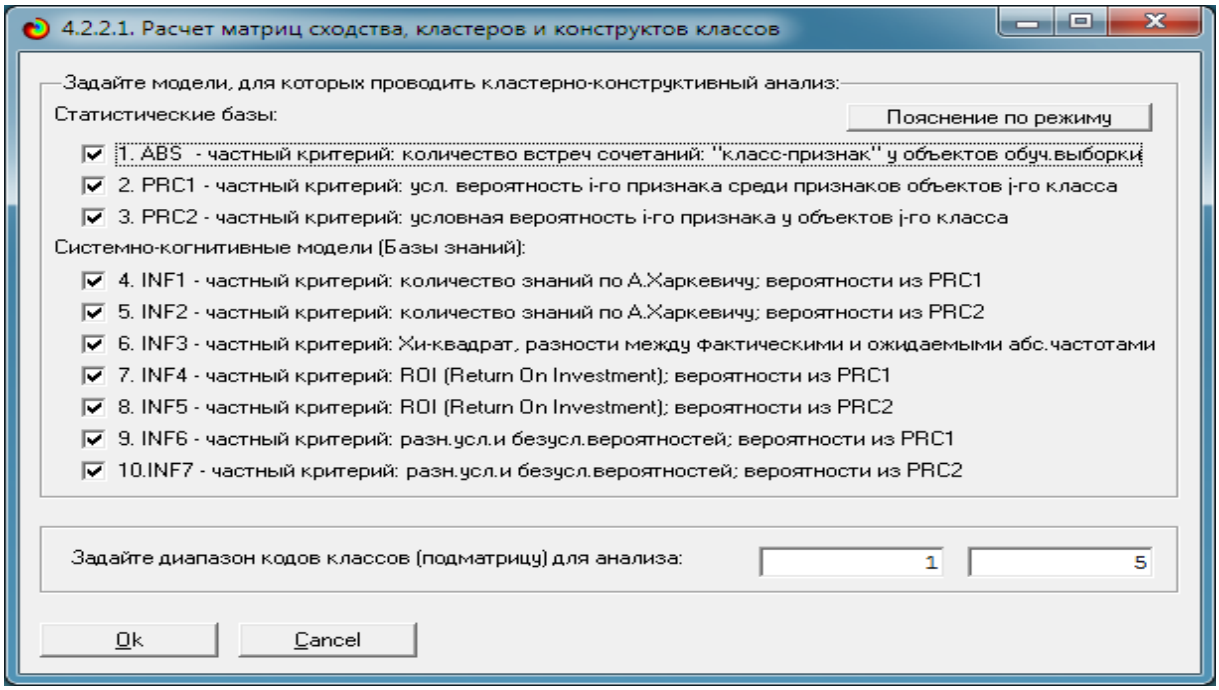


Справа птичками отмечены те объекты распознаваемой выборки (картины художников), которые действительно относятся к классам, выбранным слева.

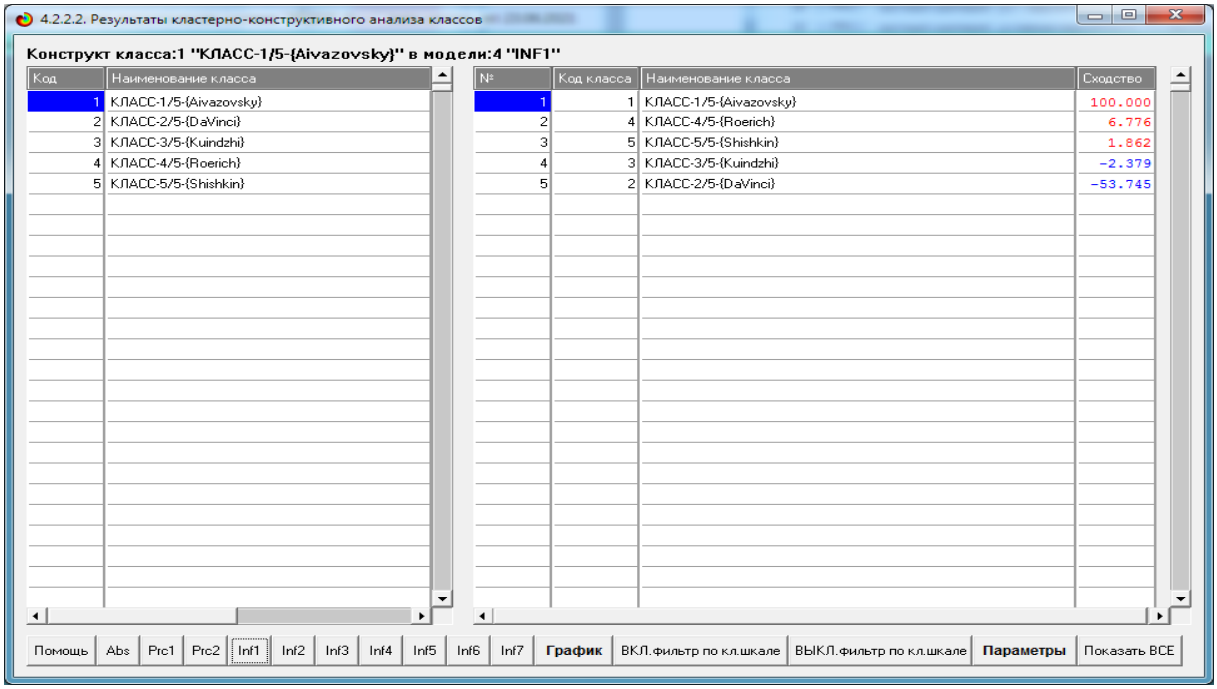
Результаты идентификации конкретных картин с классами, соответствующими художникам (их видим в левом окне), смотрим на правом нижнем окне, т.к. оно содержит результаты с интегральным критерием «Сумма знаний», который в соответствии с рекомендациями режима 3.4 используется при решении задач.

14.9.2. Решение задачи сравнения обобщенных образов классов друг с другом (задача кластерно-конструктивного анализа классов)

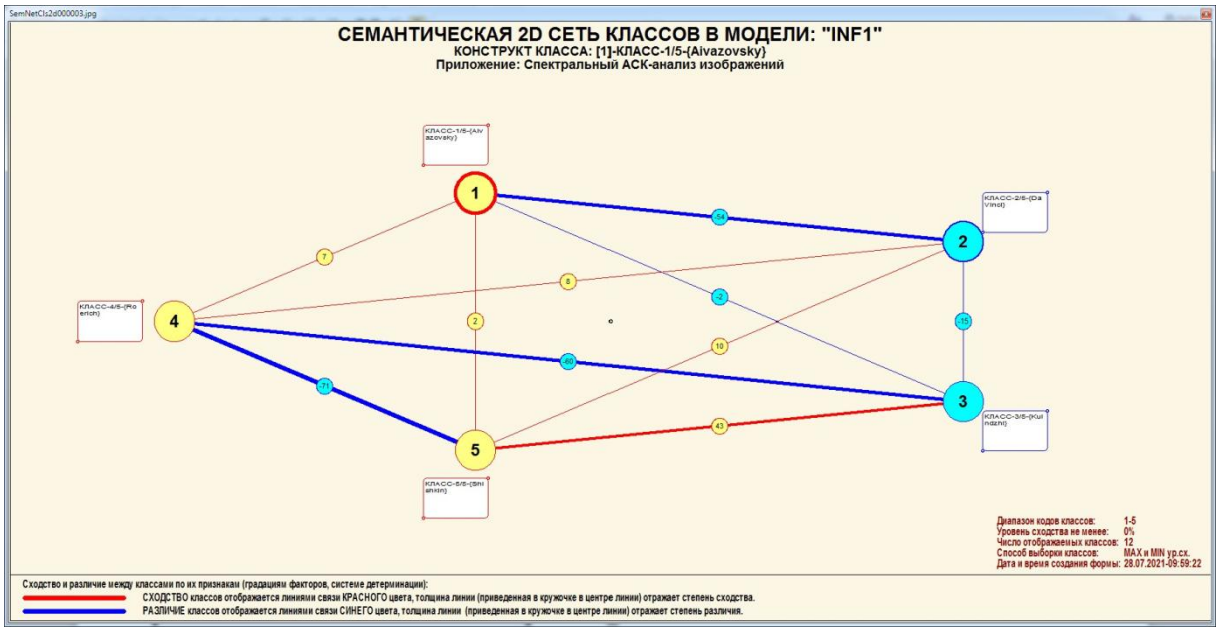
Проведем расчет матрицы сходства классов в режиме 4.2.2.1:



Результат мы видим на экранной форме режима 4.2.2.2:

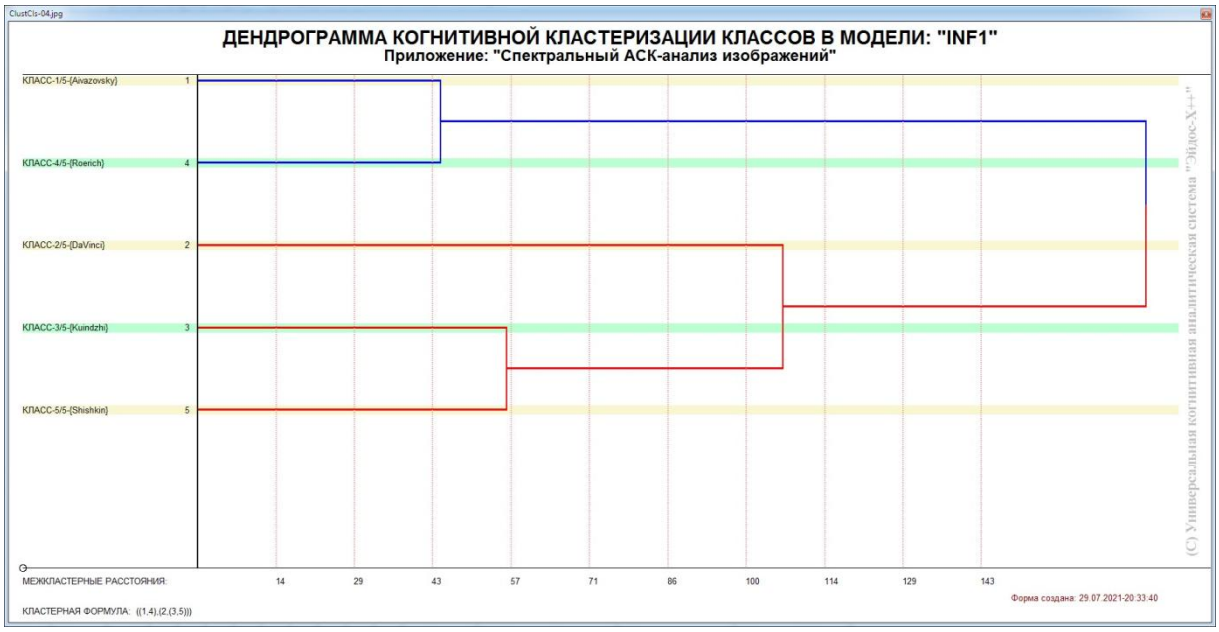


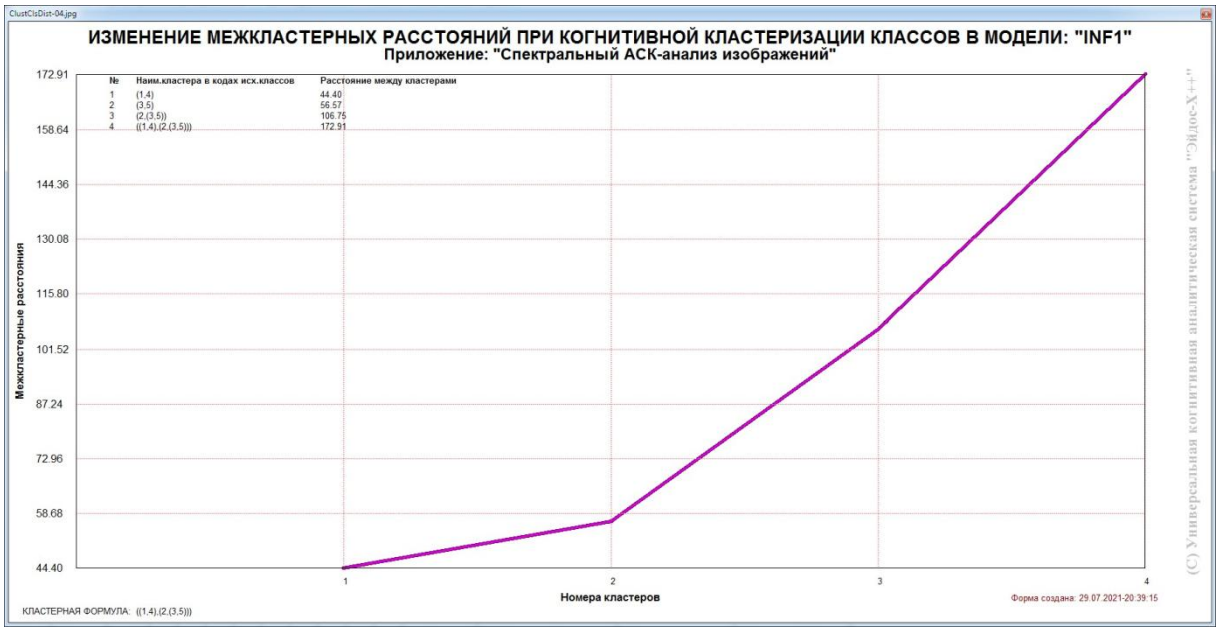
В графической форме в модели INF1:



Обратим внимание на то, что этот результат сформирован не на основе обобщения экспертных оценок (как обычно формируются подобные когнитивные диаграммы), а путем сравнения обобщенных спектров классов в системно-когнитивной модели.

Та же самая матрица сходства обобщенных образов классов может быть визуализирована также в форме дендрограммы когнитивной кластеризации [47]:

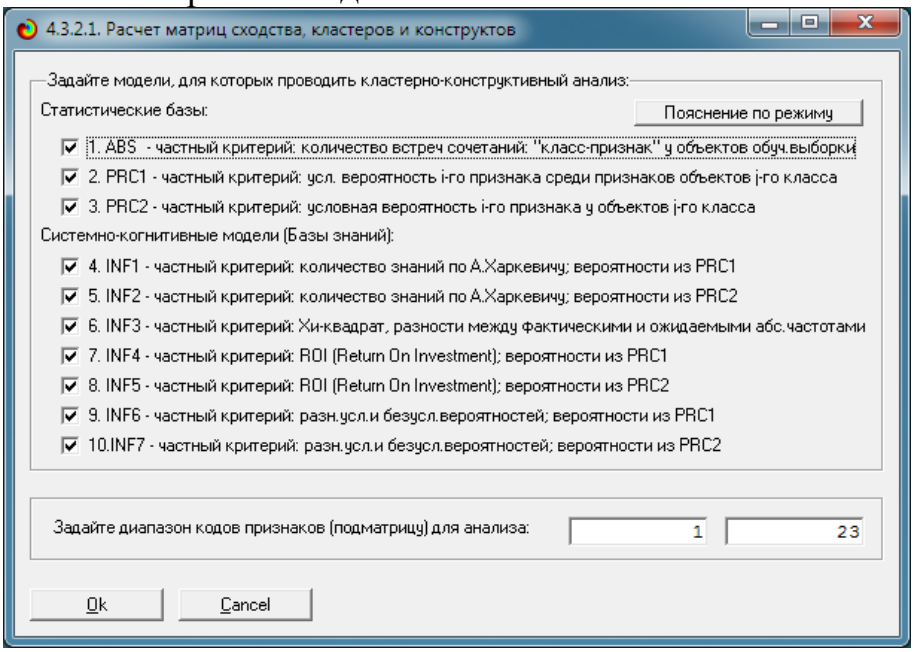




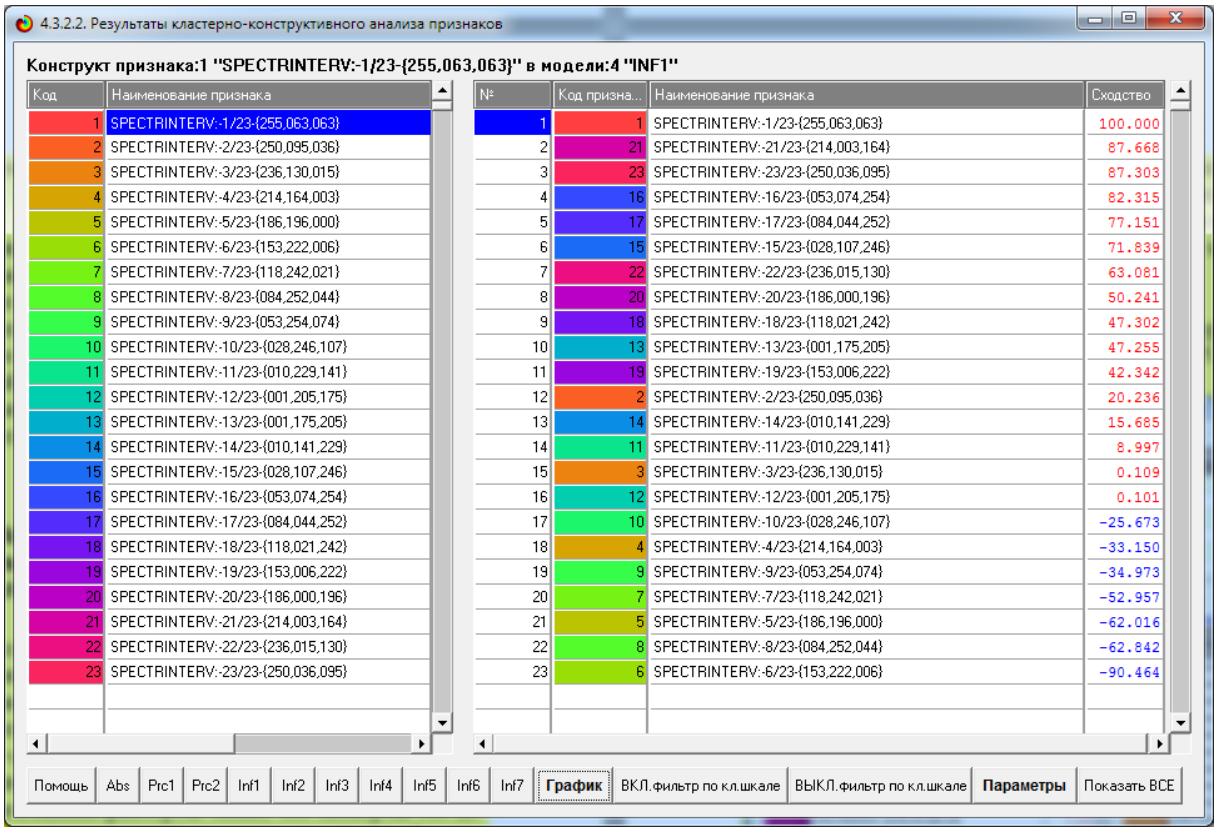
Из приведенной дендрограммы видно, что спектральные характеристики исследованных картин художников довольно сильно отличаются друг от друга, что и позволяет по спектру картины довольно точно идентифицировать ее автора. Картины Куинджи и Шишкина ожидаемо оказались более сходны друг по спектру, остальные результаты, если подумать, тоже соответствуют интуитивно ожидаемым.

14.9.3. Решение задачи сравнения обобщенных образов признаков друг с другом (задача кластерно-конструктивного анализа признаков)

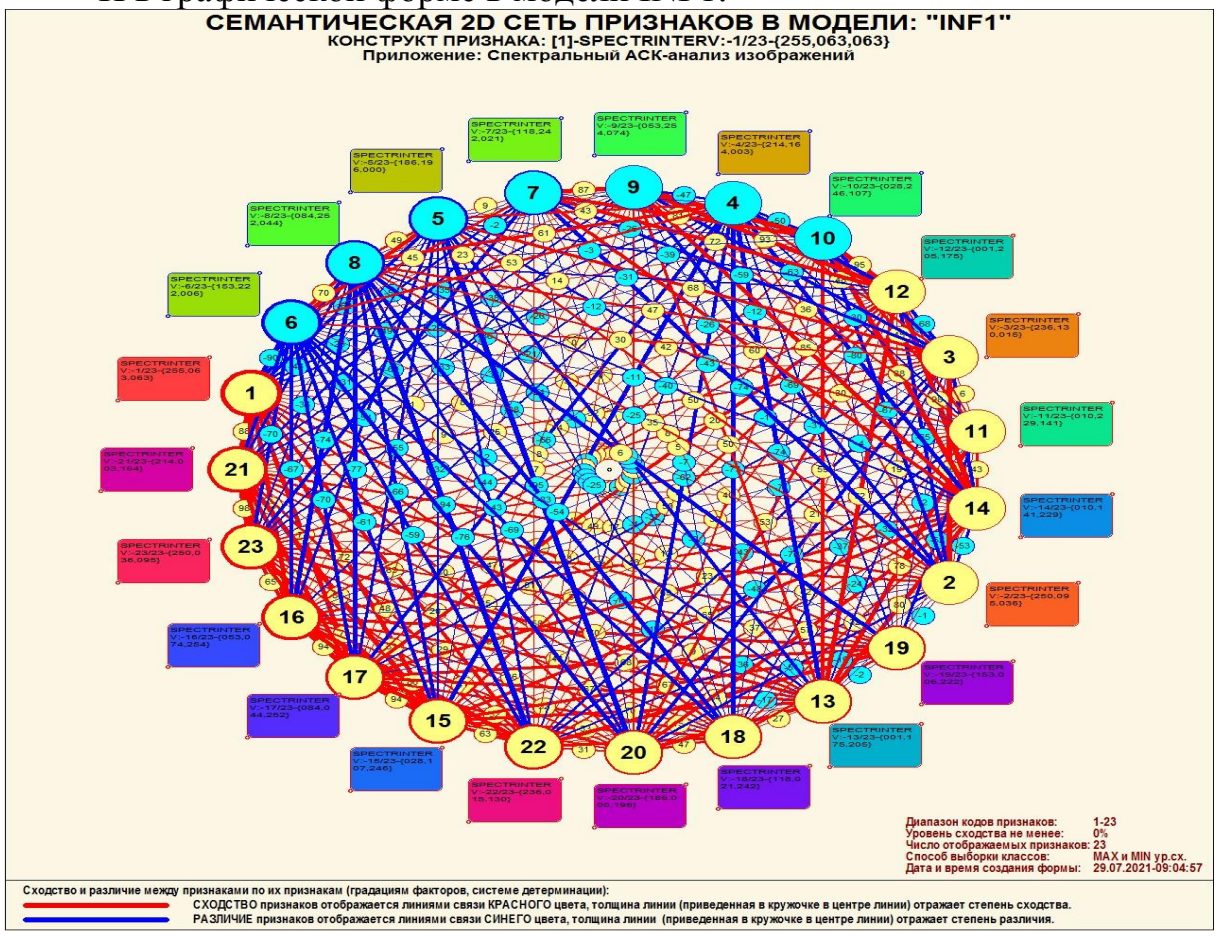
Проведем расчет матрицы сходства признаков в режиме 4.3.2.1. Отметим, что признаком в данной модели является значение цвета определенного спектрального диапазона.



Результат мы видим на экранной форме режима 4.3.2.2:



И в графической форме в модели INF1:

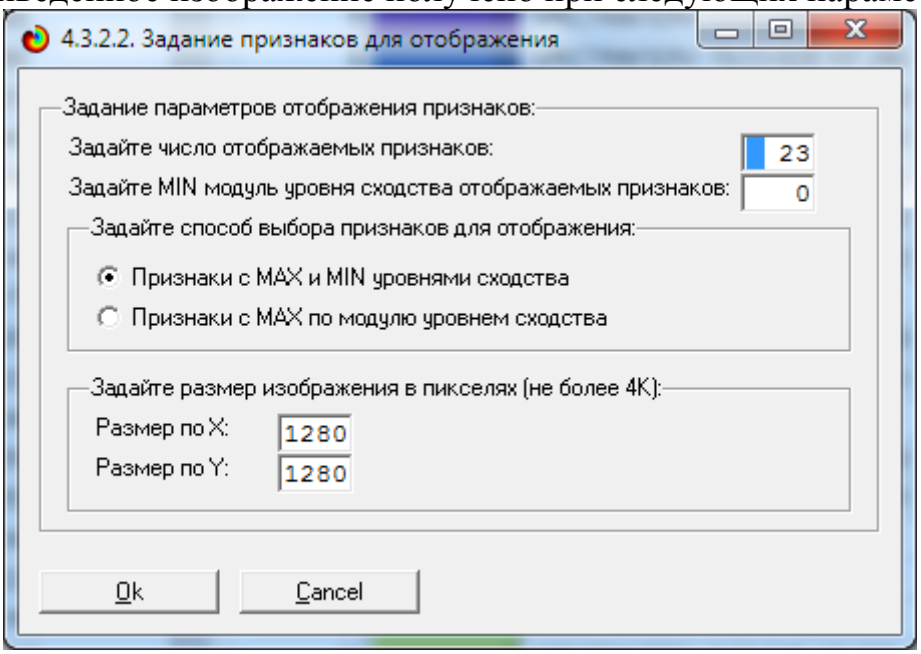


Цвет фона на наименовании признака соответствует данному признаку, т.е. цвету спектрального диапазона.

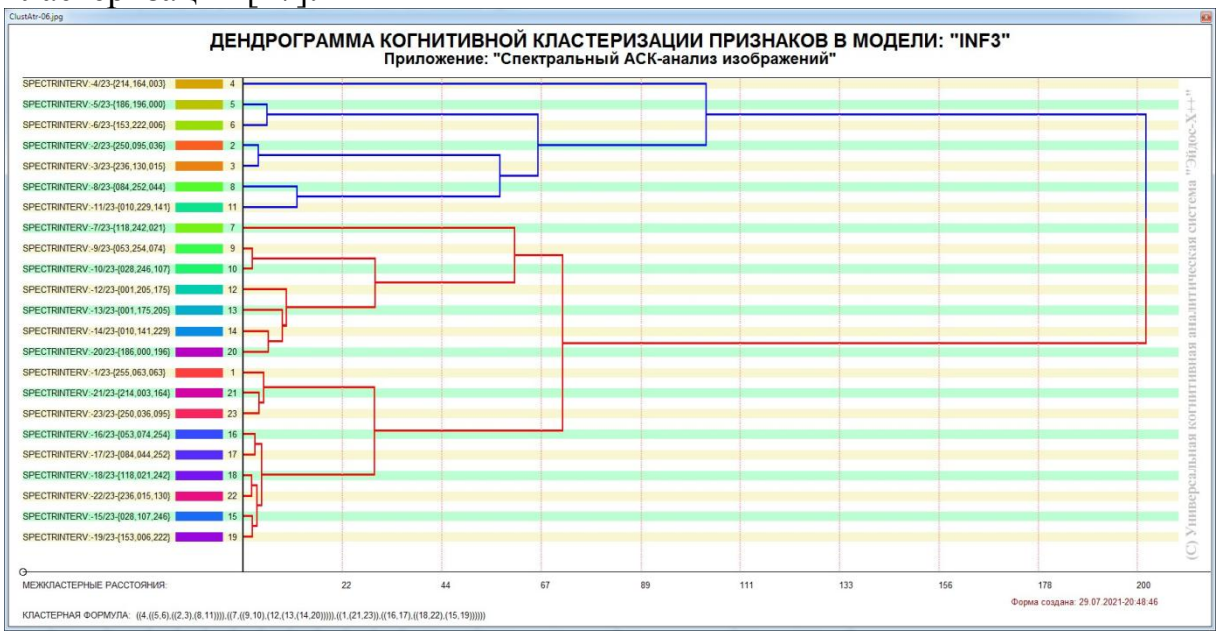
Обратим внимание на то, что этот результат сформирован не на основе обобщения экспертных оценок (как обычно формируются подобные когнитивные диаграммы), а путем сравнения обобщенных спектров классов в системно-когнитивной модели.

Видно, что спектральные диапазоны образуют два кластера, которые условно можно назвать «Зеленым» и «Красным», которые образуют полюса конструкта «Цвет».

Приведенное изображение получено при следующих параметрах:



Та же самая матрица сходства обобщенных образов классов может быть визуализирована также в форме дендрограммы когнитивной кластеризации [47]:



В приведенной дендрограмме наглядно показано сходство/различие цветов по содержащейся в них информации о принадлежности картины с этим цветом тому или иному художнику.

14.9.4. Решение задачи исследования моделируемой предметной области путем исследования ее модели (автоматизированный SWOT-анализ изображений)

SWOT-анализ является широко известным и общепризнанным метод стратегического планирования. Однако это не мешает тому, что он подвергается критике, часто вполне справедливой, обоснованной и хорошо аргументированной. В результате критического рассмотрения SWOT-анализа выявлено довольно много его слабых сторон (недостатков), источником которых является необходимость привлечения экспертов, в частности для оценки силы и направления влияния факторов. Ясно, что эксперты это делают неформализуемым путем (интуитивно), на основе своего профессионального опыта и компетенции.

Но возможности экспертов имеют свои ограничения и часто по различным причинам они не могут и не хотят это сделать, к тому же время экспертов стоит очень дорого и они не могут дать количественные оценки.

Таким образом, возникает проблема проведения SWOT-анализа без привлечения экспертов.

Эта проблема может решаться путем автоматизации функций экспертов, т.е. путем измерения силы и направления влияния факторов непосредственно на основе эмпирических данных.

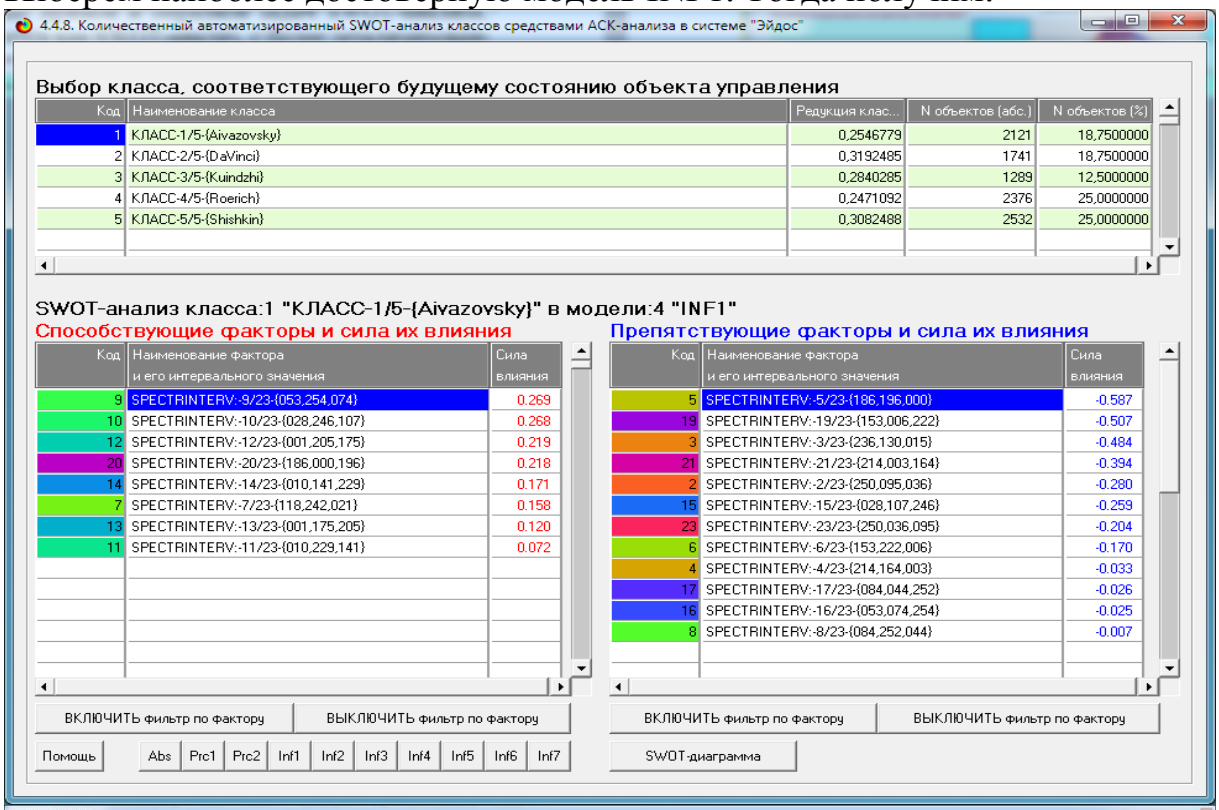
Подобная технология разработана давно, ей уже около 30 лет, но она малоизвестна – это интеллектуальная система «Эйдос» [11-33].

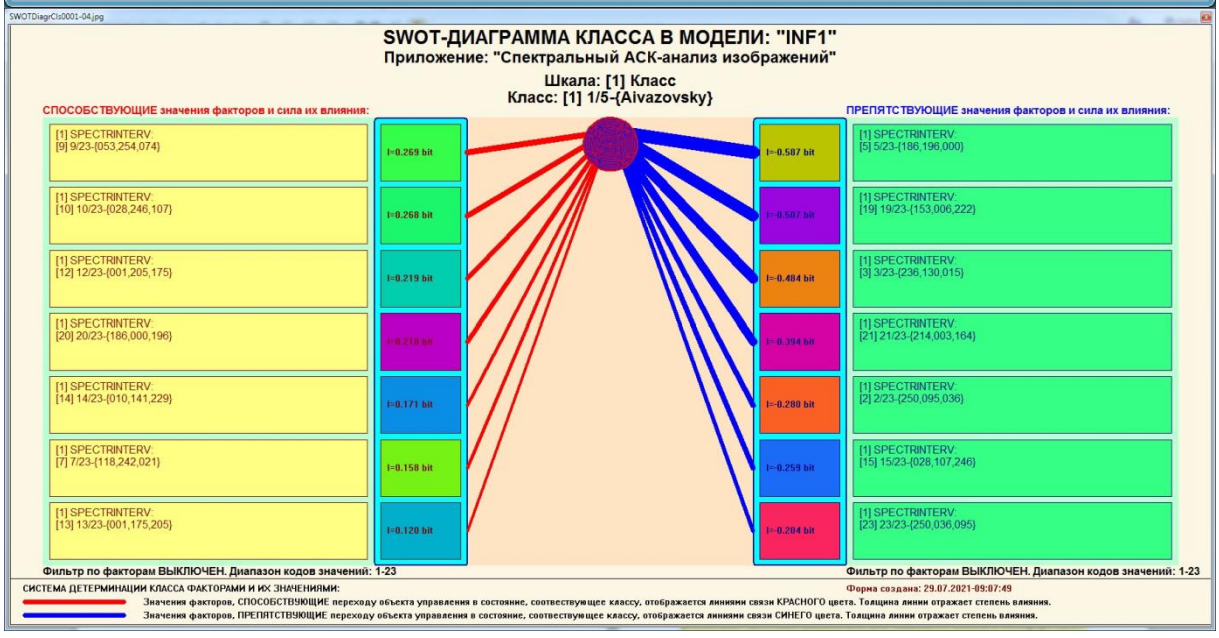
В статье [42] на реальном численном примере подробно описывается возможность проведения количественного автоматизированного SWOT-

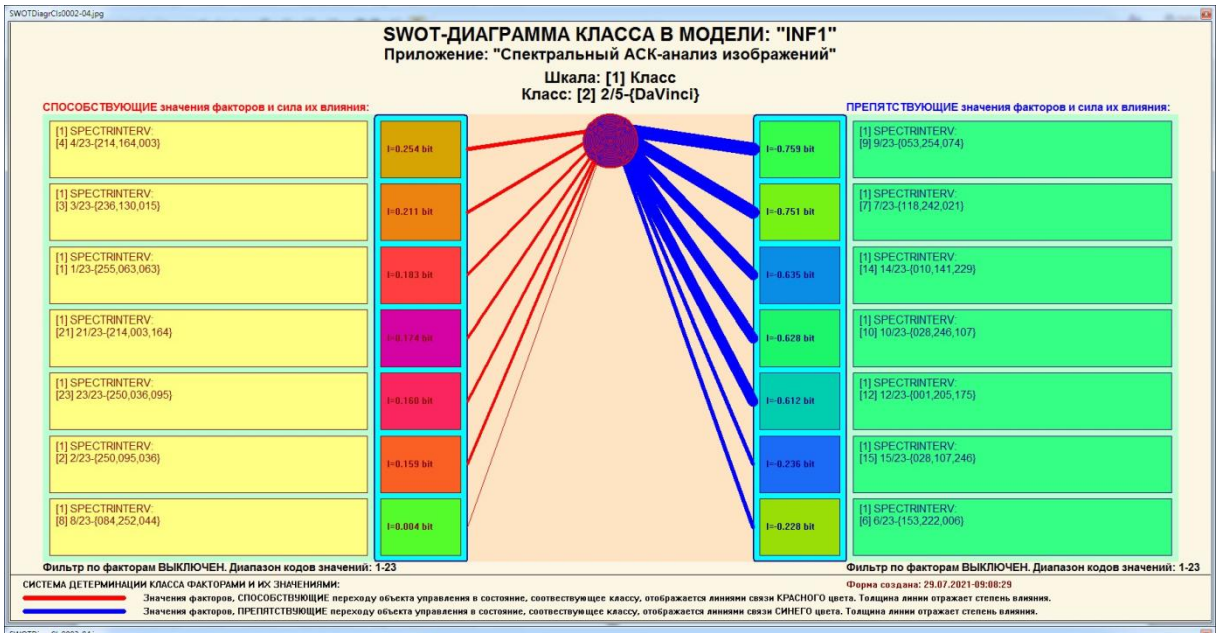
анализа средствами АСК-анализа и интеллектуальной системы «Эйдос-X++» без использования экспертных оценок непосредственно на основе эмпирических данных.

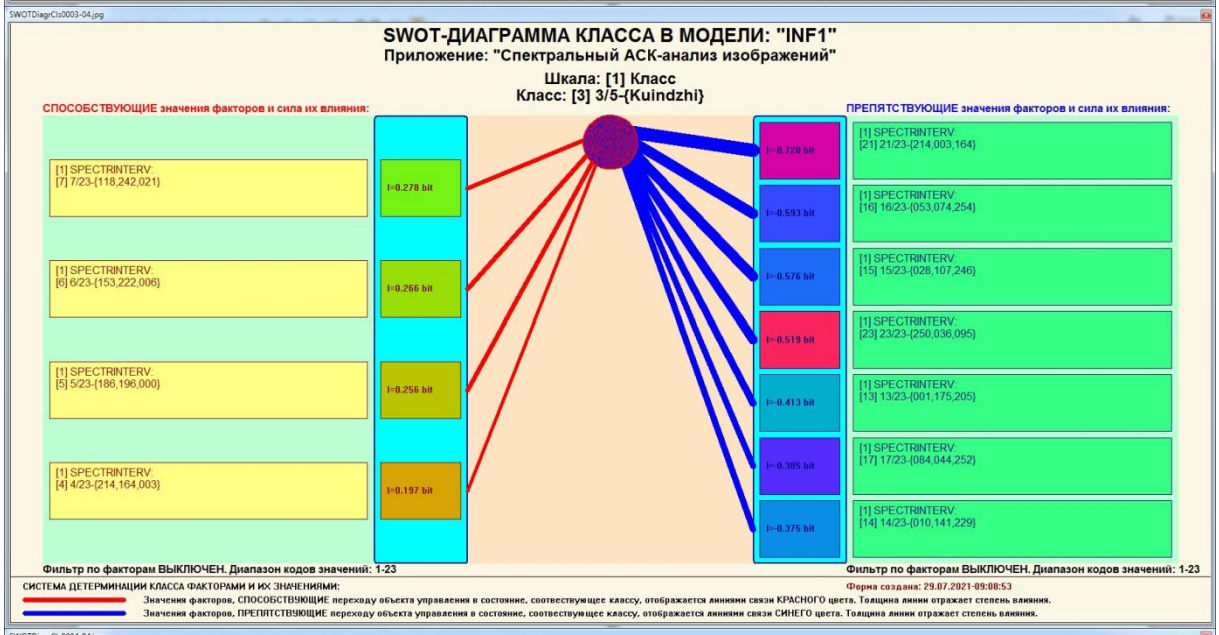
Предложено решение прямой и обратной задач SWOT-анализа. PEST-анализ рассматривается как SWOT-анализ, с более детализированной классификацией внешних факторов. Поэтому выводы, полученные в данной работе на примере SWOT-анализа, можно распространить и на PEST-анализ [46].

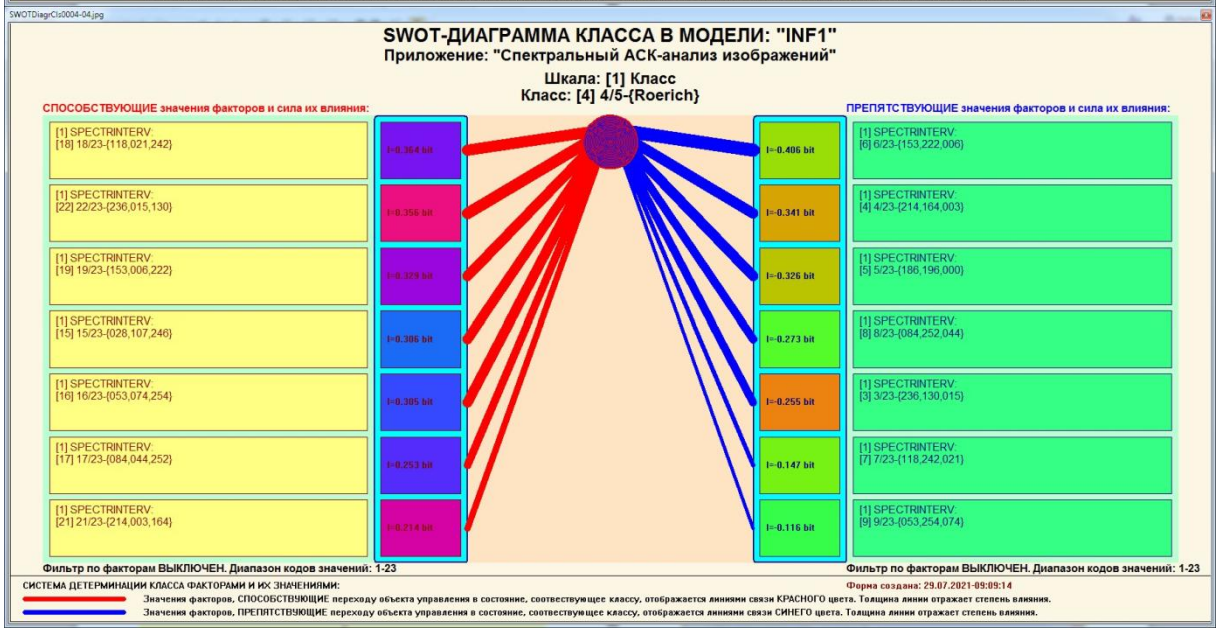
Запустим режим 4.4.8. На экранной форме режима зададим класс и выберем наиболее достоверную модель INF1. Тогда получим:

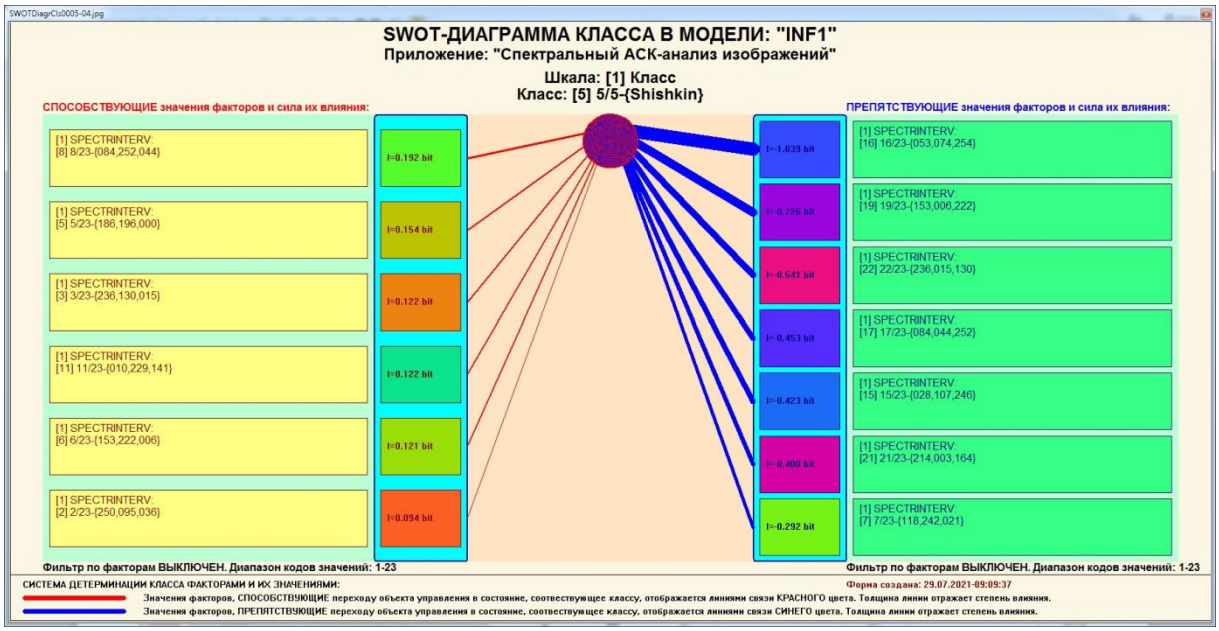




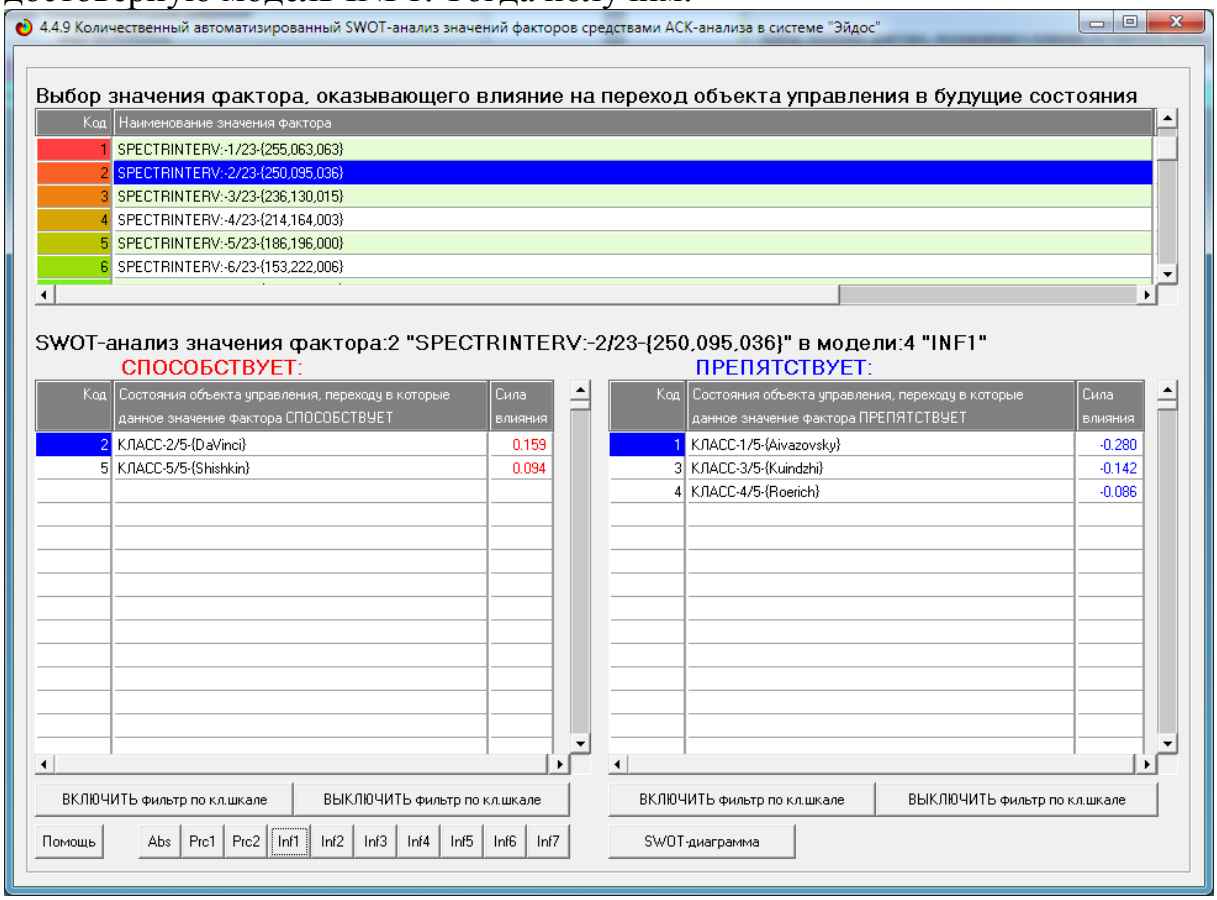

SWOTDiagrCls0002-04.jpg
SWOT-ДИАГРАММА КЛАССА В МОДЕЛИ: "INF1"
Приложение: "Спектральный АСК-анализ изображений"
Шкала: [1] Класс
Класс: [2] 2/5-{DaVinci}
СПОСОБСТВУЮЩИЕ значения факторов и сила их влияния:
ПРЕПЯТСТВУЮЩИЕ значения факторов и сила их влияния:
[1] SPECTRINTERV: [4] 4/23-{214,164,003} I=0.254 bit
[1] SPECTRINTERV: [3] 3/23-{236,130,015} I=0.211 bit
[1] SPECTRINTERV: [1] 1/23-{255,063,063} I=0.183 bit
[1] SPECTRINTERV: [21] 21/23-{214,003,164} I=0.174 bit
[1] SPECTRINTERV: [23] 23/23-{250,036,095} I=0.168 bit
[1] SPECTRINTERV: [2] 2/23-{250,095,036} I=0.159 bit
[1] SPECTRINTERV: [8] 8/23-{084,252,044} I=0.004 bit
I=0.759 bit [1] SPECTRINTERV: [9] 9/23-{053,254,074}
I=0.751 bit [1] SPECTRINTERV: [7] 7/23-{118,242,021}
I=0.635 bit [1] SPECTRINTERV: [14] 14/23-{010,141,229}
I=0.628 bit [1] SPECTRINTERV: [10] 10/23-{028,246,107}
I=0.612 bit [1] SPECTRINTERV: [12] 12/23-{001,205,175}
I=0.236 bit [1] SPECTRINTERV: [15] 15/23-{028,107,246}
I=0.228 bit [1] SPECTRINTERV: [6] 6/23-{153,222,006}
Фильтр по факторам ВЫКЛЮЧЕН. Диапазон кодов значений: 1-23
Фильтр по факторам ВЫКЛЮЧЕН. Диапазон кодов значений: 1-23
СИСТЕМА ДЕТЕРМИНАЦИИ КЛАССА ФАКТОРАМИ И ИХ ЗНАЧЕНИЯМИ:
Форма создана: 29.07.2021-09:08:29
Значения факторов, СПОСОБСТВУЮЩИЕ переходу объекта управления в состояние, соответствующее классу, отображается линиями связи КРАСНОГО цвета. Толщина линии отражает степень влияния.
Значения факторов, ПРЕПЯТСТВУЮЩИЕ переходу объекта управления в состояние, соответствующее классу, отображается линиями связи СИНЕГО цвета. Толщина линии отражает степень влияния.

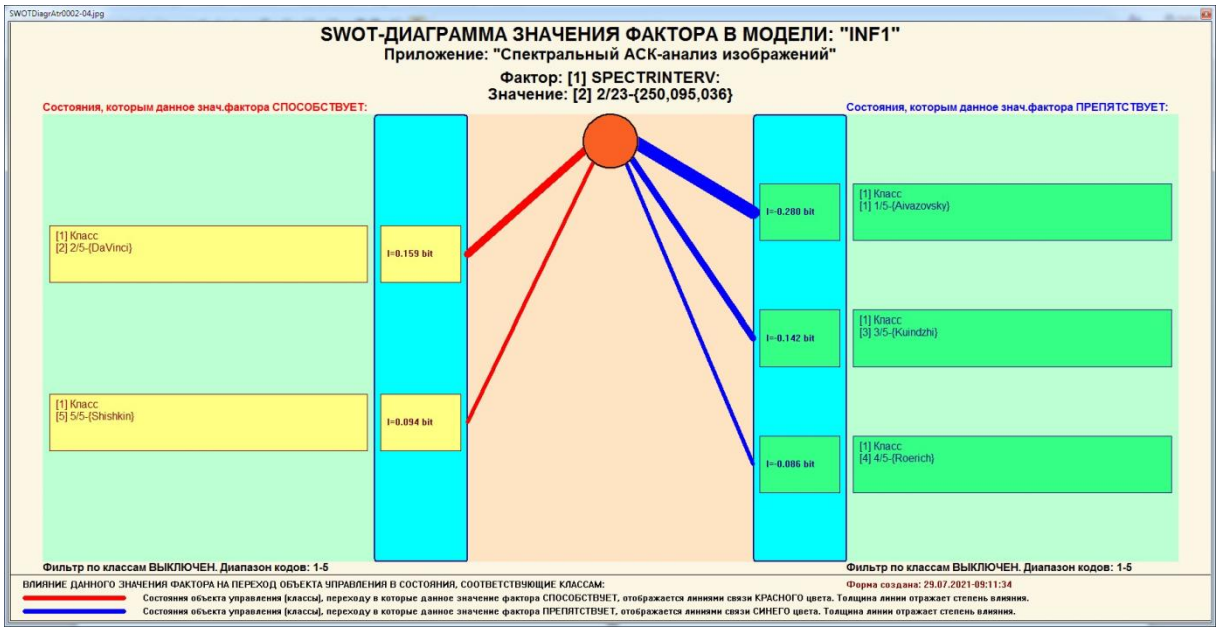

SWOTDiagrCls0003-04.jpg
SWOT-ДИАГРАММА КЛАССА В МОДЕЛИ: "INF1"
Приложение: "Спектральный АСК-анализ изображений"
Шкала: [1] Класс
Класс: [3] 3/5-{Kuindzhi}
СПОСОБСТВУЮЩИЕ значения факторов и сила их влияния:
ПРЕПЯТСТВУЮЩИЕ значения факторов и сила их влияния:
[1] SPECTRINTERV: [7] 7/23-{118,242,021} I=0.278 bit
[1] SPECTRINTERV: [6] 6/23-{153,222,006} I=0.266 bit
[1] SPECTRINTERV: [5] 5/23-{186,196,000} I=0.256 bit
[1] SPECTRINTERV: [4] 4/23-{214,164,003} I=0.197 bit
I=0.728 bit [1] SPECTRINTERV: [21] 21/23-{214,003,164}
I=0.593 bit [1] SPECTRINTERV: [16] 16/23-{053,074,254}
I=0.576 bit [1] SPECTRINTERV: [15] 15/23-{028,107,246}
I=0.519 bit [1] SPECTRINTERV: [23] 23/23-{250,036,095}
I=0.413 bit [1] SPECTRINTERV: [13] 13/23-{001,175,205}
I=0.385 bit [1] SPECTRINTERV: [17] 17/23-{084,044,252}
I=0.375 bit [1] SPECTRINTERV: [14] 14/23-{010,141,229}
Фильтр по факторам ВЫКЛЮЧЕН. Диапазон кодов значений: 1-23
Фильтр по факторам ВЫКЛЮЧЕН. Диапазон кодов значений: 1-23
СИСТЕМА ДЕТЕРМИНАЦИИ КЛАССА ФАКТОРАМИ И ИХ ЗНАЧЕНИЯМИ:
Форма создана: 29.07.2021-09:08:53
Значения факторов, СПОСОБСТВУЮЩИЕ переходу объекта управления в состояние, соответствующее классу, отображается линиями связи КРАСНОГО цвета. Толщина линии отражает степень влияния.
Значения факторов, ПРЕПЯТСТВУЮЩИЕ переходу объекта управления в состояние, соответствующее классу, отображается линиями связи СИНЕГО цвета. Толщина линии отражает степень влияния.


SWOTDiagrCls0004-04.jpg
SWOT-ДИАГРАММА КЛАССА В МОДЕЛИ: "INF1"
Приложение: "Спектральный АСК-анализ изображений"
Шкала: [1] Класс
Класс: [4] 4/5-{Roerich}
СПОСОБСТВУЮЩИЕ значения факторов и сила их влияния:
ПРЕПЯТСТВУЮЩИЕ значения факторов и сила их влияния:
[1] SPECTRINTERV: [18] 18/23-{118,021,242} I=0.364 bit
[1] SPECTRINTERV: [22] 22/23-{236,015,130} I=0.356 bit
[1] SPECTRINTERV: [19] 19/23-{153,006,222} I=0.329 bit
[1] SPECTRINTERV: [15] 15/23-{028,107,246} I=0.306 bit
[1] SPECTRINTERV: [16] 16/23-{053,074,254} I=0.305 bit
[1] SPECTRINTERV: [17] 17/23-{084,044,252} I=0.253 bit
[1] SPECTRINTERV: [21] 21/23-{214,003,164} I=0.214 bit
I=0.406 bit [1] SPECTRINTERV: [6] 6/23-{153,222,006}
I=0.341 bit [1] SPECTRINTERV: [4] 4/23-{214,164,003}
I=0.326 bit [1] SPECTRINTERV: [5] 5/23-{186,196,000}
I=0.273 bit [1] SPECTRINTERV: [8] 8/23-{084,252,044}
I=0.255 bit [1] SPECTRINTERV: [3] 3/23-{236,130,015}
I=0.147 bit [1] SPECTRINTERV: [7] 7/23-{118,242,021}
I=0.116 bit [1] SPECTRINTERV: [9] 9/23-{053,254,074}
Фильтр по факторам ВЫКЛЮЧЕН. Диапазон кодов значений: 1-23
Фильтр по факторам ВЫКЛЮЧЕН. Диапазон кодов значений: 1-23
СИСТЕМА ДЕТЕРМИНАЦИИ КЛАССА ФАКТОРАМИ И ИХ ЗНАЧЕНИЯМИ:
Форма создана: 29.07.2021-09:09:14
Значения факторов, СПОСОБСТВУЮЩИЕ переходу объекта управления в состояние, соответствующее классу, отображается линиями связи КРАСНОГО цвета. Толщина линии отражает степень влияния.
Значения факторов, ПРЕПЯТСТВУЮЩИЕ переходу объекта управления в состояние, соответствующее классу, отображается линиями связи СИНЕГО цвета. Толщина линии отражает степень влияния.



Доля решения обратной задачи SWOT-анализа запустим режим 4.4.9. На экранной форме режима зададим класс и выберем наиболее достоверную модель INF1. Тогда получим:

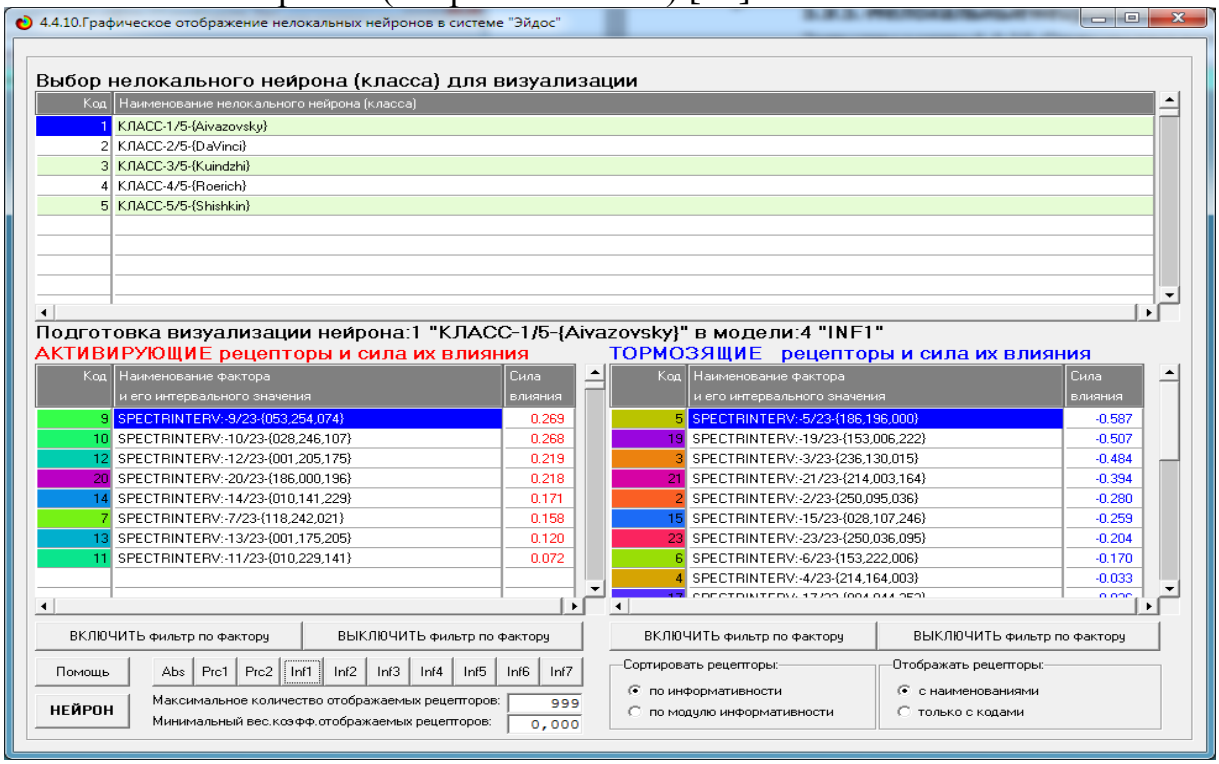




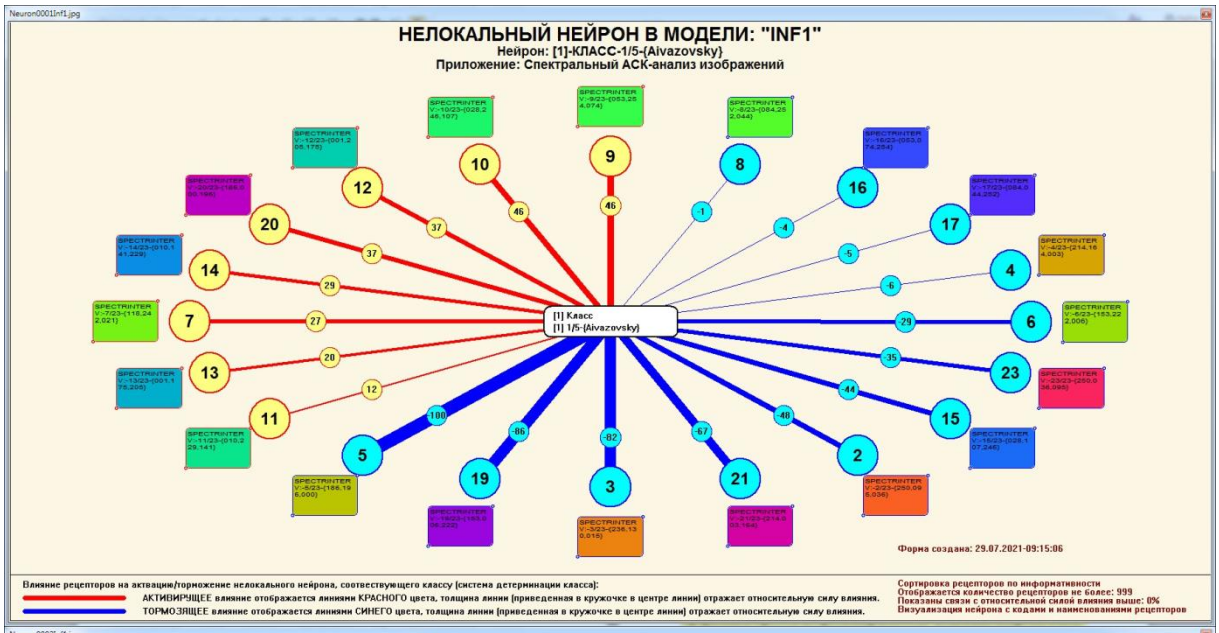
Все инвертированные SWOT-диаграммы, которых 23, по числу цветовых диапазонов, в данной работе мы не приводим из-за ограниченности ее объема.

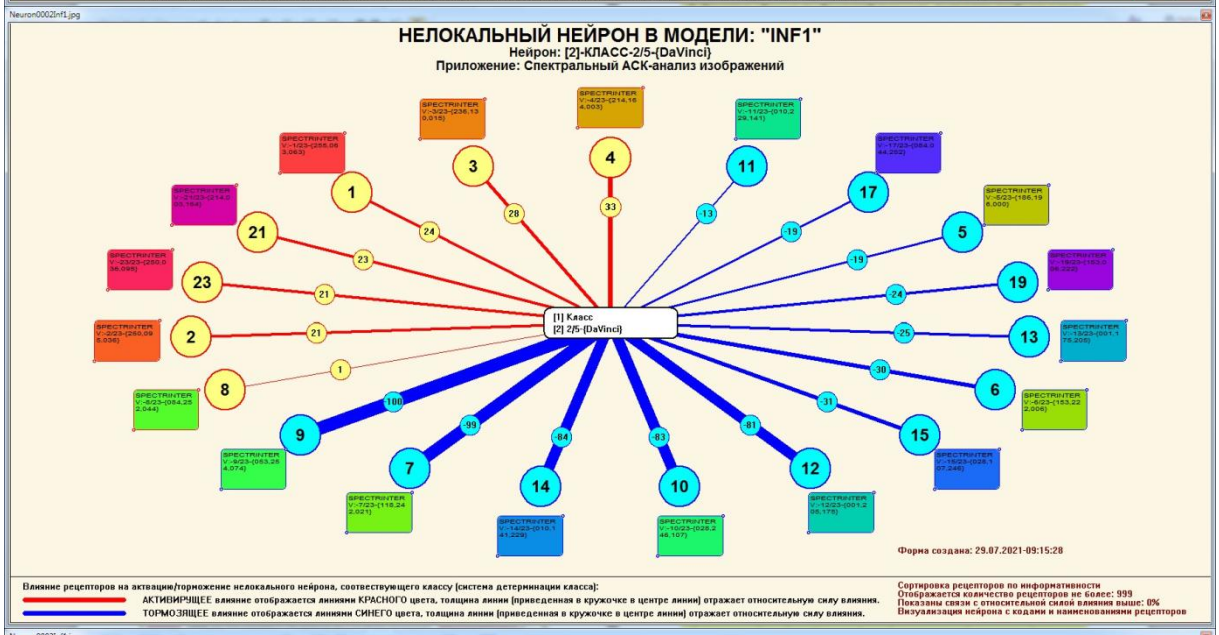
14.9.5. Нелокальные нейроны классов

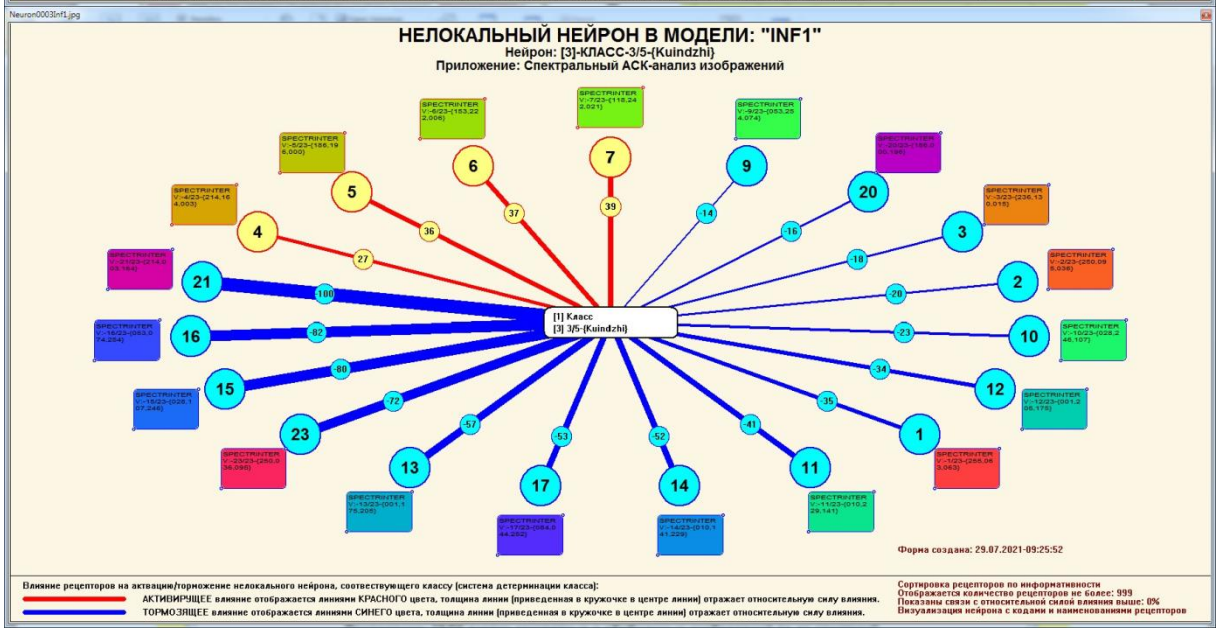
Запустим режим 4.4.10. Он позволяет вывести степень характерности различных цветовых диапазонов для заданного класса в метафоре нелокальных нейронов (нейронных сетей) [43]:

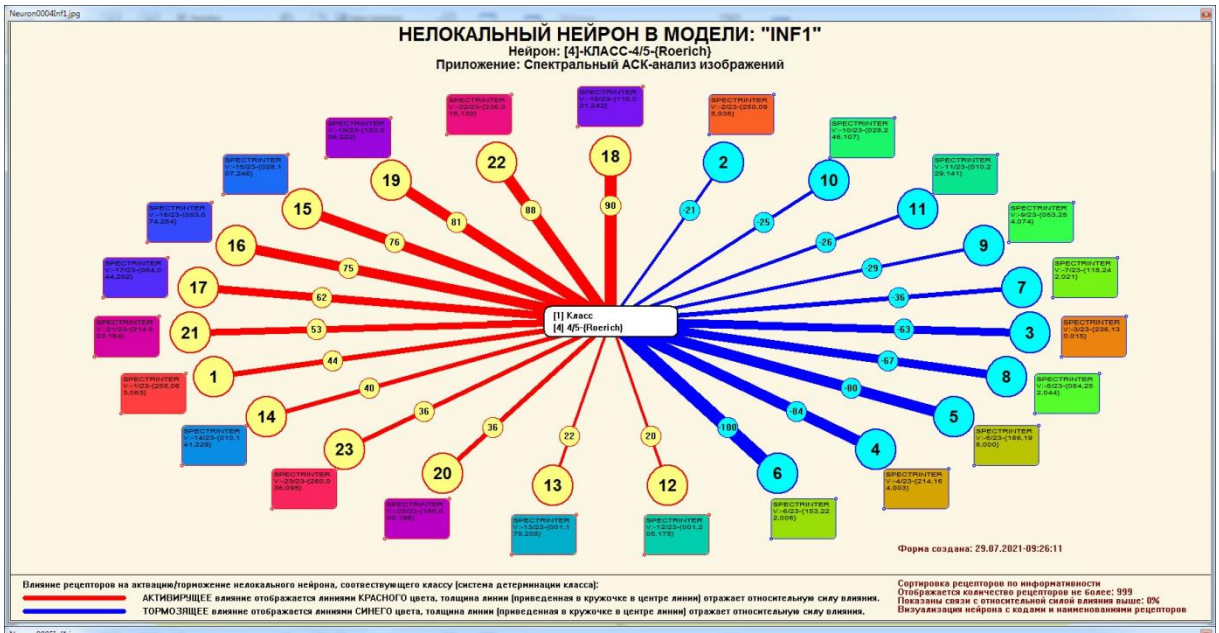


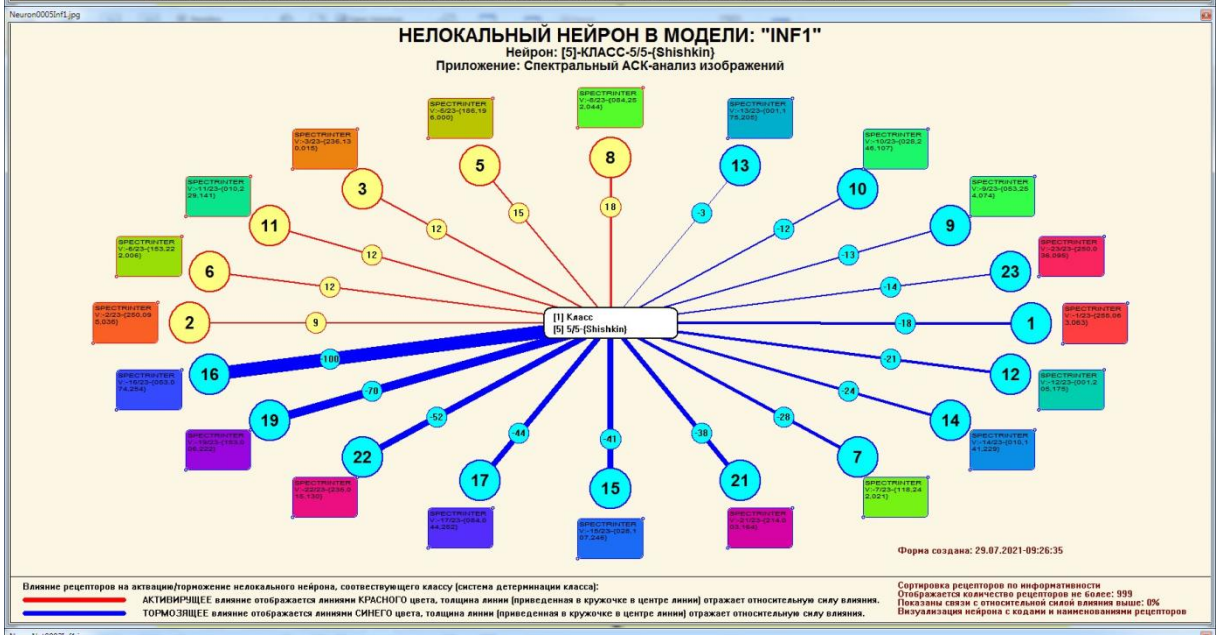
Цвет фона на наименовании признака соответствует данному признаку, т.е. цвету спектрального диапазона. Приведем нейроны по классам, соответствующим другим художникам, а также нейронную сеть.

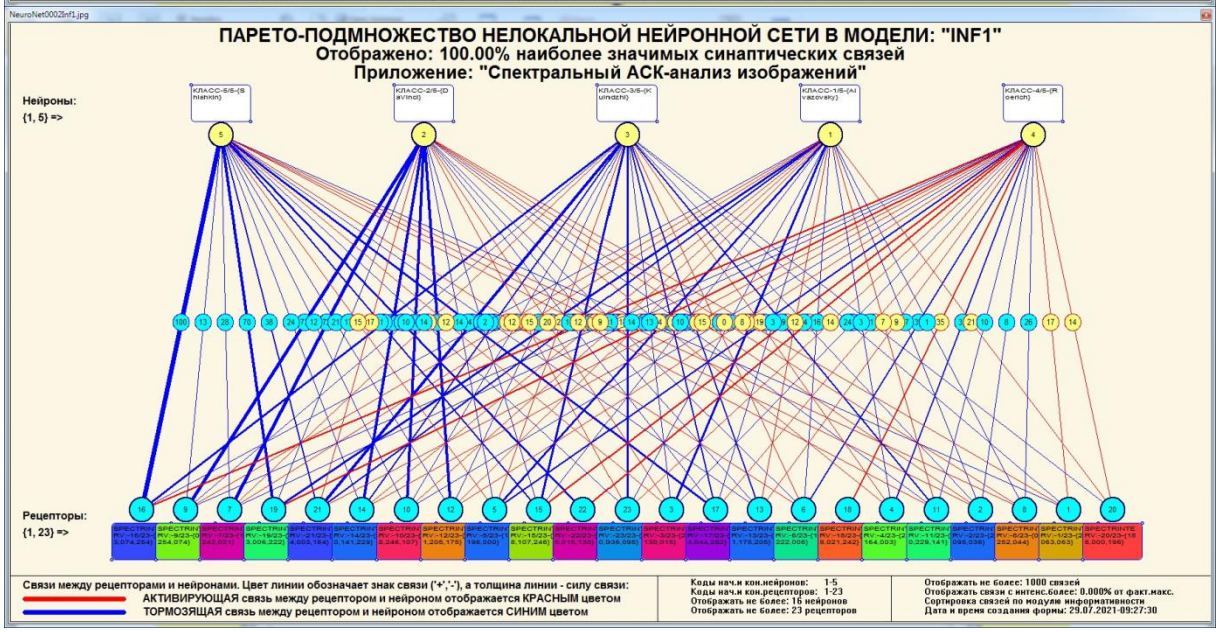

Neuron0001Inf1.jpg
НЕЛОКАЛЬНЫЙ НЕЙРОН В МОДЕЛИ: "INF1"
Нейрон: [1]-КЛАСС-1/5-{Aivazovsky}
Приложение: Спектральный АСК-анализ изображений
[1] Класс
[1] 1/5-{Aivazovsky}
Форма создана: 29.07.2021-09:15:06
Влияние рецепторов на активацию/торможение нелокального нейрона, соответствующего классу [система детерминации класса]:
АКТИВИРУЮЩЕЕ влияние отображается линиями КРАСНОГО цвета, толщина линии [приведенная в кружочке в центре линии] отражает относительную силу влияния.
ТОРМОЗЯЩЕЕ влияние отображается линиями СИНЕГО цвета, толщина линии [приведенная в кружочке в центре линии] отражает относительную силу влияния.
Сортировка рецепторов по информативности
Отображается количество рецепторов не более: 999
Показаны связи с относительной силой влияния выше: 0%
Визуализация нейрона с кодами и наименованиями рецепторов


Neuron0002Inf1.jpg
НЕЛОКАЛЬНЫЙ НЕЙРОН В МОДЕЛИ: "INF1"
Нейрон: [2]-КЛАСС-2/5-{DaVinci}
Приложение: Спектральный АСК-анализ изображений
[1] Класс
[2] 2/5-{DaVinci}
Форма создана: 29.07.2021-09:15:28
Влияние рецепторов на активацию/торможение нелокального нейрона, соответствующего классу [система детерминации класса]:
АКТИВИРУЮЩЕЕ влияние отображается линиями КРАСНОГО цвета, толщина линии [приведенная в кружочке в центре линии] отражает относительную силу влияния.
ТОРМОЗЯЩЕЕ влияние отображается линиями СИНЕГО цвета, толщина линии [приведенная в кружочке в центре линии] отражает относительную силу влияния.
Сортировка рецепторов по информативности
Отображается количество рецепторов не более: 999
Показаны связи с относительной силой влияния выше: 0%
Визуализация нейрона с кодами и наименованиями рецепторов


Neuron0003Inf1.jpg
НЕЛОКАЛЬНЫЙ НЕЙРОН В МОДЕЛИ: "INF1"
Нейрон: [3]-КЛАСС-3/5-{Kuindzhi}
Приложение: Спектральный АСК-анализ изображений
[1] Класс
[3] 3/5-{Kuindzhi}
Форма создана: 29.07.2021-09:25:52
Влияние рецепторов на активацию/торможение нелокального нейрона, соответствующего классу [система детерминации класса]:
АКТИВИРУЮЩЕЕ влияние отображается линиями КРАСНОГО цвета, толщина линии [приведенная в кружочке в центре линии] отражает относительную силу влияния.
ТОРМОЗЯЩЕЕ влияние отображается линиями СИНЕГО цвета, толщина линии [приведенная в кружочке в центре линии] отражает относительную силу влияния.
Сортировка рецепторов по информативности
Отображается количество рецепторов не более: 999
Показаны связи с относительной силой влияния выше: 0%
Визуализация нейрона с кодами и наименованиями рецепторов

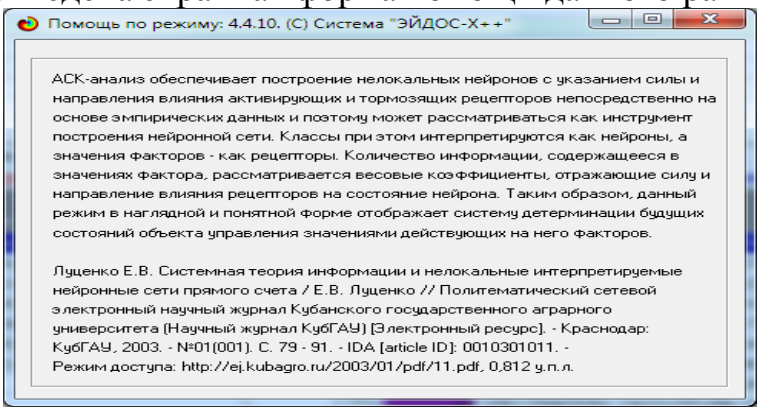

Neuron0004Inf1.jpg
НЕЛОКАЛЬНЫЙ НЕЙРОН В МОДЕЛИ: "INF1"
Нейрон: [4]-КЛАСС-4/5-{Roerich}
Приложение: Спектральный АСК-анализ изображений
[1] Класс
[4] 4/5-{Roerich}
Форма создана: 29.07.2021-09:26:11
Влияние рецепторов на активацию/торможение нелокального нейрона, соотвествующего классу (система детерминации класса):
АКТИВИРУЮЩЕЕ влияние отображается линиями КРАСНОГО цвета, толщина линии (приведенная в кружочке в центре линии) отражает относительную силу влияния.
ТОРМОЗЯЩЕЕ влияние отображается линиями СИНЕГО цвета, толщина линии (приведенная в кружочке в центре линии) отражает относительную силу влияния.
Сортировка рецепторов по информативности
Отображается количество рецепторов не более: 999
Показаны связи с относительной силой влияния выше: 0%
Визуализация нейрона с кодами и наименованиями рецепторов


Neuron0005Inf1.jpg
НЕЛОКАЛЬНЫЙ НЕЙРОН В МОДЕЛИ: "INF1"
Нейрон: [5]-КЛАСС-5/5-{Shishkin}
Приложение: Спектральный АСК-анализ изображений
[1] Класс
[5] 5/5-{Shishkin}
Форма создана: 29.07.2021-09:26:35
Влияние рецепторов на активацию/торможение нелокального нейрона, соотвествующего классу (система детерминации класса):
АКТИВИРУЮЩЕЕ влияние отображается линиями КРАСНОГО цвета, толщина линии (приведенная в кружочке в центре линии) отражает относительную силу влияния.
ТОРМОЗЯЩЕЕ влияние отображается линиями СИНЕГО цвета, толщина линии (приведенная в кружочке в центре линии) отражает относительную силу влияния.
Сортировка рецепторов по информативности
Отображается количество рецепторов не более: 999
Показаны связи с относительной силой влияния выше: 0%
Визуализация нейрона с кодами и наименованиями рецепторов


NeuroNet0002Inf1.jpg
ПАРЕТО-ПОДМНОЖЕСТВО НЕЛОКАЛЬНОЙ НЕЙРОННОЙ СЕТИ В МОДЕЛИ: "INF1"
Отображено: 100.00% наиболее значимых синаптических связей
Приложение: "Спектральный АСК-анализ изображений"
Нейроны:
{1, 5} =>
Рецепторы:
{1, 23} =>
Связи между рецепторами и нейронами. Цвет линии обозначает знак связи ('+','-'), а толщина линии - силу связи:
АКТИВИРУЮЩАЯ связь между рецептором и нейроном отображается КРАСНЫМ цветом
ТОРМОЗЯЩАЯ связь между рецептором и нейроном отображается СИНИМ цветом
Коды нач.н кон.нейронов: 1-5
Коды нач.н кон.рецепторов: 1-23
Отображать не более: 16 нейронов
Отображать не более: 23 рецепторов
Отображать не более: 1000 связей
Отображать связи с интенс.более: 0.000% от факт.макс.
Сортировка связей по модулю информативности
Дата и время создания формы: 29.07.2021-09:27:30

Отметим, что нелокальные нейроны соответствуют классам и по сути являются обобщенными определениями классов, т.е. обобщенными онтологиями, полученными путем обобщения конкретных онтологий. используя терминологию фреймовой модели Марвина Мински (1975, США), можно сказать, что это фреймы-прототипы, полученные путем обобщения относящихся к ним фреймов-экземпляров.

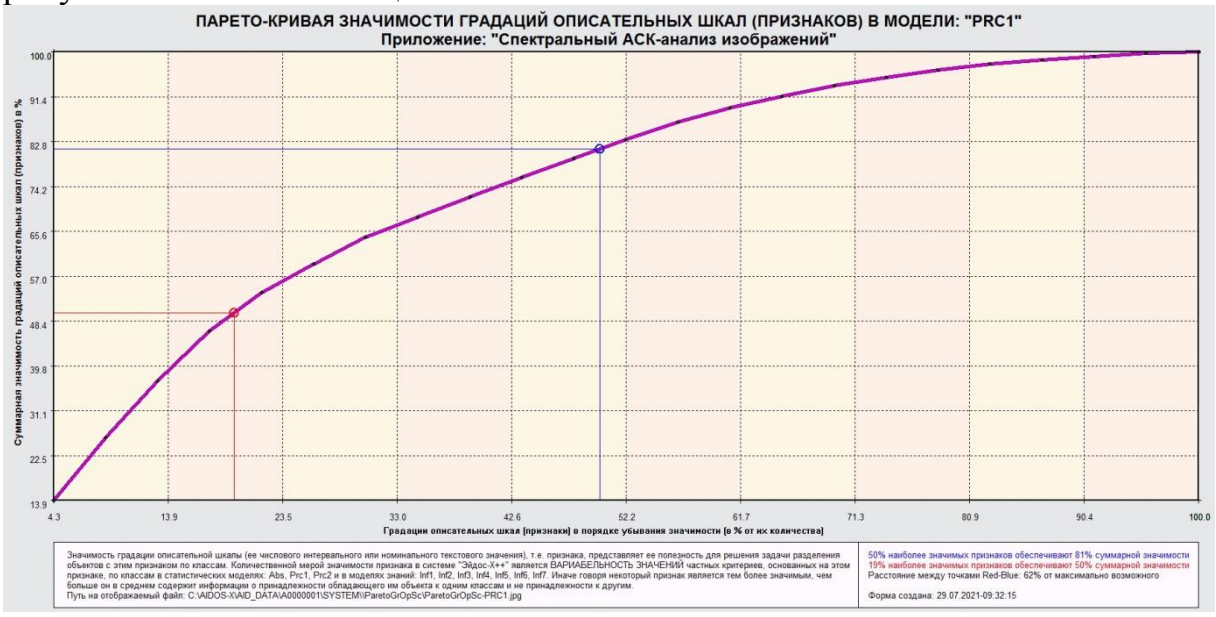
Ниже приведена экранная форма помощи данного ражима:



14.9.6. Ценность цветов для идентфкации изображений

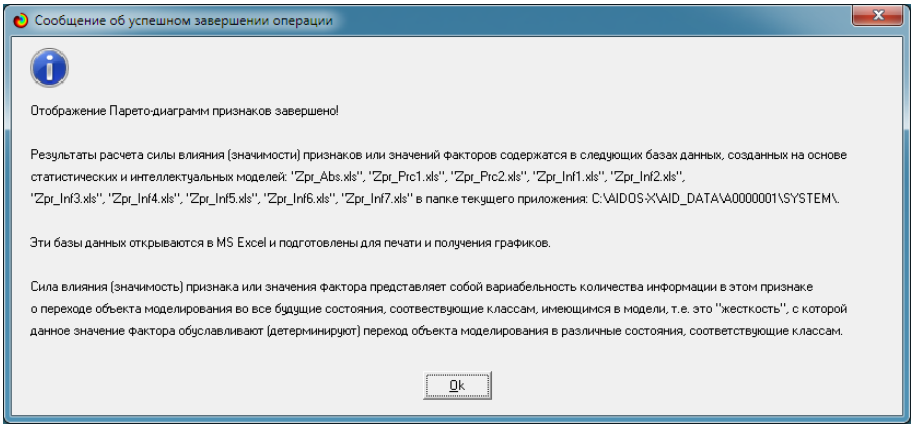
Конкретный цвет (цветовой диапазон) является тем более ценным (значимым) для идентификации конкретных картин с обобщенными образами классов, соответствующих художникам, чем сильнее отличается условная вероятность встретить этот цвет в картинах различных художников от безусловной вероятности его встречи по всей обучающей выборке картин, т.е. чем выше варибельность частного критерия в статистической или системно-когнитивной модели.

Если ранжировать (рассортировать) все цвета в порядке убывания их значимости и просуммировать эту значимость нарастающим итогом (см. таблицу), то получим логистическую парето-кривую, приведенную на рисунке ниже и в таблице.



Из рисунка мы видим, что 50% наиболее ценных для идентификации цветов вместе обеспечивают 81% суммарной значимости, а 50% суммарной значимости обеспечиваются лишь 19% наиболее значимых цветов.

Эта же информация представлена и в табличном виде:



Ценность цветов для идентификации изображений

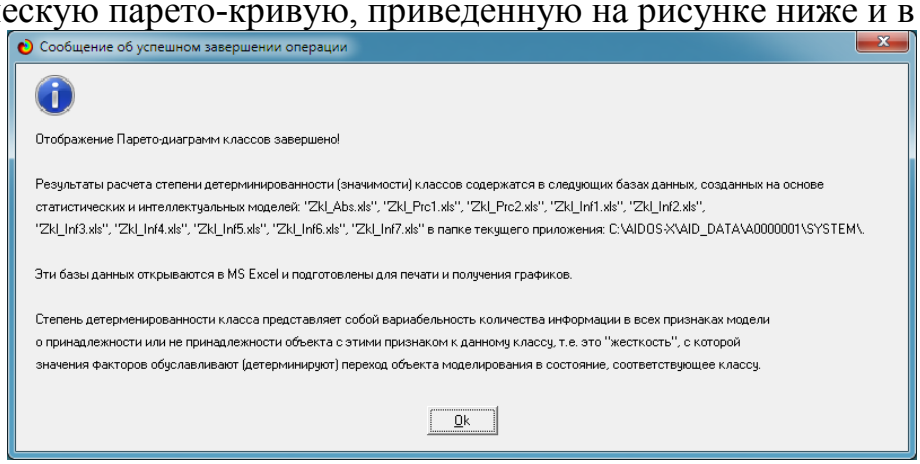
№	№%	Код	Наименование цвета (RGB)	Значимость, %	Значимость нарастающим итогом, %
1	4,348	6	SPECTRINTERV:-6/23-{153,222,006}	13,934	13,934
2	8,696	4	SPECTRINTERV:-4/23-{214,164,003}	12,128	26,062
3	13,043	3	SPECTRINTERV:-3/23-{236,130,015}	10,842	36,904
4	17,391	14	SPECTRINTERV:-14/23-{010,141,229}	9,581	46,485
5	21,739	5	SPECTRINTERV:-5/23-{186,196,000}	7,357	53,843
6	26,087	15	SPECTRINTERV:-15/23-{028,107,246}	5,421	59,264
7	30,435	8	SPECTRINTERV:-8/23-{084,252,044}	5,173	64,437
8	34,783	10	SPECTRINTERV:-10/23-{028,246,107}	3,931	68,369
9	39,130	11	SPECTRINTERV:-11/23-{010,229,141}	3,888	72,257
10	43,478	12	SPECTRINTERV:-12/23-{001,205,175}	3,771	76,028
11	47,826	2	SPECTRINTERV:-2/23-{250,095,036}	3,597	79,625
12	52,174	16	SPECTRINTERV:-16/23-{053,074,254}	3,572	83,196
13	56,522	13	SPECTRINTERV:-13/23-{001,175,205}	3,430	86,627
14	60,870	9	SPECTRINTERV:-9/23-{053,254,074}	2,643	89,270
15	65,217	17	SPECTRINTERV:-17/23-{084,044,252}	2,300	91,571
16	69,565	23	SPECTRINTERV:-23/23-{250,036,095}	2,027	93,598
17	73,913	21	SPECTRINTERV:-21/23-{214,003,164}	1,522	95,120
18	78,261	7	SPECTRINTERV:-7/23-{118,242,021}	1,439	96,559
19	82,609	19	SPECTRINTERV:-19/23-{153,006,222}	1,139	97,698
20	86,957	1	SPECTRINTERV:-1/23-{255,063,063}	0,831	98,529
21	91,304	20	SPECTRINTERV:-20/23-{186,000,196}	0,672	99,201
22	95,652	22	SPECTRINTERV:-22/23-{236,015,130}	0,627	99,828
23	100,000	18	SPECTRINTERV:-18/23-{118,021,242}	0,172	100,000

Из приведенной таблицы видно, что значимость наиболее и наименее ценных цветов отличается более чем в 81 раз, что очень существенно.

14.9.7. Степень детерминированности классов изображений цветами

Чем больше в обобщенном образе класса изображений доля цветов высокой значимости, тем безошибочнее с ним идентифицируются конкретные изображения по их спектрам, т.е. тем выше степень детерминированности классов изображений цветами.

Если ранжировать (рассортировать) все классы в порядке убывания степени их детерминированности цветами и просуммировать эту детерминированность нарастающим итогом (см. таблицу), то получим логистическую парето-кривую, приведенную на рисунке ниже и в таблице.



Степень дереминированносити классов цветами

№	№%	Код	Наименование класса	Значимость, %	Значимость нарастающим итогом, %
1	20,000	2	КЛАСС-2/5-{DaVinci}	22,589	22,589
2	40,000	5	КЛАСС-5/5-{Shishkin}	21,810	44,399
3	60,000	3	КЛАСС-3/5-{Kuindzhi}	20,097	64,496
4	80,000	1	КЛАСС-1/5-{Aivazovsky}	18,020	82,516
5	100,000	4	КЛАСС-4/5-{Roerich}	17,484	100,000

Из приведенной таблицы видно, что детерминированность классов отличается примерно на 30%, т.е. не очень существенно.

14.10. Выводы

Автоматизированный системно-когнитивный анализ (АСК-анализ) изображений обеспечивает автоматическое выявление признаков конкретных изображений из цветов пикселей и контуров изображений, синтез обобщенных образов изображений (классов), выявление наиболее характерных и нехарактерных для классов признаков изображений,

определение ценности признаков изображений для их различения, удаление из модели малоценных признаков (абстрагирование), решение задач количественного сравнения конкретных изображений с обобщенными образами классов и обобщенных образов классов друг с другом, а также задачи исследования моделируемой предметной области путем исследования ее модели. В работе рассматриваются новые возможности АСК-анализа и реализующей его интеллектуальной системы «Эйдос», обеспечивающие выявление признаков изображений путем их спектрального анализа, формирования обобщенных спектров классов, решение задач сравнения изображений конкретных объектов с классами и классов друг с другом по их спектрам. Впервые стало возможным формировать обобщенные спектры классов с весами цветов по степени их характерности и нехарактерности для классов, причем это не интенсивность цвета в спектре, а количество информации в цвете о принадлежности объекта с этим цветом к данному классу. По сути, речь идет об обобщении спектрального анализа путем применении интеллектуальных когнитивных технологий и теории информации в спектральном анализе. Во-первых, все говорят о том, что в спектральных линиях содержится информация о том, какой элемент или вещество входят в состав объекта, но никто не удосужился посчитать какое же это конкретно количество этой информации, а затем использовать его для определения состава объекта методы распознавания образов, основанные на использовании этой информации. Во-вторых, спектральный анализ традиционно используется для определения элементарного и молекулярного состава объекта, а мы предлагаем использовать его не только для этого, но и для идентификации любых изображений.

14.11. Возможные области применения и перспективы

Классический спектральный анализ традиционно применяется для определения элементного и химического состава веществ и различных объектов по спектрам их электромагнитного (или другого) излучения или поглощения без проведения химического анализа, а также для решения ряда других задач в различных предметных областях путем спектрального анализа изображений.

Например, программа: «Спектр анализатор 1.07»[40] обеспечивает:

1. Определение расовой составляющей человека, т.е. какие расы были в роду человека и каков их процент.

2. Экологический мониторинг. Определение чистоты воды, воздуха и т.д.

[40] См., например:
http://www.softportal.com/software-19743-spektr-analizator.html
http://monobit.ru/spektr-analizator.html

3. Мониторинг здоровья - выявление заболеваний, мониторинг состояния своего здоровья.

4. Анализ сетчатки глаза. Мониторинг своего состояния по изменению спектра сетчатки глаза.

5. Научные исследования. Исследование состава пород, камней, земли. Исследование растений и насекомых.

6. Позволяет вести статистику для оценки динамики изменения цветов в каком-либо объекте.

Для всех тех же целей может быть использована и предлагаемая технология в системе «Эйдос».

Но с тем отличием, что, например атомные спектры излучения химических элементов получены ведущими учеными мира путем их длительного и тщательного изучения в течение десятков лет. На знании этих спектров основаны все спектральные анализаторы, в которые эти спектры внесены в качестве образцов для сравнения при разработке и создании этих спектральных анализаторов.

В данной же работе предлагается технология, обеспечивающая:

– как создание подобных моделей на основе еще неизученных учеными образцов;

– так и применение этих моделей для идентификации новых образцов, не использованных при создании моделей.

Пример 1. Теперь же можно кинуть в костер по очереди несколько порошков различных элементов (меди, железа и т.д.) или химических соединений и сфотографировать как меняется цвет пламени. После этого создать модель на основе этих фотографий и использовать ее для определения по цвету пламени, также по фотографиям, какова доля этих элементов или соединений в смесях, которые кидают в костер. Важно, что так можно вполне успешно решать задачу идентификации даже не зная названий этих элементов или соединений и не зная являются ли они именно элементами или соединения, и тем более не зная спектральный состав их излучения.

Пример 2. Предъявляем системе изображения листьев определенного сорта растений с различной степенью повреждения определенным видом вредителя. В наименованиях файлов изображений в качестве классов указываем степень выраженности поражения площади листа вредителем в процентах. По каждой степени выраженности поражения может быть приведено несколько примеров. Для этого после имени класса необходимо указать черточку и номер примера. На основе обучающей выборки системой «Эйдос» создаются обобщенные образы классов по каждой степени выраженности поражения. После этого мы можем ввести новые изображения в систему в режиме идентификации и система покажет все классы по убыванию релевантности. На основе этих результатов можно обоснованно судить о степени повреждения [45].

По сути, речь идет о технологии создания интеллектуальных измерительных систем самых различных предметных областях [44]. Предлагается применить автоматизированный системно-когнитивный анализ (АСК-анализ) и его программный инструментарий систему «Эйдос» как для синтеза, так и для применения адаптивных интеллектуальных измерительных систем с целью измерения не значений параметров объектов, а для системной идентификации состояний сложных многофакторных нелинейных динамических систем по их спектрам. Кратко рассматривается математический метод АСК-анализа, реализованный в его программном инструментарии – универсальной когнитивной аналитической системе «Эйдос-X++». Математический метод АСК-анализа основан на системной теории информации (СТИ), которая создана в рамках реализации программной идеи обобщения всех понятий математики, в частности - теории информации, базирующихся на теории множеств, путем тотальной замены понятия множества на более общее понятие системы и тщательного отслеживания всех последствий этой замены. Благодаря математическому методу, положенному в основу АСК-анализа, этот метод является непараметрическим и позволяет сопоставимо обрабатывать десятки и сотни тысяч градаций факторов и будущих состояний объекта управления (классов) при неполных (фрагментированных), зашумленных данных числовой и нечисловой природы измеряемых в различных единицах измерения. Приводится развернутый численный пример применения АСК-анализа и системы «Эйдос-X++» как для синтеза системно-когнитивной модели, обеспечивающей многопараметрическую типизацию состояний сложных систем, так и для системной идентификации их состояний, а также для принятии решений об управляющем воздействии, так изменяющем состав объекта управления, чтобы его качество (уровень системности) максимально повышалось при минимальных затратах на это. Для численного примера в работе приняты картины известных художников. Однако необходимо отметить, что этот пример следует рассматривать шире, т.к. АСК-анализ и система «Эйдос» разрабатывались и реализовались в очень обобщенной постановке, не зависящей от предметной области, и с успехом могут быть применены в самых различных предметных областях.

Скачать систему «Эйдос-X++» (самую новую на текущий момент версию) всегда можно на сайте автора по ссылке: http://lc.kubagro.ru/aidos/_Aidos-X.htm. Это наиболее полная на данный момент незащищенная от несанкционированного копирования портативная (portable) версия системы (не требующая инсталляции) с исходными текстами, находящаяся в полном открытом бесплатном доступе (около 140 Мб). Обновление имеет объем около 10 Мб.

Численный пример для данного раздела работы можно загрузить из Эйдос-облака в режиме 1.3 приложение № **277**.

ГЛАВА 15. АВТОМАТИЗИРОВАННЫЙ СИСТЕМНО-КОГНИТИВНЫЙ АНАЛИЗ ТЕКСТОВ

Из-за жестких ограничений на объем данной монографии ниже приведем лишь краткую инфолрмацию о примении АСК-анаиза для интеллекутального анализа текстов.

15.1. Синтез семантических ядер научных специальностей ВАК РФ и автоматическая классификации статей по научным специальностям с применением АСК-анализа и интеллектуальной системы «Эйдос»

14 января 2019 года на сайте ВАК РФ http://vak.ed.gov.ru/87 появилась информация: «Об уточнении научных специальностей и соответствующих им отраслей науки, по которым издания входят в Перечень рецензируемых научных изданий, в которых должны быть опубликованы основные научные результаты диссертаций на соискание ученой степени кандидата наук, на соискание ученой степени доктора наук». Сообщается, что согласно рекомендации ВАК для остальных изданий, входящих в Перечень по группам научных специальностей, работа по уточнению научных специальностей и отраслей науки будет продолжена в 2019 году. Данная работа является продолжением серии работ автора по когнитивной лингвистике. В ней предлагается инновационная интеллектуальная технология для автоматизации решения задачи, сформулированной ВАК РФ выше. С применением автоматизированного системно-когнитивного анализа (АСК-анализ) и его программного инструментария – интеллектуальной системы «Эйдос» непосредственно на основе официальных текстов паспортов научных специальностей ВАК РФ созданы их семантические ядра, а затем реализована автоматическая классификация научных текстов (статей, монографий, учебных пособий и т.д.) по специальностям и группам специальностей ВАК РФ. Традиционно эта задача решается диссертационными советами, а также редакционными советами научных изданий, т.е. экспертами, на основе экспертных оценок, неформализованным путем, на основе опыта, интуиции и профессиональной компетенции. Однако, традиционный подход имеет ряд довольно серьезных недостатков, накладывающих на качество и объемы анализа существенные ограничения. Следовательно, актуальными является усилия исследователей и разработчиков по преодолению этих ограничений. В настоящее время уже есть все основания рассматривать эти ограничения как неприемлемые, т.к. их не только нужно, но и вполне

возможно преодолеть. Таким образом, налицо проблема, решение которой и являются предметом рассмотрения в данной статье. Приводится развернутый численный пример решения поставленной проблемы на реальных данных [1].

15.2. Формирование семантического ядра ветеринарии путем Автоматизированного системно-когнитивного анализа паспортов научных специальностей ВАК РФ и автоматическая классификация текстов по направлениям науки

Настоящее время характеризуется появлением в открытом доступе огромных объемов текстов на различных языках, сгенерированных людьми. В настоящее время эти тексты накапливаются в различных электронных библиотеках и библиографических базах данных (WoS, Скопус, РИНЦ и др), а также просто в Internet на различных сайтах. Все эти тексты имеют конкретных авторов, датировку и могут относиться одновременно ко многим не альтернативным категориям и жанрам, в частности: учебные; научные; художественные; политические; новостные; чаты; форумы и многие другие. Большой научный и практический интерес представляет решение обобщенной задачи атрибуции текстов, т.е. такого исследования этих текстов, при котором определялись бы их вероятные авторы, датировка создания, принадлежность этих текстов к перечисленным выше обобщенным группам или жанрам, а также оценка сходства- различия авторов и текстов по их содержанию, выделение в текстах ключевых слов и т.п. и т.д. Для решения всех этих задач необходимо сформировать обобщенные лингвистические образы текстов по группам (классам), т.е. сформировать семантические ядра классов. Частным случаем этой задачи является создание семантических ядер по различным научным специальностям ВАК РФ и автоматическая классификация научных текстов по направлениям науки. Традиционно эта задача решается диссертационными советами, т.е. экспертами, на основе экспертных оценок, т.е. неформализованным путем, на основе опыта, интуиции и профессиональной компетенции. Однако традиционный подход имеет ряд довольно серьезных недостатков, накладывающих на качество и объемы анализа существенные ограничения. В настоящее время уже есть все основания рассматривать эти ограничения как неприемлемые, т.к. их вполне можно преодолеть. Таким образом, налицо проблема, пути решения которой и являются предметом рассмотрения в данной статье. Следовательно, актуальными является усилия исследователей и разработчиков по их преодолению. Поэтому целью работы является разработка автоматизированной технологии (метода и инструментария), а также методики их применения для формирования семантического ядра ветеринарии путем автоматизированного системно-когнитивного анализа

паспортов научных специальностей ВАК РФ и автоматической классификация текстов по направлениям науки. Приводится развернутый численный пример решения поставленной проблемы на реальных данных [2].

15.3. Интеллектуальная привязка некорректных ссылок к литературным источникам в библиографических базах данных с применением АСК-анализа и системы «Эйдос»

Адекватная и технологичная оценка результативности, эффективности и качества научной деятельности конкретных ученых и научных коллективов является актуальной проблемой для информационного общества и общества, основанного на знаниях. Решение этой проблемы является предметом наукометрии и ее целью. Современный этап развития наукометрии существенно отличается от предыдущих появлением в открытом, а также платном on-line доступе огромного объема детализированных данных по большому числу показателей как об отдельных авторах, так и о научных организациях и вузах. В мире, это известные библиографические базы данных: Web of Science, Scopus, Astrophysics Data System, PubMed, MathSciNet, zbMATH, Chemical Abstracts, Springer, Agris или GeoRef. В России это прежде всего Российский индекс научного цитирования (РИНЦ). РИНЦ – это национальная информационно-аналитическая система, аккумулирующая более 9 миллионов публикаций российских ученых, а также информацию о цитировании этих публикаций из более 6000 российских журналов. Данных очень много, это так называемые «Большие данные» ("Big Data"). Основным первичным наукометрическим показателем, на основе которого строятся все остальные, такие, например, как индекс Хирша, является число цитирований работ автора, размещенных в библиографической базе данных. Это число цитирований определяется программным обеспечением РИНЦ путем так называемой «привязки», которая представляет собой грамматический разбор и поиск в базах данных работ автора, релевантных (соответствующих) ссылкам на них из источников литературы в работах различных авторов. Однако проблема состоит в том, что, как показывает опыт, авторы допускают очень большое количество некорректных и просто неполных ссылок в списках литературы, очень далеких от ГОСТ. В настоящее время программное обеспечение РИНЦ не может автоматически привязать эти некорректные ссылки и это требует вмешательства человека. Но централизованно, силами специалистов РИНЦ, это сделать не представляется возможным из-за огромного объема работ, а распределенная работа большого числа специалистов на местах все равно требует централизованной модерации. В результате работа по привязке ссылок к литературным источникам ведется очень медленно и

огромный объем ссылок оказывается непривязанными. Это ведет к занижению накометрических показателей как отдельных авторов, так и научных коллективов, что нельзя признать приемлемым. Решение этой проблемы предлагается путем применения автоматизированного системно-когнитивного анализа (АСК-анализ) и его программного инструментария – интеллектуальной системы «Эйдос». Приводится численный пример интеллектуальной привязки реальных некорректных ссылок к работам автора на основе небольшого объема реальных наукометрических данных, находящихся в открытом бесплатном on-line доступе в РИНЦ [3].

15.4. Применение АСК-анализа и интеллектуальной системы "Эйдос" для решения в общем виде задачи идентификации литературных источников и авторов по стандартным, нестандартным и некорректным библиографическим описаниям

Проблемы идентификации авторов и литературных источников по библиографическим описаниям в списках литературы в последнее время приобретает все большее значение научное и практическое значение. Это связано в частности с политикой Министерства образования и науки Российской Федерации в области оценки качества результатов научной деятельности, которая предполагает использование количества ссылок на публикации авторов и индекса Хирша. В России создаются соответствующие аналитические инструменты и сервисы для оценки результатов научной деятельности, функционально аналогичные известным зарубежным библиографическим базам данных Scopus, Web of Science и другим. В настоящее время наиболее известным в России сервисом подобного назначения является Российский индекс научного цитирования (РИНЦ): http://elibrary.ru/. Однако, как показывает опыт, часто ссылки в списках литературы публикаций сделаны с нарушением ГОСТ 7.1—2003, а также с ошибочными выходными данными, например, неверно указанными номерами страниц, наименованием издательства и т.п. На практике это приводит к тому, что программная система библиографической базы не может определить, на какую статью сделана данная ссылка и кто авторы этой статьи. В результате для этих авторов теряется цитирование, что приводит к занижению их индексов Хирша и оценки результатов их научной деятельности руководством. Понятно, что эти отрицательные последствия желательно преодолеть. Данная статья посвящена изложению подхода, который позволяет решить эту проблему путем применения АСК-анализа и интеллектуальной системы «Эйдос», представляющих собой современную инновационную интеллектуальную технологию (готовую к внедрению) [4].

15.5. АСК-анализ проблематики статей Научного журнала КубГАУ в динамике

В связи с выходом юбилейного 100-го номера Научного журнала КубГАУ было проведено исследование динамики проблематики научных исследований по публикациям в журнале. В качестве инструментов данного исследования применены автоматизированный системно-когнитивный анализ (АСК-анализ) и его программный инструментарий – Универсальная когнитивная аналитическая система «Эйдос-X++» [5].

15.6. Атрибуция анонимных и псевдонимных текстов в системно-когнитивном анализе

В данной статье исследуется возможность атрибуции текстов с применением технологии и инструментария системно-когнитивного анализа. Приведен подробный численный пример реализации всех этапов СК-анализа при атрибуции текстов, т. е. когнитивной структуризации и формализации предметной области; формирования обучающей выборки; синтеза семантической информационной модели; ее оптимизации и измерения адекватности; адаптации и пересинтеза; а также типологического и кластерно-конструктивного анализа. Для специалистов по атрибуции и контент-анализу текстов на естественном языке. Материал может быть использован в качестве руководства к лабораторной работе по дисциплине: "Интеллектуальные информационные системы" [6].

15.7. Атрибуция текстов, как обобщенная задача идентификации и прогнозирования

Вербальные описания объектов на естественном языке рассматриваются в статье как их иерархические лингвистические модели. Предлагаются методика и автоматизированная технология, базирующиеся на универсальной когнитивной аналитической системе "Эйдос" и обеспечивающие: автоматизированную формализацию предметной области на основе вербального описания ее объектов, автоматизированное формирование описательных шкал и градаций, а также обучающей выборки, синтез семантической информационной модели, ее оптимизацию, проверку адекватности и анализ [7].

15.8. Интеллектуальная датировка текста, определение авторства и жанра на примере русской литературы XIX и XX веков

С развитием интеллектуальных технологий появилось целое новое научное направление по их применению для атрибуции и наратологического анализа литературных текстов. Есть попытки автоматического определения авторства, датировки и жанра литературных произведений. Однако научные исследования и разработки в этой важной

области посвящены в основном разработке концептуальных подходов и математических моделей, тогда как для конкретных исследователей важно иметь реализующий эти модели программный инструментарий. В данной работе предлагается новая математическая модель, основанная на теории информации, а также соответствующая методика численных расчетов и реализующий их программный инструментарий для автоматической атрибуции и элементов наратологического анализа литературных текстов. Данная математическая модель разработана в новационном методе искусственного интеллекта: Автоматизированном системно-когнитивного анализе (АСК-анализ) и реализована в его программном инструментарии – интеллектуальной системе «Эйдос». Приводится численный пример с большим количеством выходных форм, основанный на реальных текстах [8].

15.9. Intellectual attribution of literary texts (finding the dates of the text, determining authorship and genre on the example of Russian literature of the XIX and XX centuries)

With the development of intelligent technologies, a whole new scientific direction has appeared aimed at their application for attribution and naratological analysis of literary texts. There are attempts to determine the authorship automatically, along with dates and genre of literary works. This article proposes a new mathematical model based on the theory of information, as well as a corresponding method of numerical calculations and a software tool, implementing them for automatic attribution and elements of naratological analysis of literary texts. This mathematical model is developed in an innovative method of artificial intelligence: the automated system-cognitive analysis (ASC-analysis) and a tool implemented in its software which is an intelligent system called "Eidos". We have obtained good results in testing the proposed approach on a real numerical example of Russian literature of the XIX and XX centuries. In the Eidos intellectual system, which implements the proposed mathematical model, a large number of text and table output forms are issued that provide interpretation of the results obtained. Thus, the proposed mathematical model and the software system implementing it can be successfully applied for finding dates of literary texts, determining their authorship and genre. This may be performed with texts in any language [9].

15.10. Выводы

По результатам краткого обзора применения АСК-анализа для интеллектуального анализа текстов, проведенного в данной главе, можно сделать обоснованный вывод о том, что данный метод является вполне адекватным для применения для этих целей, т.к. позволяет:

- формировать обобщенные лингвистические образы классов (семантические ядра) на основе фрагментов или примеров относящихся к ним текстов на любом языке;'

- количественно сравнивать лингвистический образ конкретного человека, или описание объекта, процесса с обобщенными лингвистическими образами групп (классов);

- сравнивать обобщенные лингвистические образы классов друг с другом и создавать их кластеры и конструкты;

- исследовать моделируемую предметную область путем исследования ее лингвистической системно-когнитивной модели;

- проводить интеллектуальную атрибуцию текстов, т.е. определять вероятное авторство анонимных и псевдонимных текстов, датировку, жанр и смысловую направленность содержания текстов;

- все это можно делать для любого естественного или искусственного языка или системы кодирования.

Ссылки на работы автора по теккстовому АСК-анализу размещены здесь: http://lc.kubagro.ru/aidos/Works_on_ASK-analysis_of_texts.htm.

ЗАЛЮЧЕНИЕ

Включенные в настоящую книгу научные результаты наглядно демонстрируют большое теоретическое и прикладное значение идей и подходов системной нечеткой интервальной математики. Эта новая область теоретической и прикладной математики позволяет успешно решать задачи различных предметных областей - экономики (прежде всего цифровой), искусственного интеллекта, управления (менеджмента), техники и технологий, кибернетики, информатики, химии, биологии, социологии, медицины, психологии, истории и др., практически всех предметных областей. Так, организационно-экономическое, математическое и программное обеспечение контроллинга, инноваций и менеджмента основано на идеях, подходах и результатах системной нечеткой интервальной математики.

Констатируем, что точки роста современной математики в большинстве случаев относятся именно к системной нечеткой интервальной математике, на ее основе разработана новая парадигма математических методов исследования. Поэтому мы обоснованно полагаем, что системная нечеткая интервальная математика - основа математики XXI века.

Основные научные результаты системной нечеткой интервальной математики должны быть включены в учебные планы обучения бакалавров, магистров, аспирантов, слушателей бизнес-школ, систем

переподготовки и других образовательных структур. В своих учебниках мы демонстрируем, как это можно сделать.

В настоящую книгу включена лишь наиболее принципиально важная и актуальная часть научных результатов авторов в области системной нечеткой интервальной математики, полученных после выхода в 2014 г. нашей предыдущей книги по этой тематике.

Желающим расширить свое знакомство с этой быстро растущей областью современной математики рекомендуем обратиться к публикациям авторов.

С ними можно ознакомиться в Российском индексе научного цитирования (РИНЦ):

– https://www.elibrary.ru/author_profile.asp?id=1844;

– https://www.elibrary.ru/author_profile.asp?id=123162;

в "Политематическом сетевом электронном научном журнале Кубанского государственного аграрного университета (Научном журнале КубГАУ)":

– *http://ej.kubagro.ru/a/viewaut.asp?id=2744*;

– *http://ej.kubagro.ru/a/viewaut.asp?id=11*,

а также на сайтах авторов:

– *https://orlovs.pp.ru/* (*https://orlovs.pp.ru/work/*)

– *http://lc.kubagro.ru/* (*http://lc.kubagro.ru/aidos/_Aidos-X.htm*)

и на страничках авторов в РесечГейт:

– *https://www.researchgate.net/profile/Alexandr-Orlov-6*;

– *https://www.researchgate.net/profile/Eugene-Lutsenko*.

Многие (практически все) разделы системной нечеткой интервальной математики заслуживают дальнейшего развития и практического применения. Приглашаем исследователей различных специальностей активно участвовать в этой работе.

Авторы

13 января 2022 г.
Москва-Краснодар

ЛИТЕРАТУРА

Литература к главе 1

1. Орлов А.И. Новая парадигма математических методов исследования // Заводская лаборатория. Диагностика материалов. 2015. Т.81. №.7 С. 5-5.

2. Орлов А.И. Новая парадигма разработки и преподавания организационно-экономического моделирования, эконометрики и статистики в техническом университете // Статистика и прикладные исследования: сборник трудов Всерос. научн. конф. – Краснодар: Издательство КубГАУ, 2011. – С.131-144.

3. Орлов А.И. Организационно-экономическое моделирование, эконометрика и статистика в техническом университете // Вестник МГТУ им. Н.Э. Баумана. Сер. «Естественные науки». 2012. №1. С. 106-118.

4. Орлов А.И. Новая парадигма организационно-экономического моделирования, эконометрики и статистики // Вторые Чарновские Чтения. Сборник тезисов. Материалы II международной научной конференции по организации производства. Москва, 7 – 8 декабря 2012 г. – М.: НП «Объединение контроллеров», 2012. – С. 116-120.

5. Орлов А.И. Организационно-экономическое моделирование, эконометрика и статистика при решении задач экономики и организации производства // Инженерный журнал: наука и инновации, 2014, вып. 1.

6. Орлов А.И. Новая парадигма прикладной статистики // Статистика и прикладные исследования: сборник трудов Всерос. научн. конф. – Краснодар: Издательство КубГАУ, 2011. – С.206-217.

7. Орлов А.И. Новая парадигма прикладной статистики // Заводская лаборатория. Диагностика материалов. 2012. Т. 78. №1, часть I. С.87-93.

8. Орлов А.И. Новая парадигма математической статистики // Материалы республиканской научно-практической конференции «Статистика и её применения – 2012». Под редакцией профессора А.А. Абдушукурова. – Ташкент: НУУз, 2012. – С.21-36.

9. Орлов А.И. Основные черты новой парадигмы математической статистики // Научный журнал КубГАУ. 2013. № 90. С. 45-71.

10. Орлов А.И. Новая парадигма математических методов экономики // Экономический анализ: теория и практика. – 2013. – № 36 (339). – С.25–30.

11. Орлов А.И. Новая парадигма анализа статистических и экспертных данных в задачах экономики и управления // Научный журнал КубГАУ. 2014. № 98. С. 1254-1260.

12. Орлов А.И. Новая парадигма анализа статистических и экспертных данных в задачах управления // Труды X Международной конференции «Идентификация систем и задачи управления» SICPRO'15. Москва, 26-29 января 2015 г. М.: Институт проблем управления им. В.А. Трапезникова, 2015. - С.34 - 42.

13. Орлов А.И. Устойчивые экономико-математические методы и модели. Разработка и развитие устойчивых экономико-математических методов и моделей для модернизации управления предприятиями. – Saarbrucken: Lambert Academic Publishing, 2011. – 436 с.

14. Кун Т. Структура научных революций. – М.: АСТ, 2009. – 320 с.

15. Орлов А.И. Организационно-экономическое моделирование. Ч.1. Нечисловая статистика. - М.: Изд-во МГТУ им. Н.Э. Баумана, 2009. - 541 с.

16. Лопатников Л.И. Экономико-математический словарь: Словарь современной экономической науки. — 5-е изд., перераб. и доп. — М.: Дело, 2003. — 520 с.

17. Большой Энциклопедический Словарь. – М.: Большая Российская Энциклопедия, 1997. – 1600 с.

18. Орлов А.И. Эконометрика. - М.: Экзамен, 2002 (1-е изд.), 2003 (2-е изд.), 2004 (3-е изд.). - 576 с.

19. Новая философская энциклопедия. В 4-х томах. Под редакцией В. С. Стёпина. – М. : Мысль, 2009.

20. Орлов А.И. Прикладная статистика. - М.: Экзамен, 2006. - 671 с.

21. Вторые Чарновские чтения. Сборник трудов. Материалы II международной научной конференции по организации производства. Москва, 7 – 8 декабря 2012 г. – М.: НП «Объединение контроллеров», 2013. –201 с.

22. Организация и планирование машиностроительного производства (производственный менеджмент) / Под ред. Ю.В. Скворцова, Л.А. Некрасова. -М.: Высшая школа, 2003. – 470 с.

23. Орлов А.И., Орлова Л.А. Применение эконометрических методов при решении задач контроллинга // Контроллинг. 2003. № 4(8). С.50-54.

24. Хрусталёв Е.Ю., Хрусталёв О.Е. Когнитивное моделирование развития наукоемкой промышленности (на примере оборонно-промышленного комплекса) // Экономический анализ: теория и практика. 2013. № 10 (313). С. 2 – 10.

25. Математическое моделирование процессов налогообложения (подходы к проблеме) (совместно с В. Г. Кольцовым, Н.Ю. Ивановой и др.). — М.: Изд-во ЦЭО Министерства общего и профессионального образования РФ, 1997. — 232 с.

26. Орлов А.И. Теория принятия решений. — М.: Экзамен, 2006. — 574 с.

27. Хрусталёв Е.Ю., Хрусталёв О.Е. Модельное обоснование инновационного развития наукоемкого сектора российской экономики // Экономический анализ: теория и практика. 2013. № 9 (312). С. 2 – 13.

28. Михненко П.А. Методология математического моделирования организационных изменений // Экономический анализ: теория и практика. 2013. № 26 (329). С. 40 – 48.

29. Карпычев В.Ю. Информационные технологии в экономических исследованиях // Экономический анализ: теория и практика. 2013. №20 (323). С. 2 – 11.

30. Рощин А.В., Тихонов И.П., Проничкин С.В. Методический подход к оценке эффективности результатов научно-технических программ // Экономический анализ: теория и практика. 2013. № 21 (324). С. 10 – 18.

31. Орлов А.И. Организационно-экономическое моделирование. Ч.2. Экспертные оценки. – М.: Изд-во МГТУ им. Н.Э. Баумана, 2011. – 486 с.

32. Орлов А.И. Теория экспертных оценок в нашей стране // Научный журнал КубГАУ. 2013. № 93. С. 1-11.

33. Демидов Я.П. Теория и практика современного рейтингования: критические заметки// Экономический анализ: теория и практика. – 2013. –№ 8 (311). – С. 14 – 19.

34. Лындина М.И., Орлов А.И. Математическая теория рейтингов // Научный журнал КубГАУ. 2015. № 114. С. 1 – 26.

35. Семенов С.С., Харчев В.Н., Иоффин А.И. Оценка технического уровня образцов вооружения и военной техники. - М.: Радио и связь, 2004. - 552 с.

36. Семенов С.С. Оценка качества и технического уровня сложных систем: Практика применения метода экспертных оценок. - М.: ЛЕНАНД, 2015. - 352 с.

37. Орлов А.И. Рецензия первая. Теория принятия решений, экспертные оценки и технический уровень сложных технических систем // Семенов С.С. Оценка качества и технического уровня сложных систем: Практика применения метода экспертных оценок. - М.: ЛЕНАНД, 2015. - С.18 - 24.

38. Дутов А.В., Калинин И.М. Формирование научно-технического задела в судостроении. - СПб.: ФГУП "Крыловский государственный научный центр", 2013. - 308 с.

39. Захаров М.Н., Омельченко И.Н., Саркисов А.С. Ситуации инженерно-экономического анализа. - М.: Издательство МГТУ им. Н.Э. Баумана, 2014. - 430 с.

40. Семенов С.С., Воронов Е.М., Полтавский А.В., Крянев А.В. Методы принятия решений в задачах оценки качества и технического уровня сложных технических систем. - М.: ЛЕНАНД, 2016. - 520 с.

41. Семенов С.С., Щербинин В.В. Оценка технического уровня систем наведения управляемых авиационных бомб. - М.: Машиностроение, 2015. - 326 с.

42. Орлов А.И. Создана единая статистическая ассоциация // Вестник Академии наук СССР. – 1991. – №7. – С. 152 – 153.

43. Бернштейн С.Н. Современное состояние теории вероятностей и ее приложений // Труды Всероссийского съезда математиков в Москве 27 апреля – 4 мая 1927 г. – М.-Л.: ГИЗ, 1928. – С. 50 – 63.

44. Орлов А.И. Распределения реальных статистических данных не являются нормальными // Научный журнал КубГАУ. 2016. № 117. С. 71–90.

45. Орлов А.И. О развитии статистики объектов нечисловой природы // Научный журнал КубГАУ. 2013. № 93. С. 41-50.

46. Новиков А.М., Новиков Д.А. Методология. – М.: СИНТЕГ, 2007. – 668 с.

47. Орлов А.И., Луценко Е.В. Системная нечеткая интервальная математика. Монография (научное издание). – Краснодар, КубГАУ. 2014. – 600 с.

48. Орлов А.И. Новый подход к изучению устойчивости выводов в математических моделях // Научный журнал КубГАУ. 2014. № 100. С. 146-176.

49. Орлов А.И. Компьютерно-статистические методы: состояние и перспективы // Научный журнал КубГАУ. 2014. № 103. С. 163 – 195.

50. Орлов А.И. Взаимосвязь предельных теорем и метода Монте-Карло // Научный журнал КубГАУ. 2015. № 114. С. 27–41.

51. Орлов А.И. Высокие статистические технологии // Заводская лаборатория. – 2003. – Т.69. – №11. – С. 55 – 60.

52. Орлов А.И. О развитии методологии статистических методов // Статистические методы оценивания и проверки гипотез. Межвузовский сборник научных трудов. – Пермь: Изд-во Пермского государственного университета, 2001. – С. 118 – 131.

53. Орлов А.И. О методологии статистических методов // Научный журнал КубГАУ. 2014. № 104. С. 53 – 80.

54. Орлов А.И. Эконометрика. Изд. 4-е, доп. и перераб. – Ростов-на-Дону: Феникс, 2009. – 572 с.

55. Орлов А.И. Принятие решений. Теория и методы разработки управленческих решений. М.: – ИКЦ «МарТ»; Ростов н/Д: Издательский центр «МарТ», 2005. – 496 с.

56. Колобов А.А., Омельченко И.Н., Орлов А.И. Менеджмент высоких технологий. Интегрированные производственно-корпоративные структуры: организация, экономика, управление, проектирование, эффективность, устойчивость. — М.: Экзамен, 2008. — 621 с.

57. Орлов А.И. Организационно-экономическое моделирование. Ч.3. Статистические методы анализа данных. — М.: Изд-во МГТУ им. Н.Э. Баумана, 2012. — 624 с.

58. Орлов А.И. Менеджмент: организационно-экономическое моделирование. — Ростов-на-Дону: Феникс, 2009. — 475 с.

59. Орлов А.И. Организационно-экономическое моделирование: теория принятия решений. — М. : КноРус, 2011. — 568 с.

60. Орлов А.И. Вероятность и прикладная статистика: основные факты: справочник. – М.: КноРус, 2010. – 192 с.

61. Орлов А.И., Федосеев В.Н. Менеджмент в техносфере. – М.: Академия, 2003. – 384 с.

62. Орлов А.И. Проблемы управления экологической безопасностью. Итоги двадцати лет научных исследований и преподавания. – Saarbrücken: Palmarium Academic Publishing. 2012. – 344 с.

63. Орлов А.И. Оптимальные методы в экономике и управлении. Учебное пособие. — М.: Изд-во МГТУ им. Н.Э. Баумана, 2007. — 44 с.

64. Орлов А.И. Эконометрика : учебное пособие. — М., Саратов : Интернет-Университет Информационных Технологий (ИНТУИТ), Ай Пи Ар Медиа, 2020. — 676 с.

65. Орлов А.И. Искусственный интеллект: нечисловая статистика : учебник. — М.: : Ай Пи Ар Медиа, 2022. — 446 с.

66. Орлов А.И. Искусственный интеллект: статистические методы анализа данных : учебник. — М.: Ай Пи Ар Медиа, 2022. — 843 с.

67. Орлов А.И. Искусственный интеллект: экспертные оценки : учебник. — М.: : Ай Пи Ар Медиа, 2022. — 436 с.

68. Орлов А.И. Основы теории принятия решений : учебное пособие. — М.: Ай Пи Ар Медиа, 2022. — 66 с.

69. Орлов А.И. Прикладной статистический анализ : учебник. — М.: Ай Пи Ар Медиа, 2022. — 812 с.

70. Орлов А.И. Проблемы управления экологической безопасностью : учебное пособие. — М.: Ай Пи Ар Медиа, 2022. — 224 с.

71. Орлов А.И. Теория принятия решений : учебник. — М.: Ай Пи Ар Медиа, 2022. — 826 с.

72. Орлов А.И. Устойчивые экономико-математические методы и модели : монография. — М.: Ай Пи Ар Медиа, 2022. — 337 с.

73. Орлов А.И. Экспертные оценки : учебное пособие. — М.: : Ай Пи Ар Медиа, 2022. — 57 с.

74. Агаларов З.С., Орлов А.И. Эконометрика : учебник. — М.: : Дашков и К, 2021. — 380 с.

75. Клейн Ф. Лекции о развитии математики в XIX столетии. Часть I. – М.-Л.: Объединенное научно-техническое издательство НКТП СССР. Главная редакция технико-теоретической литературы, 1937. – 432 с.

76. Орлов А.И. О новой парадигме математических методов исследования // Научный журнал КубГАУ. 2016. №122. С. 807–832.

77. Орлов А.И. О новой парадигме организационно-экономического моделирования, эконометрики и статистики / Стратегическое планирование и развитие предприятий. Материалы Четырнадцатого всероссийского симпозиума (Москва, 9-10 апреля 2013 г.). Под ред. чл.-корр. РАН Г.Б. Клейнера. Секция 2. - М.: ЦЭМИ РАН, 2013. - С. 140-142.

78. Орлов А.И. О новой парадигме математических методов и моделей социально-экономических процессов / Материалы республиканской научно-практической конференции «Новые теоремы молодых математиков – 2013». – Наманган: Наманганском Государственный Университет, 2013. - С. 49-52.

79. . Орлов А.И. О новой парадигме математических методов и моделей социально-экономических процессов / Актуальные вопросы экономики и финансов в условиях современных вызовов российского и мирового хозяйства: материалы международной научно-практической конференции НОУ ВПО «СИ ВШПП», 25 марта 2013 г. Ч. 2. – Самара: ООО «Издательство Ас Гард», 2013. – С. 400-404.

80. Орлов А.И. О новой парадигме прикладной математики / Философия математики: актуальные проблемы. Математика и реальность. Тезисы Третьей всероссийской научной конференции (27-28 сентября 2013 г.) Редкол.: Бажанов В.А. и др. – М.: Центр стратегической конъюнктуры, 2013. – С. 84–87.

81. Орлов А.И. О новой парадигме математического моделирования при управлении развитием крупномасштабных систем / Управление развитием крупномасштабных систем (MLSD'2013). Материалы Седьмой международной конференции (30 сентября – 2 окт. 2013 г.), в 2 т. Т.1. Пленарные доклады, секции 1 – 3. – М.: ИПУ РАН, 2013. – С. 297 – 299.

82. Орлов А.И. О новой парадигме прикладной математической статистики / Статистические методы оценивания и проверки гипотез: межвуз. сб. науч. тр. Вып. 25. – Пермь: Перм. гос. нац. иссл. ун-т, 2013. –С. 162-176.

83. Орлов А.И. Статистическое образование в соответствии с новой парадигмой прикладной статистики / Россия: тенденции и перспективы развития. Ежегодник. Вып. 13 Ч. 1. Отв. ред. В.И. Герасимов. – М.: ИНИОН РАН. Отд. науч. сотрудничества, 2018. - С. 868-874.

84. Орлов А. Статистическое образование в соответствии с новой парадигмой прикладной статистики / Экономист. 2018. №10.

85. Орлов А.И. Статистическое образование в соответствии с новой парадигмой прикладной статистики / Математические основы разработки и использования машинного интеллекта: Сборник научных статей, посвященный 70-летию со дня рождения доктора технических наук, профессора Лябаха Николая Николаевича. - Майкоп: Изд-во "ИП Кучеренко В.О.", 2018. - С. 93-108.

86. Орлов А.И. Смена парадигм в прикладной статистике // Заводская лаборатория. Диагностика материалов. 2021. Т.87. № 7. С. 6-7.

Литература к главе 2

1. Орлов А.И. Развитие математических методов исследования (2006 – 2015 гг.) / Заводская лаборатория. Диагностика материалов. 2017. Т.83. №1. Ч.1. С. 78-86.

2. Орлов А.И. Устойчивость в социально-экономических моделях. – М. : Наука, 1979. – 296 с.

3. Орлов А.И. Статистика объектов нечисловой природы и экспертные оценки / Экспертные оценки. Вопросы кибернетики. Вып.58. - М.: Научный Совет АН СССР по комплексной проблеме «Кибернетика», 1979. С. 17-33.

4. Тюрин Ю.Н., Литвак Б.Г., Орлов А.И., Сатаров Г.А., Шмерлинг Д.С. Анализ нечисловой информации / Заводская лаборатория. 1980. Т.46. №10. С. 931-935.

5. Орлов А.И. Статистика объектов нечисловой природы (Обзор) / Заводская лаборатория. 1990. Т.56. №3. С. 76-83.

6. Орлов А.И. Тридцать лет статистики объектов нечисловой природы (обзор) / Заводская лаборатория. Диагностика материалов. 2009. Т.75. №5. С. 55-64.

7. Орлов А.И. Организационно-экономическое моделирование. Часть 1. Нечисловая статистика. – М.: Изд-во МГТУ им. Н.Э. Баумана, 2009. – 544 с.

8. Кун Т. Структура научных революций. М.: АСТ, 2003. — 605 с.

9. Орлов А. И. Новая парадигма прикладной статистики / Заводская лаборатория. Диагностика материалов. 2012. Т.78. №1, часть I. С. 87-93.

10. Орлов А. И. О новой парадигме математических методов исследования / Научный журнал КубГАУ. 2016. №122. С. 807–832.

11. Орлов А.И. Объекты нечисловой природы / Заводская лаборатория. Диагностика материалов. 1995. Т.61. №3. С.43-52.

12. Дискуссия по анализу интервальных данных / Заводская лаборатория. 1990. Т.56. №7. С.75-95.

13. Скибицкий Н.В. Решение задачи аналитического описания статических характеристик в условиях интервальной неопределенности / Заводская лаборатория. Диагностика материалов. 2019. Т.85. № 3. С. 64-74.

14. Орлов А.И. Прикладная статистика. - М.: Экзамен, 2006. - 671 с.

15. Орлов А.И. Теория принятия решений.– М.: Экзамен, 2006. – 576 с.

16. Орлов А.И., Луценко Е.В. Системная нечеткая интервальная математика. – Краснодар, КубГАУ. 2014. – 600 с.

17. Орлов А.И. Статистика интервальных данных (обобщающая статья) / Заводская лаборатория. Диагностика материалов. 2015. Т. 81. № 3. С. 61 - 69.

18. Орлов А. И. Эконометрика. - М.: Экзамен, 2002. – 576 с.

19. Орлов А.И. Теория нечетких множеств – часть теории вероятностей / Научный журнал КубГАУ. 2013. № 92. С. 51-60.

20. Орлов А.И., Луценко Е.В., Лойко В.И. Организационно-экономическое, математическое и программное обеспечение контроллинга, инноваций и менеджмента: монография / под общ. ред. С. Г. Фалько. – Краснодар : КубГАУ, 2016. – 600 с.

21. Крамер Г. Математические методы статистики. - М.: Мир, 1975. - 648 с.

22. Смирнов Н.В., Дунин-Барковский И.В. Курс теории вероятностей и математической статистики для технических приложений. Изд. 3-е, стереотипное. – М.: Наука, 1969. – 512 с.

23. Большев Л.Н., Смирнов Н.В. Таблицы математической статистики / 3-е изд.- М.: Наука, 1983. - 416 с. (1-е изд. – 1965).

24. Каган А.М., Линник Ю.В., Рао С.Р. Характеризационные задачи математической статистики. - М.: Наука, 1972. - 656 с.

25. Современные проблемы кибернетики (прикладная статистика). - М.: Знание, 1981. – 64 с.

26. Орлов А.И. О перестройке статистической науки и её применений / Вестник статистики. 1990. № 1. С.65 – 71.

27. Орлов А.И. Вероятностные модели конкретных видов объектов нечисловой природы / Заводская лаборатория. Диагностика материалов. 1995. Т.61. №5. С.43-51.

28. Андреенков В.Г., Орлов А.И., Толстова Ю.Н. (отв. ред.). Анализ нечисловой информации в социологических исследованиях. - М.: Наука, 1985. - 220 с.

29. Лойко В.И., Луценко Е.В., Орлов А.И. Современные подходы в наукометрии. – Краснодар: КубГАУ, 2017. – 532 с.

30. Орлов А.И. Характеризация средних величин шкалами измерения / Научный журнал КубГАУ. 2017. №134. С. 877 – 907.

31. Психологические измерения. Сб. статей. - М.: Мир, 1967. - 196 с.

32. Пфанцагль И. Теория измерений. - М.: Мир, 1976. - 248 с.

33. Орлов А.И. Статистика нечисловых данных за сорок лет (обзор) / Заводская лаборатория. Диагностика материалов. 2019. Т.85. №11. С. 69-84.

34. Кемени Дж., Снелл Дж. Кибернетическое моделирование: Некоторые приложения. – М.: Советское радио, 1972. – 192 с.

35. Жуков М. С., Орлов А. И. Задача исследования итогового ранжирования мнений группы экспертов с помощью медианы Кемени / Научный журнал КубГАУ. 2016. № 122. С. 785 – 806.

36. Орлов А.И. Средние величины и законы больших чисел в пространствах произвольной природы / Научный журнал КубГАУ. 2013. № 89. С. 556 – 586.

37. Орлов А.И. Предельная теория решений экстремальных статистических задач / Научный журнал КубГАУ. 2017. №133. С. 579 – 600.

38. Орлов А.И. Асимптотика оценок плотности распределения вероятностей / Научный журнал КубГАУ. 2017. № 131. С. 845 – 873.

39. Ибрагимов И.А., Хасьминский Р.З. Асимптотическая теория оценивания. – М.: Наука, 1979. – 528 с.

40. Вероятность и математическая статистика: Энциклопедия / Гл. ред. Ю.В. Прохоров. – М.: Большая Российская Энциклопедия, 1999. – 910 с.

41. Орлов А.И. Скорость сходимости ядерных оценок плотности в пространствах произвольной природы / Статистические методы оценивания и проверки гипотез: межвуз. сб. науч. тр. / Перм. гос. нац. иссл. ун-т. - Пермь, 2018. - Вып.28. - С. 35-45.

42. Орлов А.И. Математические методы теории классификации / Научный журнал КубГАУ. 2014. № 95. С. 23 – 45.

43. Луценко Е.В., Орлов А.И. Методы снижения размерности пространства статистических данных / Научный журнал КубГАУ. 2016. № 119. С. 92–107.

44. Орлов А.И. Прогностическая сила – наилучший показатель качества алгоритма диагностики / Научный журнал КубГАУ. 2014. № 99. С. 33-–49.

45. Орлов А.И. Асимптотическое поведение статистик интегрального типа / Доклады АН СССР. 1974. Т.219. №4. С. 808-811.

46. Орлов А.И. Предельная теория непараметрических статистик / Научный журнал КубГАУ. 2014. № 100. С. 31-52.

47. Налимов В.В. Применение математической статистики при анализе вещества. – М.: Физматгиз, 1960. – 430 с.

48. Новицкий П.В., Зограф И.А. Оценка погрешностей результатов измерений. – Л.: Энергоатомиздат, 1985. – 248 с.

49. Орлов А.И. Распределения реальных статистических данных не являются нормальными / Научный журнал КубГАУ. 2016. № 117. С. 71–90.

50. Орлов А.И. Современное состояние непараметрической статистики / Научный журнал КубГАУ. 2015. № 106. С. 239 – 269.

51. Орлов А.И. Статистика нечетких данных / Научный журнал КубГАУ. 2016. №119. С. 75–91.

52. Орлов А.И. Теория люсианов / Научный журнал КубГАУ. 2014. № 101. С. 275 – 304.

53. Орлов А.И. Предельные теоремы в статистическом контроле / Научный журнал КубГАУ. 2016. № 116. С. 462 – 483.

54. Орлов А.И. Организационно-экономическое моделирование : учебник : в 3 ч. Ч.2. Экспертные оценки. — М.: Изд-во МГТУ им. Н. Э. Баумана, 2011. — 486 с.

55. Статистические методы анализа экспертных оценок / Ученые записки по статистике, т. 29. - М.: Наука, 1977. - 385 с.

56. Экспертные оценки / Вопросы кибернетики. - Вып.58. - М.: Научный Совет АН СССР по комплексной проблеме "Кибернетика". 1979. - 200 с.

57. Экспертные оценки в системных исследованиях / Сборник трудов. - Вып.4. - М.: ВНИИСИ, 1979. - 120 с.

58. Экспертные оценки в задачах управления / Сборник трудов. - М.: Институт проблем управления. 1982. - 106 с.

59. Орлов А.И. Теория экспертных оценок в нашей стране / Научный журнал КубГАУ. 2013. № 93. С. 1-11.

60. Гнеденко Б.В., Орлов А.И. Роль математических методов исследования в кардинальном ускорении научно-технического прогресса / Заводская лаборатория. 1988. Т.54. №1. С.1 - 4.

61. Орлов А.И. О высоких статистических технологиях / Научный журнал КубГАУ. 2015. № 105. С. 14 – 38.

62. Орлов А.И. Компьютерно-статистические методы: состояние и перспективы / Научный журнал КубГАУ. 2014. № 103. С. 163 – 195.

63. Загоруйко Н.Г., Орлов А.И. Некоторые нерешенные математические задачи прикладной статистики / Современные проблемы кибернетики (прикладная статистика). - М.: Знание, 1981. - С.53-63.

64. Орлов А.И. Некоторые нерешенные вопросы в области математических методов исследования / Заводская лаборатория. Диагностика материалов. 2002. Т.68. №3. С.52-56.

65. Никитин Я.Ю. Асимптотическая эффективность непараметрических критериев. - М.: Наука, 1995. - 240 с.

66. Орлов А.И. Проблема множественных проверок статистических гипотез / Заводская лаборатория. Диагностика материалов. 1996. Т.62. №5. С.51-54.

67. Орлов А.И. Статистика нечисловых данных за сорок лет (обзор) // Заводская лаборатория. Диагностика материалов. 2019. Т.85. №11. - С. 69-84.

68. Орлов А.И. Статистика нечисловых данных - центральная часть современной прикладной статистики // Научный журнал КубГАУ. 2020. №156. С. 111–142.

69. Орлов А.И. О методологии статистических методов / Научный журнал КубГАУ. 2014. № 104. С. 53–80.

Литература к главе 3

1. Орлов А.И. О новой парадигме прикладной математической статистики // Статистические методы оценивания и проверки гипотез: межвуз. сб. науч. тр. / Перм. гос. нац. иссл. ун-т. – Пермь, 2013. Вып. 25. С.162-176.

2. Орлов А.И. Устойчивость в социально-экономических моделях. - М.: Наука, 1979. - 296 с.

3. Орлов А.И. Статистика объектов нечисловой природы и экспертные оценки // Экспертные оценки / Вопросы кибернетики. Вып.58. - М.: Научный Совет АН СССР по комплексной проблеме "Кибернетика", 1979. С.17-33.

4. Орлов А.И. Организационно-экономическое моделирование: учебник : в 3 ч. Часть 1: Нечисловая статистика. – М.: Изд-во МГТУ им. Н.Э. Баумана. 2009. – 541 с.

5. Орлов А.И. О развитии статистики объектов нечисловой природы // Научный журнал КубГАУ. 2013. № 93. С. 41-50.

6. Орлов А.И. Оценки плотности в пространствах произвольной природы // Статистические методы оценивания и проверки гипотез: межвуз. сб. науч. тр. / Перм. гос. нац. иссл. ун-т. – Пермь, 2013. Вып. 25. С.21-33.

7. Орлов А.И. Оценки плотности распределения вероятностей в пространствах произвольной природы // Научный журнал КубГАУ. 2014. № 99. С. 15-32.

8. Орлов А.И. Ядерные оценки плотности в пространствах произвольной природы // Статистические методы оценивания и проверки гипотез: межвуз. сб. науч. тр. / Перм. гос. нац. иссл. ун-т. – Пермь, 2015. Вып. 26. С. 43-57.

9. Орлов А.И. Предельные теоремы для ядерных оценок плотности в пространствах произвольной природы // Научный журнал КубГАУ. 2015. № 108. С. 316 – 333.

10. Орлов А.И. Непараметрические ядерные оценки плотности вероятности в дискретных пространствах // Научный журнал КубГАУ. 2016. № 122. С. 833 –855.

11. Орлов А.И. Ядерные оценки плотности в конечных пространствах // Статистические методы оценивания и проверки гипотез: межвуз. сб. науч. тр. / Перм. гос. нац. иссл. ун-т. – Пермь, 2016. – Вып. 27. – С. 24-37.

12. Вероятность и математическая статистика: Энциклопедия / Гл. ред. Ю.В. Прохоров. – М.: Большая Российская Энциклопедия, 1999. – 910 с.

13. Орлов А.И. Статистика объектов нечисловой природы // Теория вероятностей и ее применения. 1980. Т.XXV. № 3. С. 655-656.

14. Орлов А.И. Непараметрические оценки плотности в топологических пространствах // Прикладная статистика. Ученые записки по статистике, т.45. – М.: Наука, 1983. – С. 12-40.

15. Rosenblatt M. Remarks on some nonparametric estimates of a density function // Ann. Math. Statist. 1956. V.27. N 5. P. 832 – 837.

16. Parzen E. On estimation of a probability density function and mode // Ann. Math. Statist. 1962. V.33. N 6. P. 1065-1076.

17. Ибрагимов И.А., Хасьминский Р.З. Асимптотическая теория оценивания. – М.: Наука, 1979. – 528 с.

18. Орлов А.И. О развитии методологии статистических методов // Статистические методы оценивания и проверки гипотез: межвуз. сб. науч. тр. / Перм. гос. нац. иссл. ун-т. – Пермь, 2001. – Вып. 15. – С.118-131.

19. Орлов А.И. О влиянии методологии на последствия принятия решений // Научный журнал КубГАУ. 2017. № 125. С. 319 – 345.

20. Смирнов Н.В. О приближении плотностей распределения случайных величин // Ученые записки МГПИ им. В.П. Потемкина. 1951. Т. XVI. Вып.3. С. 69-96.

21. Надарая Э.А. К построению доверительных областей для плотности вероятности // Сообщения АН ГрузССР. 1970. Т.59. № 1. С.33-36.

22. Надарая Э.А. О построению доверительных областей для плотности вероятности // Аннотации докладов семинара Института прикладной математики Тбилисского государственного университета. 1972. № 5. С. 27-32.

23. Мания Г.М. Статистическое оценивание распределения вероятностей. - Тбилиси: Издательство Тбилисского университета, 1974. - 240 с.

24. Конаков В.Д. Полные асимптотические разложения для максимального уклонения эмпирической функции плотности // Теория вероятностей и её применения. 1978. Т. XXIII. №3. С. 495-509.

25. Боровков А.А. Теория вероятностей. - М.: Наука, 1976. - 352 с.

26. Прохоров Ю,В., Розанов Ю.А. Теория вероятностей: Основные понятия. Предельные теоремы. Случайные процессы / Справочная математическая библиотека. - М.: Наука, 1973. - 496 с.

27. Орлов А.И. Прикладная статистика. Учебник для вузов. — М.: Экзамен, 2006. — 672 с.

28. Орлов А.И. Средние величины и законы больших чисел в пространствах произвольной природы // Научный журнал КубГАУ. 2013. № 89. С. 556 – 586.

29. Орлов А.И. Асимптотика решений экстремальных статистических задач // Анализ нечисловых данных в системных исследованиях. Сборник трудов. Вып.10. - М.: Всесоюзный научно-исследовательский институт системных исследований, 1982. С. 4-12.

30. Орлов А.И. Математические методы теории классификации // Научный журнал КубГАУ. 2014. № 95. С. 23 – 45.

31. Орлов А.И. Прогностическая сила – наилучший показатель качества алгоритма диагностики // Научный журнал КубГАУ. 2014. № 99. С. 33-–49.

32. Орлов А.И.Асимптотика оценок плотности распределения вероятностей // Научный журнал КубГАУ. 2017. №131. С. 845 – 873.

33. Орлов А.И. Скорость сходимости ядерных оценок плотности в пространствах произвольной природы // Статистические методы оценивания и проверки гипотез: межвуз. сб. науч. тр. / Перм. гос. нац. иссл. ун-т. - Пермь, 2018. - Вып.28. - С. 35-45.

Литература к главе 4

1. Дискуссия по анализу интервальных данных // Заводская лаборатория. 1990. Т. 56. №.7. С. 75–95.

2. Сборник трудов Международной конференции по интервальным и стохастическим методам в науке и технике (ИНТЕРВАЛ-92). Т. 1, 2. — М.: МЭИ, 1992. — 216 с., 152 с.

3. Орлов А.И. Устойчивость в социально-экономических моделях. — М.: Наука, 1979. — 296 с.

4. ГОСТ 11.011-83. Прикладная статистика. Правила определения оценок и доверительных границ для параметров гамма-распределения. — М.: Изд-во стандартов, 1984. — 53 с.

5. Orlov A.I. Interval statistics // Interval Computations, 1992, №.1(3). P. 44–52.

6. Орлов А.И. Основные идеи интервальной математической статистики // Наука и технология в России. — 1994. №.4(6). С. 8–9.

7. Шокин Ю.И. Интервальный анализ. — Новосибирск: Наука, 1981. — 112 с.

8. Орлов А.И. О развитии реалистической статистики. — В сб.: Статистические методы оценивания и проверки гипотез. Межвузовский сборник научных трудов. Пермь: Изд-во Пермского государственного университета, 1990. С.89–99.

9. Орлов А.И. Некоторые алгоритмы реалистической статистики. — В сб.: Статистические методы оценивания и проверки гипотез. Межвузовский сборник научных трудов. — Пермь: Изд-во Пермского государственного университета, 1991. С.77–86.

10. Орлов А.И. О влиянии погрешностей наблюдений на свойства статистических процедур (на примере гамма-распределения). — В сб.: Статистические методы оценивания и проверки гипотез. Межвузовский сборник научных трудов. — Пермь: Изд-во Пермского государственного университета, 1988. С. 45–55.

11. Орлов А.И. Интервальная статистика: метод максимального правдоподобия и метод моментов. — В сб.: Статистические методы оценивания и проверки гипотез. Межвузовский сборник научных трудов. — Пермь: Изд-во Пермского государственного университета, 1995. С.114–124.

12. Орлов А.И. Интервальный статистический анализ. — В сб.: Статистические методы оценивания и проверки гипотез. Межвузовский сборник научных трудов. — Пермь: Пермский государственный университет, 1993. С.149–158.

13. Биттар А.Б. Метод наименьших квадратов для интервальных данных. Дипломная работа. — М.: МЭИ, 1994. — 38 с.

14. Пузикова Д.А. Об интервальных методах статистической классификации // Наука и технология в России. 1995. № 2(8). С. 12–13.

15. Орлов А.И. Пути развития статистических методов: непараметрика, робастность, бутстреп и реалистическая статистика // Надежность и контроль качества, 1991. № 8. С. 3–8.

16. Орлов А.И. Современная прикладная статистика // Заводская лаборатория. Диагностика материалов. 1998. Т. 64. № 3. С. 52–60.

17. Вощинин А.П. Метод оптимизации объектов по интервальным моделям целевой функции. — М.: МЭИ, 1987. —109 с.

18. Вощинин А.П., Сотиров Г.Р. Оптимизация в условиях неопределенности. — М.: МЭИ; София: Техника, 1989. — 224 с.

19. Вощинин А.П., Акматбеков Р.А. Оптимизация по регрессионным моделям и планирование эксперимента. — Бишкек: Илим, 1991. — 164 с.

20. Вощинин А.П. Метод анализа данных с интервальными ошибками в задачах проверки гипотез и оценивания параметров неявных и линейно параметризованных функций // Заводская лаборатория. Диагностика материалов. 2000. Т. 66, № 3. С. 51–65.

21. Вощинин А.П. Интервальный анализ данных: развитие и перспективы // Заводская лаборатория, 2002. Т. 68, № 1. С. 118–126.

22. Дывак Н.П. Разработка методов оптимального планирования эксперимента и анализа интервальных данных. Автореф. дисс. канд.. технич. наук. — М.: МЭИ, 1992. — 20 с.

23. Симов С.Ж. Разработка и исследование интервальных моделей при анализе данных и проектировании экспертных систем. Автореф. дисс. канд. технич. наук. — М.: МЭИ, 1992. — 20 с.

24. Орлов А.И. Вероятность и прикладная статистика: основные факты: справочник. – М.: КНОРУС, 2010. – 192 с.

25. Орлов А.И. Часто ли распределение результатов наблюдений является нормальным? // Заводская лаборатория, 1991. Т. 57. № 7. С. 64–66.

26. Новицкий П.В., Зограф И.А. Оценка погрешностей результатов измерений. — Л.: Энергоатомиздат, 1985. — 248 с.

27. Гнеденко Б.В., Хинчин А.Я. Элементарное введение в теорию вероятностей. — М.: Наука, 1970.

28. Боровков А.А. Математическая статистика. — М.: Наука, 1984. — 472 с.

29. Орлов А.И. Эконометрика. Изд. 3-е, испр. и доп. — М.: Экзамен, 2004. — 576 с.

30. Орлов А.И. О развитии статистики объектов нечисловой природы / А.И. Орлов // Научный журнал КубГАУ. 2013. №93. С. 273–309.

31. Орлов А.И. Прикладная статистика. — М.: Экзамен, 2006.— 671 с.

32. Орлов А.И. Устойчивые экономико-математические методы и модели. Saarbrücken (Germany), Lambert Academic Publishing, 2011. 436 с.

33. Орлов А.И. Устойчивые математические методы и модели // Заводская лаборатория. Диагностика материалов. 2010. Т.76. №3. С.59-67.

34. Орлов А.И. Теория принятия решений. — М.: Экзамен, 2006. — 574 с.

35. Орлов А.И. Организационно-экономическое моделирование : учебник : в 3 ч. Ч. 1. Нечисловая статистика. – М.: Изд-во МГТУ им. Н.Э. Баумана, 2009. — 541 с.

36. Вощинин А.П., Бронз П.В. Построение аналитических моделей по данным вычислительного эксперимента в задачах анализа чувствительности и оценки экономических рисков // Заводская лаборатория. Диагностика материалов. 2007. Т.73. №1. С.101-109.

37. Вощинин А.П., Скибицкий Н.В. Интервальный подход к выражению неопределенности измерений и калибровке цифровых измерительных систем // Заводская лаборатория. Диагностика материалов. 2007. Т.73. №11. С.66-71.

38. Орлов А.И. Об оценивании параметров гамма-распределения // Обозрение прикладной и промышленной математики. 1997. Т. 4. Вып. 3. С. 471–482.

39. Гуськова Е.А., Орлов А.И. Интервальная линейная парная регрессия // Заводская лаборатория. Диагностика материалов. 2005. Т. 71. №3. С. 57–63.

40. Орлов А.И., Луценко Е.В. О развитии системной нечеткой интервальной математики // Философия математики: актуальные проблемы. Математика и реальность. Тезисы Третьей всероссийской научной конференции; 27-28 сентября 2013 г.– М.: Центр стратегической конъюнктуры, 2013. – С.190–193.

41. Луценко Е.В., Орлов А.И. Системная нечеткая интервальная математика (СНИМ) – перспективное направление теоретической и вычислительной математики // Научный журнал КубГАУ. №91. С. 255–308.

42. Орлов А.И. Основные идеи статистики интервальных данных // Научный журнал КубГАУ. 2013. №94. С. 867 – 892.

43. Новиков Д.А., Орлов А.И. Математические методы анализа интервальных данных // Заводская лаборатория. Диагностика материалов. 2014. Т.80. №7. С. 5 – 6.

44. Орлов А.И. Оценка погрешностей характеристик финансовых потоков инвестиционных проектов в ракетно-космической промышленности // Научный журнал КубГАУ. 2015. №109. С. 238 – 264.

45. Орлов А.И. Статистика интервальных данных (обобщающая статья) // Заводская лаборатория. Диагностика материалов. 2015. Т. 81. № 3. С. 61 - 69.

Литература к главе 5

1. Орлов А.И. Прикладная статистика. — М.: Экзамен, 2006. — 671 с.

2. Орлов А.И. Устойчивость в социально-экономических моделях. — М.: Наука, 1979. — 296 с.

3. Налимов В.В. Теория эксперимента. — М.: Наука, 1971. — 208 с.

4. Ермаков С.М., Бродский В.З., Жиглявский А.А. и др. Математическая теория планирования эксперимента. — М.: Физматлит, 1983. — 392 с.

5. Бернштейн С.Н. Об одном элементарном свойстве коэффициента корреляции / Зап. Харьк. матем. тов. 1932. Т. 5. С. 65-66.

6. Колмогоров А.Н. К вопросу о пригодности найденных статистическим путем формул прогноза / Журн. геофиз. 1933. Т.3. С. 78-82.

7. Орлов А.И. Методы поиска наиболее информативных множеств признаков в регрессионном анализе / Заводская лаборатория. Диагностика материалов. 1995. Т.61. № 1. С. 56-58.

8. Орлов А.И. Проблема множественных проверок статистических гипотез / Заводская лаборатория. Диагностика материалов. 1996. Т.62. № 5. С. 51-54.

9. Сердобольский В.И., Орлов А.И. Статистический анализ при большом числе параметров / Программно-алгоритмическое обеспечение прикладного многомерного статистического анализа. Тезисы докладов III Всесоюзной школы-семинара. — М.: ЦЭМИ АН СССР, 1987. — С. 151-160.

10. Орлов А.И. Организационно-экономическое моделирование: : учебник : в 3 ч. Ч.1: Нечисловая статистика. — М.: Изд-во МГТУ им. Н. Э. Баумана, 2009. — 542 с.

11. Орлов А.И. Статистический контроль по двум альтернативным признакам и метод проверки их независимости по совокупности малых выборок / Заводская лаборатория. Диагностика материалов. 2000. Т.66. № 1. С. 58-62.

12. Лойко В. И., Луценко Е. В., Орлов А. И. Современные подходы в наукометрии: монография / Под науч. ред. проф. С. Г. Фалько. – Краснодар: КубГАУ, 2017. – 532 с.

13. Орлов А.И. Статистические пакеты – инструменты исследователя / Заводская лаборатория. Диагностика материалов. 2008. Т.74. № 5. С. 76-78.

14. Орлов А.И. Первый Всемирный конгресс Общества математической статистики и теории вероятностей им. Бернулли / Заводская лаборатория. Диагностика материалов. 1987. Т.53. №3. С.90-91.

15. Тырсин А.Н., Максимов К.Е. Оценивание линейных регрессионных уравнений с помощью метода наименьших модулей // Заводская лаборатория. Диагностика материалов. 2012. Том 78. № 7. С. 65-71.

16. Орлов А.И. Распределения реальных статистических данных не являются нормальными // Научный журнал КубГАУ. 2016. № 117. С. 71–90.
17. Орлов А.И. Основные черты новой парадигмы математической статистики / Научный журнал КубГАУ. 2013. № 90. С. 45-71.
18. Орлов А.И. Современное состояние непараметрической статистики / Научный журнал КубГАУ. 2015. № 106. С. 239 – 269.
19. Королюк В.С., Портенко Н.И., Скороход А.В., Турбин А.Ф. Справочник по теории вероятностей и математической статистике. — М.: Наука, 1985. — 640 с.
20. Орлов А.И. Асимптотика оценок плотности распределения вероятностей / Научный журнал КубГАУ. 2017. № 131. С. 845 – 873.
21. Орлов А.И. Восстановление зависимости методом наименьших квадратов на основе непараметрической модели с периодической составляющей / Научный журнал КубГАУ. 2013. № 91. С. 133-162.
22. Себер Дж. Линейный регрессионный анализ. — М.: Мир, 1980. — 456 с.
23. Орлов А.И. Асимптотика некоторых оценок размерности модели в регрессии / Прикладная статистика. Ученые записки по статистике. Т. 45. — М.: Наука, 1983. — С. 260–265.
24. Орлов А.И. Об оценивании регрессионного полинома / Заводская лаборатория. Диагностика материалов. 1994. Т.60. №5. С. 43-47.
25. Орлов А.И. Статистика интервальных данных (обобщающая статья) / Заводская лаборатория. Диагностика материалов. 2015. Т.81. №3. С. 61 - 69.
26. Гуськова Е.А., Орлов А.И. Интервальная линейная парная регрессия (обобщающая статья) / Заводская лаборатория. Диагностика материалов. 2005. Т.71. №3. С.57-63.
27. Орлов А.И. Теория принятия решений. — М.: Экзамен, 2006. — 576 с.
28. Орлов А.И., Луценко Е.В. Системная нечеткая интервальная математика. – Краснодар, КубГАУ. 2014. – 600 с.
29. Орлов А.И. Ошибки при использовании коэффициентов корреляции и детерминации / Заводская лаборатория. Диагностика материалов. 2018. Т.84. № 3. С. 68-72.
30. Орлов А.И. Многообразие моделей регрессионного анализа (обобщающая статья) / Заводская лаборатория. Диагностика материалов. 2018. Т.84. №5. С. 63-73.
31. Орлов А.И. Вероятностно-статистические модели корреляции и регрессии / Научный журнал КубГАУ. 2020. №160. С. 130–162.

Литература к главе 6

1. Орлов А.И. Вероятностно-статистические модели корреляции и регрессии / Научный журнал КубГАУ. 2020. №160. С. 130–162.
2. Орлов А.И. Многообразие моделей регрессионного анализа (обобщающая статья) / Заводская лаборатория. Диагностика материалов. 2018. Т.84. №5. С. 63-73.
3. Кендалл М.Дж., Стьюарт А. Статистические выводы и связи. - М: Наука, 1973. - 900 с.
4. Демиденко Е.З. Линейная и нелинейная регрессия. - М.: Финансы и статистика, 1982. - 126 с.
5. Алгоритмы и программы восстановления зависимостей / Под ред. В.Н. Вапника. - М.: Наука, 1984. - 816 с.
6. Петрович М.Л. Регрессионный анализ и его математическое обеспечение на ЕС ЭВМ: Практическое руководство. - М.: Финансы и статистика, 1982. - 193 с.
7. Математическая теория планирования эксперимента / Справочная математическая библиотека. - М.: Наука, 1983. - 392 с.
8. Себер Дж. Линейный регрессионный анализ. - М.: Мир, 1980. - 456 с.
9. Дрейпер Н., Смит Г. Прикладной регрессионный анализ: Книга 2. - М.: Финансы и статистика, 1987. - 351 с.
10. Орлов А.И. Предельная теория решений экстремальных статистических задач / Научный журнал КубГАУ. 2017. №133. С. 579–600.
11. Орлов А.И. Организационно-экономическое моделирование: : учебник : в 3 ч. Ч.1: Нечисловая статистика. — М.: Изд-во МГТУ им. Н. Э. Баумана, 2009. — 542 с.
12. Орлов А.И. Прикладная статистика. - М.: Экзамен, 2006 - 671 с.
13. Орлов А.И. Распределения реальных статистических данных не являются нормальными / Научный журнал КубГАУ. 2016. №117. С. 71–90.
14. Налимов В.В. Теория эксперимента. - М.: Наука, 1971. - 208 с.
15. Орлов А.И. Оценка размерности модели в регрессии / Алгоритмическое и программное обеспечение прикладного статистического анализа. - М.: Наука, 1980. - С. 92-99.
16. Митропольский А.К. Техника статистических вычислений. - М.: Наука, 1971. - 570 с.

17. Пустыльник Е.И. Статистические методы анализа и обработки наблюдений. - М.: Наука, 1968. - 288 с.

18. Колмогоров А.Н. К обоснованию метода наименьших квадратов / Успехи математических наук. 1946. Т.1. Вып. 1. С.57-70.

19. Тутубалин В.Н. Теория вероятностей. - М.: МГУ, 1972. - 232 с.

20. Гальченко М.В., Гуревич А.В. Почти параметрическая оценка регрессии / Статистические методы оценивания и проверки гипотез: Межвузовский сборник научных трудов. - Пермь: Пермский ун-т, 1984. - С. 52-59.

21. Орлов А.И. Предельное распределение одной оценки числа базисных функций в регрессии / Прикладной многомерный статистический анализ. - М.: Наука, 1978. -= С. 380-381.

22. Уилкс С. Математическая статистика. - М.: Наука, 1967. - 632 с.

23. Ширяев А.Н. Статистический последовательный анализ: Оптимальные правила остановки. 2-е изд., перераб. - М.: Физматлит, 1976. - 272 с.

24. Арнольд В.И. О локальных задачах анализа / Вестник МГУ. Сер. матем. и мех. 1970. №2. С. 52-56.

25. Орлов А.И. Асимптотика некоторых оценок размерности модели в регрессии / Прикладная статистика. - М.: Наука, 1983. - С. 260-265.

26. Каган А.М., Линник Ю.В., Рао С.Р. Характеризационные задачи математической статистики. - М.: Наука, 1972. - 656 с.

27. Бернштейн С.Н. Об одном свойстве, характеризующем закон Гаусса / Труды Ленинградского политехн. ин-та. 1941. №3. С. 21-22. - Перепеч. в кн.: Бернштейн С.Н. Собрание сочинений: Т.IV: Теория вероятностей и математическая статистика. - М.: Наука, 1964. - С. 394-395, 569.

28. Гнеденко Б.В. Об одной теореме С.Н. Бернштейна / Известия АН СССР, Сер. матем. 1948. Т.12. №1. С. 97-100.

29. Боганик Г.Н. Об установлении порядка уравнения параболической регрессии / Теория вероятностей и её применения. 1967. Т.XII. №4. С. 750-763.

30. Киричук В.С. Выбор степени полинома, сглаживающего результаты измерений / Автометрия. 1970. №3. С. 26-71. 31.

31. Ковалерчук Б.Я., Лавков В.В. Поиск максимального верхнего нуля для минимизации числа признаков в регрессионном анализе / Журнал вычислительной математики и математической физики. 1984. Т.24. №3. С. 1241-1249.

32. Крамер Г. Математические методы статистики. - М.: Мир, 1975. - 648 с.

33. Орлов А.И. Математические методы теории классификации / Научный журнал КубГАУ. 2014. №95. С. 423 – 459.

34. Орлов А.И. Базовые результаты математической теории классификации / Научный журнал КубГАУ. 2015. №110. С. 219 – 239.

35. Эренбург Э.С. Смеси распределений в надежности. - М.: Знание, 1983. - 48 с.

36. Орлов А.И. Оценивание параметров: одношаговые оценки предпочтительнее оценок максимального правдоподобия / Научный журнал КубГАУ. 2015. №109. С. 208–237.

37. White H. Maximum likelihood estimation of misspecified models / Econometrics. 1982. V.50. N 1. P.1-25.

38. Орлов А.И. Некоторые вероятностные вопросы теории классификации / Прикладная статистика. - М.: Наука, 1983. - С.166-179.

39. Никифоров А.М. Исследование некоторых вопросов статистической теории распознавания образов с самообучением и анализа данных с пропусками применительно к задаче обработки клинических данных / Дисс. ... канд. физ.-мат. наук. - М.: МФТИ, 1987. - 144 с.

40. Волынский Ю.Д., Курочкина А.И. Многомерный анализ клинических данных / Вестник АМН СССР. 1987. № 1. С. 84-93.

41. Перекрест В.Т. Нелинейный типологический анализ социально-экономической информации: Математические и вычислительные методы. - Л.: Наука, 1983. - 176 с.

42. Терехина А.Г. Анализ данных методами многомерного шкалирования. - М.: Наука, 1986. - 168 с.

43. Енюков И.С. Методы, алгоритмы, программы многомерного статистического анализа: Пакет ППСА. - М.: Финансы и статистика, 1986. - 232 с.

44. Huber P.J. Projection Pursuit / Ann. Statist. 1985. V/13. N 3. P. 435-476.

45. Орлов А.И. Первый Всемирный конгресс Общества математической статистики и теории вероятностей им. Бернулли // Надежность и контроль качества. 1987. №6. С. 54-59.

46. Орлов А.И. Общий взгляд на статистику объектов нечисловой природы / Анализ нечисловой информации в социологических исследованиях. - М.: Наука, 1985. - С. 58-92.

47. Тюрин Ю.Н., Литвак Б.Г., Орлов А.И., Сатаров Г.А., Шмерлинг Д.С. Анализ нечисловой информации / Препринт. - М.: Научный Совет АН СССР по комплексной проблеме "Кибернетика", 1981. - 80 с.

48. Орлов А.И. Методы снижения размерности / Приложение 2 к книге: Толстова Ю.Н. Основы многомерного шкалирования. - М.: Издательство КДУ, 2006. С. 113-120.

49. Орлов А.И., Луценко Е.В. Методы снижения размерности пространства статистических данных / Научный журнал КубГАУ. 2016. №119. С. 92–107.

50. Смоляк С.А., Титаренко Б.П. Устойчивые методы оценивания: Статистическая обработка неоднородных совокупностей. - М.: Статистика, 1980. - 208 с.

51. Рекомендации: Прикладная статистика. Методы обработки данных. Основные требования и характеристики / Орлов А.И., Миронова Н.Г., Фомин В.Н., Черчинцев А.Н. - М.: ВАНИИСБ 1987. - 64 с.

52. Орлов А.И. Основные требования к методам анализа данных (на примере задач классификации) / Научный журнал КубГАУ. 2020. №159. С. 239–267.

53. Орлов А.И. Новый подход к изучению устойчивости выводов в математических моделях / Научный журнал КубГАУ. 2014. № 100. С. 146-176.

54. Reise R.D. Consistency of minimum contrast estimators in nonstandart case / Metriks. 1978. V.25. N 3. P. 129-142.

55. Миркин Б.Г. Анализ качественных признаков и структур. - М.: Статистика, 1980. - 319 с.

56. Андрукович П.Ф. Некоторые свойства метода главных компонент / Многомерный статистический анализ в социально-экономических исследованиях. - М.: Наука, 1974. - С. 189-228.

57. Орлов А.И. Организационно-экономическое моделирование : учебник : в 3 ч. Ч.2. Экспертные оценки. — М.: Изд-во МГТУ им. Н. Э. Баумана, 2011. — 486 с.

58. Орлов А.И., Луценко Е.В. Системная нечеткая интервальная математика. Монография (научное издание). – Краснодар, КубГАУ. 2014. – 600 с.

59. Лойко В.И., Луценко Е.В., Орлов А.И. Высокие статистические технологии и системно-когнитивное моделирование в экологии : монография. – Краснодар : КубГАУ, 2019. – 258 с.

60. Орлов А.И. Оценивание размерности вероятностно-статистической модели // Научный журнал КубГАУ. 2020. №162. С. 1–36.

Литература к главе 7

1. Орлов А. И. Прикладная статистика. - М.: Экзамен, 2006. - 671 с.

2. Орлов А. И. Характеризация средних величин шкалами измерения // Научный журнал КубГАУ. 2017. №134. С. 877 – 907.

3. Енюков И. С. Методы оцифровки неколичественных признаков // Алгоритмическое и программное обеспечение прикладного статистического анализа. - М.: Наука, 1980. - С. 309-316.

4. Александров В. В., Горский Н. Д. Алгоритмы и программы структурного метода обработки данных. - Л.: Наука, 1983. - 208 с.

5. Саати Т. Л. Принятие решений. Метод анализа иерархий. — М.: Радио и связь, 1989. — 316 с.

6. Гаек Я., Шидак З. Теория ранговых критериев / Пер. с англ. - М.: Наука, 1971. – 376 с.

7. Холлендер М., Вульф Д. Непараметрические методы статистики. – М.: Финансы и статистика, 1983. - 518 с.

8. Алимов Ю.И. Альтернатива методу математической статистики. - М.: знание, 1980. - 64 с.

9. Малиновский Л. Г. Анализ статистических связей: модельно-конструктивный подход / Отв. ред. Н. А. Кузнецов, Л. И. Титомир ; Рос. акад. наук, Ин-т проблем передачи информации. - Москва : Наука, 2002. - 687 с.

10. Орлов А. И. О методах проверки однородности двух независимых выборок // Заводская лаборатория. Диагностика материалов. 2020. Т.86. №3. С. XX-XX.

11. Большев Л. Н., Смирнов Н. В. Таблицы математической статистики. – М.: Наука, 1983. - 416 с.

12. Орлов А. И. Структура непараметрической статистики (обобщающая статья) // Заводская лаборатория. Диагностика материалов. 2015. Т.81. №7. С. 62-72.

13. Загоруйко Н. Г., Орлов А. И. Некоторые нерешенные математические задачи прикладной статистики // Современные проблемы кибернетики (прикладная статистика). - М.: Знание, 1981. - С. 53-63.

14. Орлов А. И., Миронова Н. Г., Фомин В. Н., Черномордик О .М. Методика. Проверка однородности двух выборок параметров продукции при оценке ее технического уровня и качества. - М.: ВНИИСтандартизации, 1987. - 116 с.

15. Орлов А. И. Реальные и номинальные уровни значимости при проверке статистических гипотез // Научный журнал КубГАУ. 2015. № 114. С. 42–54.

16. Орлов А. И. Состоятельные критерии проверки абсолютной однородности независимых выборок // Заводская лаборатория. Диагностика материалов. 2012. Т.78. №11. С.66-70.

17. Орлов А. И. Устойчивость в социально-экономических моделях. — М.: Наука, 1979. — 296 с.
18. Орлов А. И. Устойчивые экономико-математические методы и модели. Разработка и развитие устойчивых экономико-математических методов и моделей для модернизации управления предприятиями. — Saarbrücken (Germany), LAP (Lambert Academic Publishing), 2011. — 436 с.
19. Орлов А. И. Некоторые вероятностные вопросы теории классификации // Прикладная статистика. - М.: Наука, 1983. - С. 166-179.
20. Тихонов А. Н., Арсенин В. Я. Методы решения некорректных задач. - М.: Наука,1986. – 288 с.
21. Куперштох В. Л., Миркин Б. Г., Трофимов В. А. Сумма внутренних связей как показатель качества классификации // Автоматика и телемеханика. 1976. №3. С. 91-98.
22. Орлов А. И. Метод статистических испытаний в прикладной статистике // Заводская лаборатория. Диагностика материалов. 2019. Т.85. №5. С. 67-79.
23. Блехман И. И., Мышкис А. Д., Пановко Я. Г. Механика и прикладная математика: логика и особенности приложений математики / 2-ое изд., испр. и доп. - М: Наука, 1990. - 360 с.
24. Орлов А. И. Статистические пакеты – инструменты исследователя // Заводская лаборатория. Диагностика материалов. 2008. Т.74. № 5. С. 76–78.
25. Орлов А. И. Сертификация и статистические методы (обобщающая статья) // Заводская лаборатория. Диагностика материалов. 1997. Т.63. № 3. С. 55-62.
26. Орлов А. И. Прогностическая сила – наилучший показатель качества алгоритма диагностики // Научный журнал КубГАУ. 2014. № 99. С. 33-49.
27. Тутубалин В. Н. Теория вероятностей в естествознании. - М.: Знание, 1972. - 64 с.
28. Орлов А. И. Эконометрика. Изд. 4-е, доп. и перераб. — Ростов-на-Дону: Феникс, 2009. — 572 с.
29. Чесноков С. В. Детерминационный анализ социально-экономических данных. Изд. 2, испр. и доп. - М.: URSS. 2009. - 168 с.
30. Лбов Г. С. Методы обработки разнотипных экспериментальных данных. - Новосибирск: Наука, 1981. - 160 с.
31. Хайтун С. Д. Наукометрия: Состояние и перспективы. - М.: Наука, 1983. - 344 с.
32. Орлов А. И. Непараметрические критерии согласия Колмогорова, Смирнова, омега-квадрат и ошибки при их применении // Научный журнал КубГАУ. 2014. №97. С. 32-45.
33. Джини К. Логика в статистике. - М.: Статистика, 1973. — 128 с.
34. Вентцель Е. С. Методологические особенности прикладной математики на современном этапе // Математики о математике. - М.: Знание, 1982. - С.37-55.
35. Миркин Б. Г. Анализ качественных признаков и структур. - М.: Статистика, 1980. — 319 с.
36. Орлов А. И. Предельная теория решений экстремальных статистических задач // Научный журнал КубГАУ. 2017. № 133. С. 579–600.
37. Орлов А. И. Оценка размерности модели в регрессии // Алгоритмическое и программное обеспечение прикладного статистического анализа. - М.: Наука, 1980. - С. 92-99.
38. Рабухин А. Е., Сильвестров В. П., Орлов А. И. и др. Результаты лечения больных острой пневмонией // Актуальные вопросы клинической и экспериментальной медицины. - М.: 4 ГУ МЗ СССР, 1978. - С. 132-138.
39. Орлов А. И., Миронова Н. Г., Фомин В. Н., Черчинцев А. Н. Рекомендации. Прикладная статистика. Методы обработки данных. Основные требования и характеристики. - М.: ВНИИСтандартизации, 1987. - 62 с.
40. Купцов В. И. Детерминизм и вероятность. - М.: Политиздат, 1976. - 256 с.
41. Сачков Ю. В. Вероятностная революция в науке (Вероятность, случайность, независимость, иерархия). - М.: Научный мир, 1999. - 144 с.
42. Сачков Ю. В. Введение в вероятностный мир. - М.: Наука, 1971. — 208 с.
43. Тутубалин В. Н. Границы применимости (вероятностно-статистические методы и их возможности). - М.: Знание, 1977. - 64 с.
44. Орлов А. И. О развитии прикладной статистики // Современные проблемы кибернетики (прикладная статистика). - М.: Знание, 1981. - С. 3-14.
45. Орлов А. И. Математика нечеткости // Наука и жизнь. 1982. № 7. С. 60-67.
46. Колмогоров А. Н. Основные понятия теории вероятностей. Изд. 2-е. - М.: Наука, 1974. - 120 с.
47. Моргенштерн О. О точности экономико-статистических наблюдений. - М.: Статистика, 1968. - 293 с.
48. Кендалл М. Дж., Стьюарт А. Многомерный статистический анализ и временные ряды. - М.: Наука, 1976. — 736 с.
49. Адлер Ю. П. Управление качеством: статистический подход. - М.: Знание, 1979. - 51 с.

50. Орлов А. И., Гусейнов Г. А. Математические методы в изучении способных к математике школьников // Исследования по вероятностно-статистическому моделированию реальных систем. - М.: ЦЭМИ АН СССР, 1977. - С. 80-93.

51. Тюрин Ю. Н. О математических задачах в экспертных оценках // Экспертные оценки. Вопросы кибернетики, вып.58. - М.: Научный совет АН СССР по комплексной проблеме "Кибернетика", 1979. - С. 7-16.

52. Орлов А. И. Теория люсианов // Научный журнал КубГАУ. 2014. № 101. С. 275–304.

53. Бурбаки Н. Очерки по истории математики. - М.: ИЛ, 1963. - 292 с.

54. Орлов А. И. Роль методологии в математических методах исследования // Заводская лаборатория. Диагностика материалов. 2019. Т.85. №7. С. 5-6.

55. Орлов А.И. Основные требования к математическим методам классификации // Заводская лаборатория. Диагностика материалов. 2020. Т.86. №11. С. 67-78.

56. Орлов А.И. Основные требования к методам анализа данных (на примере задач классификации) / Научный журнал КубГАУ. 2020. №159. С. 239–267.

Литература к главе 8

1. Горский В.Г., Орлов А.И. Математические методы исследования: итоги и перспективы // Заводская лаборатория. Диагностика материалов. 2002. Т.68. №1. С.108-112.

2. Орлов А.И. Новая парадигма прикладной статистики // Заводская лаборатория. Диагностика материалов. 2012. Т.78. №1. С. 87-93.

3. Орлов А.И. О новой парадигме математических методов исследования // Научный журнал КубГАУ. 2016. №122. С. 807–832.

4. Орлов А.И., Луценко Е.В. Системная нечеткая интервальная математика.– Краснодар, КубГАУ. 2014. – 600 с.

5. Орлов А.И. Развитие математических методов исследования (2006 – 2015 гг.) // Заводская лаборатория. Диагностика материалов. 2017. Т.83. №1. Ч.1. С. 78-86.

6. Колмогоров А.Н. Теория информации и теория алгоритмов. - М. Наука, 1987. - 304 с.

7. Григорьев Ю.Д. Метод Монте-Карло: вопросы точности асимптотических решений и качества генераторов псевдослучайных чисел // Заводская лаборатория. Диагностика материалов. 2016. Т.82. №7. С. 72-84.

8. Орлов А.И. Предельные теоремы и метод Монте-Карло // Заводская лаборатория. Диагностика материалов. 2016. Т.82. №7. С. 67-72.

9. Кутузов О.И., Татарникова Т.М. Из практики применения метода Монте-Карло // Заводская лаборатория. Диагностика материалов. 2017. Т.83. №3. С. 65-70.

10. Аронов И.З., Максимова О.В. Анализ времени достижения консенсуса в работе технических документов по стандартизации по результатам статистического моделирования // Заводская лаборатория. Диагностика материалов. 2017. Т.83. №3. С. 71-77.

11. Орлов А.И. Консенсус и истина (комментарий к опубликованной выше статье И.З. Аронова и О.В. Максимовой) // Заводская лаборатория. Диагностика материалов. 2017. Т.83. №3. С. 78-79.

12. Орлов А.И. Значение информационно-коммуникационных технологий для математических методов исследования // Заводская лаборатория. Диагностика материалов. 2017. Т.83. №7. С. 5-6.

13. Жуков М.С. Об алгоритмах расчета медианы Кемени // Заводская лаборатория. Диагностика материалов. 2017. Т.83. №7. С. 72-78.

14. Гадолина И.В., Лисаченко Н.Г. Разработка метода построения доверительных интервалов для процентилей случайной выборки прочности композитов с применением бутстреп-моделирования // Заводская лаборатория. Диагностика материалов. 2017. Т.83. №11. С. 73-77.

15. Орлов А.И. О проверке однородности двух независимых выборок // Заводская лаборатория. Диагностика материалов. 2003. Т.69. №1. С.55-60.

16. Большев Л.Н., Смирнов Н.В. Таблицы математической статистики / 3 изд. - М.: Наука, 1983. - 416 с.

17. Орлов А.И. Какие гипотезы можно проверять с помощью двухвыборочного критерия Вилкоксона? // Заводская лаборатория. Диагностика материалов. 1999. Т.65. № 1. С.51-55.

18. Орлов А.И. Состоятельные критерии проверки абсолютной однородности независимых выборок // Заводская лаборатория. Диагностика материалов. 2012. Т.78. №11. С.66-70.

19. Lehmann E.L. Consistency and unbiasedness of certain nonparametric tests // Ann. Math. Statist. 1951. V.22. N 2. P.165-179.

20. Rosenblatt M. Limit theorems associated with variants of the von Mises statistic // Ann. Math. Statist. 1952. V.23. N 4. P.617-623.

21. Орлов А.И. Организационно-экономическое моделирование : учебник : в 3 ч. Ч.3. Статистические методы анализа данных. - М.: Изд-во МГТУ им. Н.Э. Баумана, 2012. - 624 с.

22. Гаек Я., Шидак З. Теория ранговых критериев. - М.: Наука, 1971. - 374 с.

23. Холлендер М., Вулф Д. Непараметрические методы статистики. - М. : Финансы и статистика, 1983. - 520 с.

24. Парджанадзе А.М., Хмаладзе Э.В. Об асимптотической теории статистик от последовательных рангов // Теория вероятностей и её применения. 1986. Т.XXXI. Вып.4. С. 758-772.

25. Орлов А.И. Реальные и номинальные уровни значимости при проверке статистических гипотез // Научный журнал КубГАУ. 2015. № 114. С. 42–54.

26. Форсайт Дж., Малькольм М., Моулер К. Машинные методы математических вычислений. - М.: Мир, 1980. - 144 с.

27. Шеннон Р. Имитационное моделирование систем: Искусство и наука. - М.: Мир, 1978. - 418 с.

28. Орлов А.И. Первый Всемирный конгресс Общества математической статистики и теории вероятностей им. Бернулли // Заводская лаборатория. Диагностика материалов. 1987. Т.53. №3. С.90-91.

29. Ермаков С.М. Метод Монте-Карло и смежные вопросы. - М.: Наука, 1971. - 328 с.

30. Ермаков С.М., Михайлов Г.А. Статистическое моделирование. - М.: Наука, 1982. - 296 с.

31. Журбенко И.Г., Кожевникова И.А., Клиндухова О.В. Определение критической длины последовательности псевдослучайных чисел // Вероятностно-статистические методы исследования. - М.: МГУ им. М.В. Ломоносова, 1983. - С.18-39.

32. Журбенко И.Г., Кожевникова И.А., Смирнова О.С. О построении и исследовании псевдослучайных последовательностей различными методами // Заводская лаборатория. Диагностика материалов. 1985. Т.51. № 5. С. 47-51.

33. Журбенко И.Г. Анализ стационарных и однородных случайных систем. - М.: МГУ им. М.В. Ломоносова, 1987. - 240 с.

34. Рыданова Г.В. Методика изучения временных зависимостей в последовательностях псевдослучайных чисел // Заводская лаборатория. Диагностика материалов. 1986. Т.52. № 1. С. 56-58.

35. Орлов А.И. Вероятностно-статистическое моделирование помех, создаваемых электровозами // Научный журнал КубГАУ. 2015. № 106. С. 225 – 238.

36. Орлов А.И. Теория люсианов // Научный журнал КубГАУ. 2014. № 101. С. 275 – 304.

37. Орлов А.И. О реальных возможностях бутстрепа как статистического метода // Заводская лаборатория. Диагностика материалов. 1987. Т.53. № 10. С. 82-85.

38. Айвазян С.А., Енюков И.С., Мешалкин Л.Д. Прикладная статистика. Основы моделирования и первичная обработка данных. - М.: Финансы и статистика, 1983. - 472 с.

39. Хастингс Н., Пикок Дж. Справочник по статистическим распределениям. - М.: Статистика, 1980. — 95 с.

40. Фомин В.Н. Нормирование показателей надежности. - М.: Изд-во стандартов, 1986. - 140 с.

41. Кокс Д., Хинкли Д. Теоретическая статистика. - М.: Мир, 1978. - 560 с.

42. Орлов А.И. Устойчивые математические методы и модели // Заводская лаборатория. Диагностика материалов. 2010. Т.76. № 3. С. 59-67.

43. Орлов А.И. Взаимосвязь предельных теорем и метода Монте-Карло // Научный журнал КубГАУ. 2015. № 114. С. 27–41.

44. Орлов А.И. Метод статистических испытаний в прикладной статистике // Заводская лаборатория. Диагностика материалов. 2019. Т.85. №5. С. 67-79.

45. Орлов А.И. Применение метода Монте-Карло при изучении свойств статистических критериев однородности двух независимых выборок // Научный журнал КубГАУ. 2019. №154. С. 55–83.

Литература к главе 9

1. Орлов А.И. Высокие статистические технологии - из науки в преподавание // Образование через науку. Тезисы докладов Международной конференции (Москва, 2005 г.). - М.: МГТУ им. Н.Э. Баумана, 2005. - С. 555-556.

2. Орлов А.И. Эконометрика. Изд. 3-е, перераб. и доп. Учебник для вузов. – М.: Экзамен, 2004. – 576 с.

3. Орлов А.И. Эконометрика. Изд. 4-е, доп. и перераб. Учебник для вузов.– Ростов-на-Дону: Феникс, 2009. - 572 с.

4. Агаларов З.С,, Орлов А.И. Эконометрика. Учебник. - М.: Издательско-торговая корпорация «Дашков и К°», 2021. — 380 с.

5. Орлов А.И. Научная школа кафедры «Экономика и организация производства» в области эконометрики / Четвёртые Чарновские Чтения. Сборник трудов. Материалы IV международной научной конференции по организации производства. Москва, 5-6 декабря 2014 г. – М.: НП «Объединение контроллеров», 2014. – С. 326 -337.

6. Орлов А.И. Отечественная научная школа в области эконометрики // Научный журнал КубГАУ. 2016. №121. С. 235–261.

7. Орлов А.И. Отечественная научная школа в области организационно-экономического моделирования, эконометрики и статистики // Контроллинг. 2019. №73. С. 28-35.

8. Эконометрика : учебник для вузов / И.И. Елисеева [и др.] ; под редакцией И.И. Елисеевой. - М.: Издательство Юрайт, 2020. - 449 с.

9. Магнус Я.Р., Катышев П.К., Пересецкий А.А. Эконометрика. Начальный курс: Учеб. — 6-е изд., перераб. и доп. - М.: Дело, 2004. - 576 с.

10. Орлов А.И. Новая парадигма разработки и преподавания организационно-экономического моделирования, эконометрики и статистики в техническом университете / Статистика и прикладные исследования: сборник трудов Всерос. научн. конф. – Краснодар: Издательство КубГАУ, 2011. – С. 131-144.

11. Орлов А.И. Новая парадигма прикладной статистики // Статистика и прикладные исследования: сборник трудов Всерос. научн. конф. – Краснодар: Издательство КубГАУ, 2011. – С. 206-217.

12. Орлов А.И. Новая парадигма прикладной статистики // Заводская лаборатория. Диагностика материалов. 2012. Том 78. №1, часть I. С. 87-93.

13. Орлов А.И. Новая парадигма организационно-экономического моделирования, эконометрики и статистики / Вторые Чарновские Чтения. Материалы II международной научной конференции по организации производства (Москва, 7 – 8 декабря 2012 г.). Сборник тезисов. – М.: НП «Объединение контроллеров», 2012. – С. 116-120.

14. Орлов А.И. О новой парадигме организационно-экономического моделирования, эконометрики и статистики // Стратегическое планирование и развитие предприятий. Материалы Четырнадцатого всероссийского симпозиума (Москва, 9-10 апреля 2013 г.). Под ред. чл.-корр. РАН Г.Б. Клейнера. Секция 2. - М.: ЦЭМИ РАН, 2013. - С. 140-142.

15. Орлов А.И. О новой парадигме математических методов и моделей социально-экономических процессов // Материалы республиканской научно-практической конференции «Новые теоремы молодых математиков – 2013». – Наманган: Наманганском Государственный Университет, 2013. - С. 49-52.

16. Орлов А.И. Основные положения новой парадигмы организационно-экономического моделирования, эконометрики и статистики/ / Вторые Чарновские чтения. Материалы II международной научной конференции по организации производства (Москва, 7 – 8 декабря 2012 г.). Сборник трудов. – М.: НП «Объединение контроллеров», 2013. – С. 106-117.

17. Орлов А.И. Основные черты новой парадигмы математической статистики / Научный журнал КубГАУ. 2013. №90. С. 188-214.

18. Орлов А.И. О новой парадигме математических методов и моделей социально-экономических процессов / Актуальные вопросы экономики и финансов в условиях современных вызовов российского и мирового хозяйства: материалы международной научно-практической конференции НОУ ВПО «СИ ВШПП», 25 марта 2013 г. [редкол.: А.В. Бирюков, А.А. Бельцер, М.Н. Коростелева, К.Н. Ермолаев, О.А. Подкопаев (отв. ред.)] Ч. 2. – Самара: ООО «Издательство Ас Гард», 2013. – С. 400-404.

19. Орлов А.И. Новая парадигма математических методов экономики / Экономический анализ: теория и практика. 2013. №339. С.25–30.

20. Орлов А.И. О новой парадигме прикладной математики // Философия математики: актуальные проблемы. Математика и реальность. Тезисы Третьей всероссийской научной конференции (27-28 сентября 2013 г.) Редкол.: Бажанов В.А. и др. – М.: Центр стратегической конъюнктуры, 2013. – С. 84–87.

21. Орлов А.И. О новой парадигме математического моделирования при управлении развитием крупномасштабных систем // Управление развитием крупномасштабных систем (MLSD'2013). Материалы Седьмой международной конференции (30 сентября – 2 окт. 2013 г.), в 2 т. Под общ. ред. С.Н. Васильева, А.Д. Цвиркуна. Т.1. Пленарные доклады, секции 1 – 3. – М.: ИПУ РАН, 2013. – С. 297 – 299.

22. Орлов А.И. О новой парадигме прикладной математической статистики / Статистические методы оценивания и проверки гипотез: межвуз. сб. науч. тр. Вып. 25. – Пермь: Перм. гос. нац. иссл. ун-т, 2013. –С. 162-176.

23. Орлов А.И. Новая парадигма анализа статистических и экспертных данных в задачах экономики и управления / Научный журнал КубГАУ. 2014. №98. С. 105–125.

24. Орлов А.И. Новая парадигма математических методов исследования / Заводская лаборатория. Диагностика материалов. 2015. Т.81. №.7 С. 5-5.

25. Орлов А.И. О новой парадигме математических методов исследования / Научный журнал КубГАУ. 2016. №122. С. 807–832.

26. Орлов А.И. Статистическое образование в соответствии с новой парадигмой прикладной статистики / Россия: тенденции и перспективы развития. Ежегодник. Вып. 13 Ч. 1. Отв. ред. В.И. Герасимов. – М.: ИНИОН РАН. Отд. науч. сотрудничества, 2018. - С. 868-874.

27 Орлов А. Статистическое образование в соответствии с новой парадигмой прикладной статистики / Экономист. 2018. №10.

28. Орлов А.И. Смена парадигм в прикладной статистике // Заводская лаборатория. Диагностика материалов. 2021. Т.87. № 7. С. 6-7.

29. Орлов А.И. Эконометрическая поддержка контроллинга // Контроллинг. 2002. №1. С. 42-53.

30. Орлов А.И., Орлова Л.А. Применение эконометрических методов при решении задач контроллинга // Контроллинг. 2003. №4(8). С. 50-54.

31. Орлов А.И., Орлова Л.А. Эконометрика в обучении контроллеров // Контроллинг. 2004. №3 (11). С. 68-73.

32. Орлов А.И. Эконометрические инструменты контроллинга // Научный журнал КубГАУ. 2015. № 107. С. 1073–1101.

33. Орлов А.И. Эконометрика для контроллеров // Научный журнал КубГАУ. 2015. № 107. С. 1049–1072.

34. Орлов А.И. Современные эконометрические методы - интеллектуальные инструменты инженера, управленца и экономиста // Научный журнал КубГАУ. 2016. № 116. С. 484 – 514.

35. Орлов А.И. Эконометрика как учебная дисциплина // Научный журнал КубГАУ. 2017. №128. С. 679 – 709.

36. Орлов А.И. Метод ценообразования на основе оценивания функции спроса // Научный журнал КубГАУ. 2020. №158. С. 250 – 267.

37. Орлов А.И. Прикладная статистика. Учебник для вузов. — М.: Экзамен, 2006. — 672 с.

38.Орлов А.И. Распределения реальных статистических данных не являются нормальными // Научный журнал КубГАУ. 2016. № 117. С. 71–90.

39. Орлов А.И. Многообразие моделей регрессионного анализа (обобщающая статья) / Заводская лаборатория. Диагностика материалов. 2018. Т.84. №5. С. 63-73.

40. Орлов А.И. Вероятностно-статистические модели корреляции и регрессии / Научный журнал КубГАУ. 2020. №160. С. 130–162.

41. Орлов А.И. Оценка инфляции по независимой информации // Научный журнал КубГАУ. 2015. № 108. С. 259–287.

42. Куликова С.Ю., Муравьева В.С., Орлов А.И. Контроллинг динамики потребительских цен и прожиточного минимума // Научный журнал КубГАУ. 2017. №126. С. 403–421.

43. Орлов А.И. Организационно-экономическое моделирование : учебник : в 3 ч. Ч.2. Экспертные оценки. — М.: Изд-во МГТУ им. Н. Э. Баумана, 2011. — 486 с.

44. Орлов А.И. Теория экспертных оценок в нашей стране // Научный журнал КубГАУ. 2013. № 93. С. 1-11.

45. Орлов А.И. О развитии теории принятия решений и экспертных оценок // Научный журнал КубГАУ. 2021. № 167. С. 177–198.

46. Орлов А.И. Анализ экспертных упорядочений // Научный журнал КубГАУ. 2015. №112. С. 21–51.

47. Орлов А.И. Прикладная теория измерений / Прикладной многомерный статистический анализ. Ученые записки по статистике, т.33. - М.: Наука, 1978. - С. 68-138.

48. Орлов А.И. Репрезентативная теория измерений и ее применения / Заводская лаборатория. Диагностика материалов. 1999. Т.65. №3. С. 57-62.

49. Орлов А.И. Математические методы исследования и теория измерений // Заводская лаборатория. Диагностика материалов. 2006. Т.72. №1. С. 67-70.

50. Орлов А.И. Теория измерений как часть методов анализа данных: размышления над переводом статьи П.Ф. Веллемана и Л. Уилкинсона / Социология: методология, методы, математическое моделирование. 2012. № 35. С. 155-174.

51. Орлов А.И. Формализация логики правдоподобных рассуждений на основе теории измерений // Научный журнал КубГАУ. 2020. №164. С. 304–317.

52. Орлов А.И. О средних величинах / Управление большими системами. Выпуск 46. - М.: ИПУ РАН, 2013. - С. 88-117.

53. Орлов А.И. Характеризация средних величин шкалами измерения // Научный журнал КубГАУ. 2017. №134. С. 877–907.

54. Орлов А.И. Многообразие рисков // Научный журнал КубГАУ. 2015. № 111. С. 53-80.

55. Орлов А.И. Современное состояние контроллинга рисков // Научный журнал КубГАУ. 2014. № 98. С. 933-942.

56. Орлов А.И. Контроллинг рисков как научная, практическая и учебная дисциплина // Научный журнал КубГАУ. 2021. – №04(168). С. 154 – 185.

57. Орлов А.И. Аддитивно-мультипликативная модель оценки рисков при создании ракетно-космической техники // Научный журнал КубГАУ. 2014. № 102. С. 78–111.

58. Орлов А.И. Организационно-экономическое моделирование: учебник : в 3 ч. Часть 1: Нечисловая статистика. – М.: Изд-во МГТУ им. Н.Э. Баумана. – 2009. – 541 с.

59. Орлов А.И. Средние величины и законы больших чисел в пространствах произвольной природы // Научный журнал КубГАУ. 2013. № 89. С. 556 – 586.

60. Орлов А.И. О развитии статистики объектов нечисловой природы // Научный журнал КубГАУ. 2013. № 93. С. 41-50.

61. Орлов А.И. Многообразие объектов нечисловой природы // Научный журнал КубГАУ. 2014. № 102. С. 32 – 63.

62. Орлов А.И. Вероятностные модели порождения нечисловых данных // Научный журнал КубГАУ. 2015. № 105. С. 39–66.

63. Орлов А.И. Статистика нечисловых данных - центральная часть современной прикладной статистики // Научный журнал КубГАУ. 2020. №156. С. 111–142.

64. Кара-Мурза С.Г., Батчиков С.А., Глазьев С.Ю. Куда идет Россия. Белая книга реформ. — М.: Алгоритм, 2008. — 448 с.

65. Лившиц В.Н., Лившиц С.В. Системный анализ нестационарной экономики России (1992—2009): рыночные реформы, кризис, инвестиционная политика. — М.: Поли Принт Сервис, 2010. - 444 с.

66. Кара-Мурза С.Г., Гражданкин А.И. Белая книга России. Строительство, перестройка и реформы. 1950-2014. - М.: ООО «ТД Алгоритм», 2016. - 728 с.

67. Орлов А.И. Теория принятия решений. Учебник для вузов. — М.: Экзамен, 2006. — 576 с.

68. Орлов А.И. Всегда ли нужен контроль качества продукции у поставщика? // Научный журнал КубГАУ. 2014. № 96. С. 709-724.

69. Орлов А.И. Основные проблемы контроллинга качества // Научный журнал КубГАУ. 2015. № 111. С. 20-52.

70. Орлов А.И. Асимптотические методы статистического контроля // Научный журнал КубГАУ. 2014. № 102. С. 1–31.

71. Орлов А.И. Метод проверки гипотез по совокупности малых выборок и его применение в теории статистического контроля // Научный журнал КубГАУ. 2014. № 104. С. 38–52.

72. Орлов А.И. Предельные теоремы в статистическом контроле // Научный журнал КубГАУ. 2016. № 116. С. 462 – 483.

73. Орлов А.И. Организационно-экономическое моделирование и искусственный интеллект в цифровой экономике (на примере управления качеством) // Научный журнал КубГАУ. 2021. №169. С. 216 – 242.

74. Орлов А.И. О проверке симметрии распределения / Теория вероятностей и ее применения.1972. Т.17. №2. С.372-377.

75. Орлов А.И. О проверке однородности связанных выборок // Научный журнал КубГАУ. 2016. № 123. С. 708–726.

76. Орлов А.И. Модель анализа совпадений при расчете непараметрических ранговых статистик // Заводская лаборатория. Диагностика материалов. 2017. Т.83. №11. С. 66-72.

77. Орлов А.И., Луценко Е.В. Системная нечеткая интервальная математика. Монография (научное издание). – Краснодар, КубГАУ. 2014. – 600 с.

78. Орлов А.И. Задачи оптимизации и нечеткие переменные. — М.: Знание, 1980. — 64 с.

79. Орлов А.И., Луценко Е.В. Системная нечеткая интервальная математика (СНИМ) – перспективное направление теоретической и вычислительной математики // 2013. № 91. С. 163-215.

80. Луценко Е.В., Орлов А.И. Когнитивные функции как обобщение классического понятия функциональной зависимости на основе теории информации в системной нечеткой интервальной математике // Научный журнал КубГАУ 2014. №95. С. 122 – 183.

81. Орлов А.И. Статистика нечетких данных // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета. 2016. № 119. С. 75–91.

82. Орлов А.И. Системная нечеткая интервальная математика - основа математики XXI века // Научный журнал КубГАУ. 2021. №165. С. 111–130.

83. Орлов А.И. Теория нечетких множеств – часть теории вероятностей // Научный журнал КубГАУ. 2013. № 92. С. 51-60.

84. Орлов А.И. Теория принятия решений. Учебник для вузов. — М.: Экзамен, 2006. — 576 с.

85. Орлов А.И. Основные идеи статистики интервальных данных // Научный журнал КубГАУ. 2013. № 94. С. 55-70.

86. Орлов А.И. Оценка погрешностей характеристик финансовых потоков инвестиционных проектов в ракетно-космической промышленности // Научный журнал КубГАУ. 2015. № 109. С. 238–264.

87. Орлов А.И. Базовые результаты математической теории классификации // Научный журнал КубГАУ. 2015. № 110. С. 219–239.

88. Орлов А.И. Оценки плотности распределения вероятностей в пространствах произвольной природы // Научный журнал КубГАУ. 2014. № 99. С. 15-32.

89. Орлов А.И. Предельные теоремы для ядерных оценок плотности в пространствах произвольной природы // Научный журнал КубГАУ. 2015. № 108. С. 316 – 333.

90. Орлов А.И. Непараметрические ядерные оценки плотности вероятности в дискретных пространствах // Научный журнал КубГАУ. 2016. № 122. С. 833 –855.

91. Орлов А.И. Асимптотика оценок плотности распределения вероятностей // Научный журнал КубГАУ. 2017. №131. С. 845–873.

92. Орлов А.И. Прогностическая сила – наилучший показатель качества алгоритма диагностики // Научный журнал КубГАУ. 2014. № 99. С. 33-–49.

93. Орлов А.И. Основные требования к методам анализа данных (на примере задач классификации) // Научный журнал КубГАУ. 2020. №159. С. 239 – 267.

94. Лындина М.И., Орлов А.И. Математическая теория рейтингов // Научный журнал КубГАУ. 2015. № 114. С. 1 – 26.

95. Орлов А.И. Методы принятия управленческих решений: учебник. - М.: КНОРУС, 2018. - 286 с.

96. Орлов А.И., Цисарский А.Д. Определение приоритетности реализации НИОКР на предприятиях ракетно-космической отрасли // Контроллинг. 2020. № 2(76). С. 58-65.

97. Орлов А.И. Основные этапы становления статистических методов // Научный журнал КубГАУ. 2014. № 97. С. 73-85.

98. Орлов А.И. Непараметрическая и прикладная статистика в нашей стране // Научный журнал КубГАУ. 2014. № 101. С. 197–226.

99. Лойко В.И., Луценко Е.В., Орлов А.И. Высокие статистические технологии и системно-когнитивное моделирование в экологии : монография. – Краснодар : КубГАУ, 2019. – 258 с.

100. Орлов А.И. Организационно-экономическое моделирование и искусственный интеллект в организации производства в эпоху цифровой экономики // Инновации в менеджменте. 2021. № 2(28). С. 36-45.

101. Загонова Н.С. Разработка организационной системы информационной поддержки управления продуктовыми инновациями на промышленных предприятиях на основе эконометрических методов : 08. 00. 05 : дис. кэн / Загонова Н. С. ; МГТУ им. Н. Э. Баумана. - М. : Изд-во МГТУ им. Н. Э. Баумана, 2004. - 160 с.

102. Емельянова Е.А., Орлов А.И. Методы прогнозирования продаж на предприятиях оптовой торговли // Контроллинг. 2018. №1 (67). С. 68-76.

103. Куликова С.Ю., Муравьева В.С., Орлов А.И. Структура современной эконометрики в ее преподавании // Актуальные вопросы экономики, менеджмента и инноваций: материалы Международной научно-практической конференции. – Нижегород. гос. техн. ун-т им. Р.Е. Алексеева. – Нижний Новгород, 2021. – С. 304-316.

104. Куликова С.Ю., Муравьева В.С., Орлов А.И. Современная эконометрика и ее преподавание // Контроллинг. 2022. №1 (83).

Литература к главе 10

1. Налимов В.В., Мульченко З.М. Наукометрия. Изучение развития науки как информационного процесса. - М.: Наука, 1969. - 192 с.

2. Маркс К., Энгельс Ф. Сочинения. 2 изд. Т. 20, с. 37.

2. Орлов А.И., Луценко Е.В. Системная нечеткая интервальная математика. Монография (научное издание). – Краснодар, КубГАУ. 2014. – 600 с.

4. Луценко Е.В. Автоматизированный системно-когнитивный анализ [Электронный ресурс] URL: http://lc.kubagro.ru/aidos/ASK-analysis.htm (дата обращения 20.09.2020).

5. Кульбак С. Теория информации и статистика: Пер. с англ. - М. : Наука, 1967. - 408 с.

6. Гнеденко Б.В. Курс теории вероятностей. 8-е изд., испр. и доп.—М.: Едиториал УРСС, 2005.— 448 с.

7. Орлов А.И. Прикладная статистика. — М.: Экзамен, 2006. — 672 с.

8. Орлов А.И. Теория принятия решений. — М.: Экзамен, 2006. — 576 с.

9. Орлов А.И. Организационно-экономическое моделирование: : учебник : в 3 ч. Ч.1: Нечисловая статистика. — М.: Изд-во МГТУ им. Н. Э. Баумана, 2009. — 542 с.

10. Орлов А.И. Статистика нечисловых данных - центральная часть современной прикладной статистики / Научный журнал КубГАУ. 2020. № 156. С. 111–142.

11. Орлов А.И. Основные идеи статистики интервальных данных / Научный журнал КубГАУ. 2013. №94. С. 867–892.

12. Орлов А.И. Статистика интервальных данных (обобщающая статья) / Заводская лаборатория. Диагностика материалов. 2015. Т. 81. № 3. С. 61-69.

13. Орлов А.И. Организационно-экономическое моделирование: теория принятия решений. — М. : КноРус, 2020. — 568 с.

14. Орлов А.И. Методы принятия управленческих решений. - М.: КНОРУС, 2018. - 286 с.

15. Орлов А.И., Луценко Е.В., Лойко В.И. Перспективные математические и инструментальные методы контроллинга. Под научной ред. проф. С.Г. Фалько. Монография (научное издание). – Краснодар, КубГАУ. 2015. – 600 с.

16. Орлов А.И., Луценко Е.В., Лойко В.И. Организационно-экономическое, математическое и программное обеспечение контроллинга, инноваций и менеджмента: монография / под общ. ред. С. Г. Фалько. – Краснодар : КубГАУ, 2016. – 600 с.

17. Лойко В. И., Луценко Е. В., Орлов А. И. Современные подходы в наукометрии: монография / Под науч. ред. проф. С. Г. Фалько. – Краснодар: КубГАУ, 2017. – 532 с.

18. Лойко В.И., Луценко Е.В., Орлов А.И. Современная цифровая экономика. – Краснодар: КубГАУ, 2018. – 508 с.

19. Лойко В.И., Луценко Е.В., Орлов А.И. Высокие статистические технологии и системно-когнитивное моделирование в экологии : монография. – Краснодар : КубГАУ, 2019. – 258 с.

20. Орлов А.И. Системная нечеткая интервальная математика - основа математики XXI века // Научный журнал КубГАУ. 2021. №165. С. 111–130.

Литература к главе 12

1. Луценко Е.В. Автоматизированный системно-когнитивный анализ в управлении активными объектами (системная теория информации и ее применение в исследовании экономических, социально-психологических, технологических и организационно-технических систем): Монография (научное издание). – Краснодар: КубГАУ. 2002. – 605 с. http://elibrary.ru/item.asp?id=18632909

2. Lutsenko E.V. Automated system-cognitive analysis in the management of active objects (a system theory of information and its application in the study of economic, socio-psychological, technological and organizational-technical systems) // March 2019, Publisher: KubSAU, ISBN: 5-94672-020-1, https://www.researchgate.net/publication/331745417

3. Lutsenko E.V. Theoretical foundations, mathematical model and software tools for Automated system-cognitive analysis // July 2020, DOI: 10.13140/RG.2.2.21918.15685, License CC BY-SA 4.0, https://www.researchgate.net/publication/343057312

4. Сайт проф.Е.В.Луценко: http://lc.kubagro.ru/

5. Блог Е.В.Луценко в RG https://www.researchgate.net/profile/Eugene-Lutsenko

6. Луценко Е.В. Метризация измерительных шкал различных типов и совместная сопоставимая количественная обработка разнородных факторов в системно-когнитивном анализе и системе «Эйдос» / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2013. – №08(092). С. 859 – 883. – IDA [article ID]: 0921308058. – Режим доступа: https://www.researchgate.net/publication/331801337, 1,562 у.п.л.

7. Орлов А.И., Луценко Е.В. Системная нечеткая интервальная математика. Монография (научное издание). – Краснодар, КубГАУ. 2014. – 600 с. ISBN 978-5- 94672-757-0. http://elibrary.ru/item.asp?id=21358220/.

8. Луценко Е.В. Инвариантное относительно объемов данных нечеткое мультиклассовое обобщение F-меры достоверности моделей Ван Ризбергена в АСКанализе и системе «Эйдос» / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2017. – №02(126). С. 1 – 32. – IDA [article ID]: 1261702001. – Режим доступа: http://ej.kubagro.ru/2017/02/pdf/01.pdf 2 у.п.л.

9. Луценко Е.В. Количественный автоматизированный SWOT- и PEST-анализ средствами АСК-анализа и интеллектуальной системы «Эйдос-X++» / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2014. – №07(101). С. 1367 – 1409. – IDA [article ID]: 1011407090. – Режим доступа: http://ej.kubagro.ru/2014/07/pdf/90.pdf 2,688 у.п.л.

10. Луценко Е.В. Развитый алгоритм принятия решений в интеллектуальных системах управления на основе АСК-анализа и системы «Эйдос» / Е.В. Луценко, Е.К. Печурина, А.Э. Сергеев // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2020. – №06(160). С. 95 – 114. – IDA [article ID]: 1602006009. – Режим доступа: http://ej.kubagro.ru/2020/06/pdf/09.pdf, 1,25 у.п.л.

11. Луценко Е.В. Метод когнитивной кластеризации или кластеризация на основе знаний (кластеризация в системно-когнитивном анализе и интеллектуальной системе «Эйдос») / Е.В. Луценко,

В.Е. Коржаков // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2011. – №07(071). С. 528 – 576. – Шифр Информрегистра: 0421100012\0253, IDA [article ID]: 0711107040. – Режим доступа: http://ej.kubagro.ru/2011/07/pdf/40.pdf 3,062 у.п.л.

12. Луценко Е.В. Системная теория информации и нелокальные интерпретируемые нейронные сети прямого счета / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2003. – №01(001). С. 79 – 91. – IDA [article ID]: 0010301011. – Режим доступа: http://ej.kubagro.ru/2003/01/pdf/11.pdf 0,812 у.п.л.

13. Пойа Дьердь. Математика и правдоподобные рассуждения. // под редакцией С.А.Яновской. Пер. с английского И.А.Вайнштейна., М., Наука, 1975 — 464 с., http://ilib.mccme.ru/djvu/polya/rassuzhdenija.htm

14. Луценко Е.В. Системно-когнитивный анализ как развитие концепции смысла Шенка-Абельсона / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2004. – №03(005). С. 65 – 86. – IDA [article ID]: 0050403004. – Режим доступа: http://ej.kubagro.ru/2004/03/pdf/04.pdf, 1,375 у.п.л.

15. Луценко Е.В. АСК-анализ как метод выявления когнитивных функциональных зависимостей в многомерных зашумленных фрагментированных данных / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2005. – №03(011). С. 181 – 199. – IDA [article ID]: 0110503019. – Режим доступа: http://ej.kubagro.ru/2005/03/pdf/19.pdf, 1,188 у.п.л.

16. Луценко Е.В. Когнитивные функции как обобщение классического понятия функциональной зависимости на основе теории информации в системной нечеткой интервальной математике / Е.В. Луценко, А.И. Орлов // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2014. – №01(095). С. 122 – 183. – IDA [article ID]: 0951401007. – Режим доступа: http://ej.kubagro.ru/2014/01/pdf/07.pdf, 3,875 у.п.л.

17. Луценко Е.В. Когнитивные функции как адекватный инструмент для формального представления причинно-следственных зависимостей / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2010. – №09(063). С. 1 – 23. – Шифр Информрегистра: 0421000012\0233, IDA [article ID]: 0631009001. – Режим доступа: http://ej.kubagro.ru/2010/09/pdf/01.pdf, 1,438 у.п.л.

18. Луценко Е.В. Когнитивные функции как обобщение классического понятия функциональной зависимости на основе теории информации в системной нечеткой интервальной математике / Е.В. Луценко, А.И. Орлов // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2014. – №01(095). С. 122 – 183. – IDA [article ID]: 0951401007. – Режим доступа: http://ej.kubagro.ru/2014/01/pdf/07.pdf, 3,875 у.п.л.

19. Луценко Е.В., Система восстановления и визуализации значений функции по признакам аргумента (Система «Эйдос-map»). Пат. № 2009616034 РФ. Заяв. № 2009614932 РФ. Опубл. от 30.10.2009. – Режим доступа: http://lc.kubagro.ru/aidos/2009616034.jpg, 3,125 у.п.л.

20. Луценко Е.В. Системно-когнитивный анализ функций и восстановление их значений по признакам аргумента на основе априорной информации (интеллектуальные технологии интерполяции, экстраполяции, прогнозирования и принятия решений по картографическим базам данных) / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2009. – №07(051). С. 130 – 154. – Шифр Информрегистра: 0420900012\0066, IDA [article ID]: 0510907006. – Режим доступа: http://ej.kubagro.ru/2009/07/pdf/06.pdf, 1,562 у.п.л.

21. Луценко Е.В., Бандык Д.К., Подсистема визуализации когнитивных (каузальных) функций системы «Эйдос» (Подсистема «Эйдос-VCF»). Пат. № 2011612056 РФ. Заяв. № 2011610347 РФ 20.01.2011. – Режим доступа: http://lc.kubagro.ru/aidos/2011612056.jpg, 3,125 у.п.л.

22. Луценко Е.В. Метод визуализации когнитивных функций – новый инструмент исследования эмпирических данных большой размерности / Е.В. Луценко, А.П. Трунев, Д.К. Бандык // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2011. – №03(067). С. 240 – 282. – Шифр Информрегистра: 0421100012\0077, IDA [article ID]: 0671103018. – Режим доступа: http://ej.kubagro.ru/2011/03/pdf/18.pdf, 2,688 у.п.л.

23. Луценко Е.В. Системно-когнитивный анализ изображений (обобщение, абстрагирование, классификация и идентификация) / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2009. – №02(046). С. 146 – 164. – Шифр Информрегистра: 0420900012\0017, IDA [article ID]: 0460902010. – Режим доступа: http://ej.kubagro.ru/2009/02/pdf/10.pdf, 1,188 у.п.л.

24. Луценко Е.В. Системно-когнитивный подход к синтезу эффективного алфавита / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2009. – №07(051). С. 109 – 129. – Шифр Информрегистра: 0420900012\0067, IDA [article ID]: 0510907005. – Режим доступа: http://ej.kubagro.ru/2009/07/pdf/05.pdf, 1,312 у.п.л.

25. Луценко Е.В. Автоматизированный системно-когнитивный анализ изображений по их внешним контурам (обобщение, абстрагирование, классификация и идентификация) / Е.В. Луценко, Д.К. Бандык // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2015. – №06(110). С. 138 – 167. – IDA [article ID]: 1101506009. – Режим доступа: http://ej.kubagro.ru/2015/06/pdf/09.pdf, 1,875 у.п.л.

26. Луценко Е.В. Автоматизированный системно-когнитивный анализ изображений по их пикселям (обобщение, абстрагирование, классификация и идентификация) / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2015. – №07(111). С. 334 – 362. – IDA [article ID]: 1111507019. – Режим доступа: http://ej.kubagro.ru/2015/07/pdf/19.pdf, 1,812 у.п.л.

27. Луценко Е.В. Решение задач ампелографии с применением АСК-анализа изображений листьев по их внешним контурам (обобщение, абстрагирование, классификация и идентификация) / Е.В. Луценко, Д.К. Бандык, Л.П. Трошин // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2015. – №08(112). С. 862 – 910. – IDA [article ID]: 1121508064. – Режим доступа: http://ej.kubagro.ru/2015/08/pdf/64.pdf, 3,062 у.п.л.

28. Луценко Е.В. Идентификация видов жуков-жужелиц (Coleoptera, Carabidae) путем АСК-анализа их изображений по внешним контурам (обобщение, абстрагирование, классификация и идентификация) / Е.В. Луценко, В.Ю. Сердюк // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2016. – №05(119). С. 1 – 30. – IDA [article ID]: 1191605001. – Режим доступа: http://ej.kubagro.ru/2016/05/pdf/01.pdf, 1,875 у.п.л.

29. Луценко Е.В. Классификация жуков-жужелиц (Coleoptera, Carabidae) по видам и родам путем АСК-анализа их изображений / Е.В. Луценко, В.Ю. Сердюк // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2016. – №07(121). С. 166 – 201. – IDA [article ID]: 1211607004. – Режим доступа: http://ej.kubagro.ru/2016/07/pdf/04.pdf, 2,25 у.п.л.

30. Сердюк В.Ю. Создание обобщенных изображений родов жуков-жужелиц (Coleoptera, Carabidae) на основе изображений входящих в них видов, методом АСК-анализа / В.Ю. Сердюк, Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2016. – №09(123). С. 30 – 66. – IDA [article ID]: 1231609002. – Режим доступа: http://ej.kubagro.ru/2016/09/pdf/02.pdf, 2,312 у.п.л.

31. Луценко Е.В. Автоматизированный системно-когнитивный спектральный анализ конкретных и обобщенных изображений в системе "Эйдос" (применение теории информации и когнитивных технологий в спектральном анализе) / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2017. – №04(128). С. 1 – 64. – IDA [article ID]: 1281704001. – Режим доступа: http://ej.kubagro.ru/2017/04/pdf/01.pdf, 4 у.п.л.

32. Луценко Е.В. Идентификация типов и моделей самолетов путем АСК-анализа их силуэтов (контуров) (обобщение, абстрагирование, классификация и идентификация) / Е.В. Луценко, Д.К. Бандык // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2015. – №10(114). С. 1316 – 1367. – IDA [article ID]: 1141510099. – Режим доступа: http://ej.kubagro.ru/2015/10/pdf/99.pdf, 3,25 у.п.л.

33. Луценко Е.В. Решение задачи классификации боеприпасов по типам стрелкового нарезного оружия методом АСК-анализа / Е.В. Луценко, С.В. Швец, Д.К. Бандык // Политематический сетевой

электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2016. – №03(117). С. 838 – 872. – IDA [article ID]: 1171603055. – Режим доступа: http://ej.kubagro.ru/2016/03/pdf/55.pdf, 2,188 у.п.л.

34. Луценко Е.В. Определение типа и модели стрелкового нарезного оружия по боеприпасам методом АСК-анализа / Е.В. Луценко, С.В. Швец // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2016. – №04(118). С. 1 – 40. – IDA [article ID]: 1181604001. – Режим доступа: http://ej.kubagro.ru/2016/04/pdf/01.pdf, 2,5 у.п.л.

35. Луценко Е. В. , Лаптев В. Н., Сергеев А. Э. Системно-когнитивное моделирование в АПК : учеб. пособие / Е. В. Луценко, В. Н. Лаптев, А. Э. Сергеев, – Краснодар : Экоинвест, 2018. – 518 с. ISBN 978-5-94215-416-5. https://elibrary.ru/item.asp?id=35649123

36. Луценко Е.В., Бандык Д.К., Интерфейс ввода изображений в систему "Эйдос" (Подсистема «Эйдос-img»). Свид. РосПатента РФ на программу для ЭВМ, Заявка № 2015614954 от 11.06.2015, Гос.рег.№ 2015618040, зарегистр. 29.07.2015. – Режим доступа: http://lc.kubagro.ru/aidos/2015618040.jpg, 2 у.п.л.

37. Lutsenko E.V. Scenario and spectral automated system-cognitive analysis // July 2021, DOI: 10.13140/RG.2.2.22981.37608, License CC BY-SA 4.0, https://www.researchgate.net/publication/353555996

Литература к разделам 13.1, 13.2 главы-13

1.Колмогоров А. Н. . О представлении непрерывных функций нескольких переменных в виде суперпозиций непрерывных функций одной переменной и сложения // ДАН СССР. — 1957. — Т. 114, вып. 5. — С. 953—956. URL: http://www.mathnet.ru/links/b6b5d33ca466fc59252c653a3020d6c2/dan22050.pdf

2.Hecht-Nielsen R. Kolmogorov's Mapping Neural Network Existence Theorem // IEEE First Annual Int. Conf. on Neural Networks, San Diego, 1987, Vol. 3, pp. 11-13.

3.Будак, Б.М. Кратные интегралы и ряды : учебник / Б.М. Будак, С.В. Фомин. – Москва : Физматлит, 2002. – 550 с. – Режим доступа: по подписке. – URL: http://isf.pskgu.ru/ebooks/bulakfomma.html (дата обращения: 25.06.2020). – ISBN 978-5-9221-0300-8. – Текст : электронный.

4.George; Lorentz. Metric entropy, widths, and superpositions of functions (англ.) // American Mathematical Monthly : journal. — 1962. — Vol. 69. — P. 469—485.

5.↑ David A. Sprecher. On the structure of continuous functions of several variables (англ.) // Transactions of the American Mathematical Society : journal. — 1965. — Vol. 115. — P. 340—355.

6.↑ Phillip A. Ostrand. Dimension of metric spaces and Hilbert's problem 13 (англ.) // Bulletin of the American Mathematical Society : journal. — 1965. — Vol. 71. — P. 619—622.

7.Лебедев Н.Н., Специальные функции и их разложения. 2-е издание , Москва.: Учпедгиз. – 1963.– 359с.

8.Пойя Д. Математика и правдоподобные рассуждения, в двух томах // Под редакцией С. А. ЯНОВСКОЙ, Перевод с английского И. А. ВАЙНШТЕЙНА, Издание 2е, исправленное, М., 'Наука', 1975г., режим доступа: https://www.mathedu.ru/text/poya_matematika_i_pravdopodobnye_rassuzhdeniya_1975/p0/

9.Луценко Е.В. Автоматизированный системно-когнитивный анализ в управлении активными объектами (системная теория информации и ее применение в исследовании экономических, социально-психологических, технологических и организационно-технических систем): Монография (научное издание). – Краснодар: КубГАУ. 2002. – 605 с. http://elibrary.ru/item.asp?id=18632909

10. Орлов А.И., Луценко Е.В. Системная нечеткая интервальная математика. Монография (научное издание). – Краснодар, КубГАУ. 2014. – 600 с. ISBN 978-5-94672-757-0. http://elibrary.ru/item.asp?id=21358220

11. Луценко Е.В. Метризация измерительных шкал различных типов и совместная сопоставимая количественная обработка разнородных факторов в системно-когнитивном анализе и системе «Эйдос» / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2013. – №08(092). С. 859 – 883. – IDA [article ID]: 0921308058. – Режим доступа: http://ej.kubagro.ru/2013/08/pdf/58.pdf, 1,562 у.п.л.

12. Симанков В.С., Луценко Е.В. Адаптивное управление сложными системами на основе теории распознавания образов. Монография (научное издание). – Краснодар: ТУ КубГТУ, 1999. - 318с. http://elibrary.ru/item.asp?id=18828433

13. Луценко Е.В. Семантическая информационная модель СК-анализа / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2008. –

№02(036). С. 193 – 211. – Шифр Информрегистра: 0420800012\0015, IDA [article ID]: 0360802012. – Режим доступа: http://ej.kubagro.ru/2008/02/pdf/12.pdf, 1,188 у.п.л.

14. Луценко Е.В. Универсальная автоматизированная система распознавания образов "ЭЙДОС". Свидетельство РосАПО №940217. Заяв. № 940103. Опубл. 11.05.94. – Режим доступа: http://lc.kubagro.ru/aidos/1994000217.jpg, 3,125 у.п.л.

15. Луценко Е.В., Шульман Б.Х., Универсальная автоматизированная система анализа и прогнозирования ситуаций на фондовом рынке «ЭЙДОС-фонд». Свидетельство РосАПО №940334. Заяв. № 940336. Опубл. 23.08.94. – Режим доступа: http://lc.kubagro.ru/aidos/1994000334.jpg, 3,125 / 3,063 у.п.л.

16. Луценко Е.В. Универсальная автоматизированная система анализа, мониторинга и прогнозирования состояний многопараметрических динамических систем "ЭЙДОС-Т". Свидетельство РосАПО №940328. Заяв. № 940324. Опубл. 18.08.94. – Режим доступа: http://lc.kubagro.ru/aidos/1994000328.jpg, 3,125 у.п.л.

17. Луценко Е.В. Универсальная когнитивная аналитическая система «Эйдос". Монография (научное издание). – Краснодар, КубГАУ. 2014. – 600 с. ISBN 978-5-94672-830-0. http://elibrary.ru/item.asp?id=22401787

18. Луценко Е.В. Развитый алгоритм принятия решений в интеллектуальных системах управления на основе АСК-анализа и системы «Эйдос» / Е.В. Луценко, Е.К. Печурина, А.Э. Сергеев // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2020. – №06(160). С. 95 – 114. – IDA [article ID]: 1602006009. – Режим доступа: http://ej.kubagro.ru/2020/06/pdf/09.pdf, 1,25 у.п.л.

19. Луценко Е.В. Количественный автоматизированный SWOT- и PEST-анализ средствами АСК-анализа и интеллектуальной системы «Эйдос-Х++» / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2014. – №07(101). С. 1367 – 1409. – IDA [article ID]: 1011407090. – Режим доступа: http://ej.kubagro.ru/2014/07/pdf/90.pdf, 2,688 у.п.л.

20. Луценко Е.В. Метод когнитивной кластеризации или кластеризация на основе знаний (кластеризация в системно-когнитивном анализе и интеллектуальной системе «Эйдос») / Е.В. Луценко, В.Е. Коржаков // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2011. – №07(071). С. 528 – 576. – Шифр Информрегистра: 0421100012\0253, IDA [article ID]: 0711107040. – Режим доступа: http://ej.kubagro.ru/2011/07/pdf/40.pdf, 3,062 у.п.л.

21. Луценко Е. В. Методология системно-когнитивного прогнозирования сейсмичности : монография / Е. В. Луценко, А. П. Трунев, Н. А. Чередниченко; под общ. ред. В. И. Лойко. – Краснодар : КубГАУ, 2020. – 532 с., ISBN 978-5-907294-89-9, DOI 10.13140/RG.2.2.29617.33122, https://www.researchgate.net/publication/340116509_METHODOLOGY_OF_SYSTEM-COGNITIVE_FORECASTING_OF_SEISMICITY

22. Луценко Е.В. Инвариантное относительно объемов данных нечеткое мультиклассовое обобщение F-меры достоверности моделей Ван Ризбергена в АСК-анализе и системе «Эйдос» / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2017. – №02(126). С. 1 – 32. – IDA [article ID]: 1261702001. – Режим доступа: http://ej.kubagro.ru/2017/02/pdf/01.pdf, 2 у.п.л.

23. Луценко Е.В. Системная теория информации и нелокальные интерпретируемые нейронные сети прямого счета / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2003. – №01(001). С. 79 – 91. – IDA [article ID]: 0010301011. – Режим доступа: http://ej.kubagro.ru/2003/01/pdf/11.pdf, 0,812 у.п.л.

24. Луценко Е.В. Системно-когнитивный анализ как развитие концепции смысла Шенка-Абельсона / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2004. – №03(005). С. 65 – 86. – IDA [article ID]: 0050403004. – Режим доступа: http://ej.kubagro.ru/2004/03/pdf/04.pdf, 1,375 у.п.л.

25. Луценко Е.В. АСК-анализ как метод выявления когнитивных функциональных зависимостей в многомерных зашумленных фрагментированных данных / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2005. – №03(011). С. 181 – 199. – IDA [article ID]: 0110503019. – Режим доступа: http://ej.kubagro.ru/2005/03/pdf/19.pdf, 1,188 у.п.л.

26. Луценко Е.В. Когнитивные функции как обобщение классического понятия функциональной зависимости на основе теории информации в системной нечеткой интервальной математике / Е.В.

Луценко, А.И. Орлов // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2014. – №01(095). С. 122 – 183. – IDA [article ID]: 0951401007. – Режим доступа: http://ej.kubagro.ru/2014/01/pdf/07.pdf, 3,875 у.п.л.

27. Луценко Е.В. Решение задач статистики методами теории информации / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2015. – №02(106). С. 1 – 47. – IDA [article ID]: 1061502001. – Режим доступа: http://ej.kubagro.ru/2015/02/pdf/01.pdf, 2,938 у.п.л.

28. Луценко Е.В. Модификация взвешенного метода наименьших квадратов путем применения в качестве весов наблюдений количества информации в аргументе о значении функции (математические аспекты) / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2015. – №01(105). С. 814 – 845. – IDA [article ID]: 1051501050. – Режим доступа: http://ej.kubagro.ru/2015/01/pdf/50.pdf, 2 у.п.л.

29. Луценко Е.В. Универсальный информационный вариационный принцип развития систем / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2008. – №07(041). С. 117 – 193. – Шифр Информрегистра: 0420800012\0091, IDA [article ID]: 0410807010. – Режим доступа: http://ej.kubagro.ru/2008/07/pdf/10.pdf, 4,812 у.п.л.

30. Орлов А.И. Системная нечеткая интервальная математика (СНИМ) – перспективное направление теоретической и вычислительной математики / А.И. Орлов, Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2013. – №07(091). С. 255 – 308. – IDA [article ID]: 0911307015. – Режим доступа: http://ej.kubagro.ru/2013/07/pdf/15.pdf, 3,375 у.п.л.

31. Луценко Е.В. Модификация взвешенного метода наименьших квадратов путем применения в качестве весов наблюдений количества информации в аргументе о значении функции (математические аспекты) / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2015. – №01(105). С. 814 – 845. – IDA [article ID]: 1051501050. – Режим доступа: http://ej.kubagro.ru/2015/01/pdf/50.pdf, 2 у.п.л.

32. Луценко Е.В. Модификация взвешенного метода наименьших квадратов путем применения в качестве весов наблюдений количества информации в аргументе о значении функции (алгоритм и программная реализация) / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2014. – №10(104). С. 1371 – 1421. – IDA [article ID]: 1041410100. – Режим доступа: http://ej.kubagro.ru/2014/10/pdf/100.pdf, 3,188 у.п.л.

33. Луценко Е.В. Универсальный информационный вариационный принцип развития систем / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2008. – №07(041). С. 117 – 193. – Шифр Информрегистра: 0420800012\0091, IDA [article ID]: 0410807010. – Режим доступа: http://ej.kubagro.ru/2008/07/pdf/10.pdf, 4,812 у.п.л.

34. Луценко Е.В. Проблемы и перспективы теории и методологии научного познания и автоматизированный системно-когнитивный анализ как автоматизированный метод научного познания, обеспечивающий содержательное феноменологическое моделирование / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2017. – №03(127). С. 1 – 60. – IDA [article ID]: 1271703001. – Режим доступа: http://ej.kubagro.ru/2017/03/pdf/01.pdf, 3,75 у.п.л.

35. Луценко Е.В. Асимптотический информационный критерий качества шума / Е.В. Луценко, А.И. Орлов // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2016. – №02(116). С. 1569 – 1618. – IDA [article ID]: 1161602100. – Режим доступа: http://ej.kubagro.ru/2016/02/pdf/100.pdf, 3,125 у.п.л.

36. Луценко Е.В. Исследование символьных и цифровых рядов методами теории информации и АСК-анализа (на примере числа Пи с одним миллионом знаков после запятой) / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2014. –

№05(099). С. 319 – 355. – IDA [article ID]: 0991405022. – Режим доступа: http://ej.kubagro.ru/2014/05/pdf/22.pdf, 2,312 у.п.л.

37. Сайт проф.Е.В.Луценко: http://lc.kubagro.ru/

38. Проф.Е.В.Луценко в RG: https://www.researchgate.net/profile/Eugene_Lutsenko

Литература к разделу 13.3 главы-13

1. Lutsenko E.V. Automated system-cognitive analysis in the management of active objects (a system theory of information and its application in the study of economic, socio-psychological, technological and organizational-technical systems) // March 2019, Publisher: KubSAU, ISBN: 5-94672-020-1, https://www.researchgate.net/publication/331745417

2. Lutsenko E.V. Theoretical foundations, mathematical model and software tools for Automated system-cognitive analysis // July 2020, DOI: 10.13140/RG.2.2.21918.15685, License CC BY-SA 4.0, https://www.researchgate.net/publication/343057312

3. Lutsenko E.V. Methods of writing scientific papers, logic and the manner in which scientific statements // February 2021, DOI: 10.13140/RG.2.2.23546.41920, License CC BY-SA 4.0, https://www.researchgate.net/publication/349039044

4. Луценко Е.В. Метризация измерительных шкал различных типов и совместная сопоставимая количественная обработка разнородных факторов в системно-когнитивном анализе и системе «Эйдос» / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2013. – №08(092). С. 859 – 883. – IDA [article ID]: 0921308058. – Режим доступа: https://www.researchgate.net/publication/331801337, 1,562 у.п.л.

5. Луценко Е.В. Проблемы и перспективы теории и методологии научного познания и автоматизированный системно-когнитивный анализ как автоматизированный метод научного познания, обеспечивающий содержательное феноменологическое моделирование / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2017. – №03(127). С. 1 – 60. – IDA [article ID]: 1271703001. – Режим доступа: http://ej.kubagro.ru/2017/03/pdf/01.pdf, 3,75 у.п.л.

6. Lutsenko E.V. Script ASC-analysis as a method for developing generalized basic functions and weight coefficients for the decomposition of a state function of an arbitrary concrete object or situation in the theorem by A.N.Kolmogorov (1957) // August 2020, DOI: 10.13140/RG.2.2.28017.92007, License CC BY-SA 4.0, https://www.researchgate.net/publication/343365649

7. Луценко Е.В., Коржаков В.Е., «Подсистема интеллектуальной системы «Эйдос-Х++», реализующая сценарный метод системно-когнитивного анализа ("Эйдос-сценарии"). Свид. РосПатента РФ на программу для ЭВМ, Гос.рег.№ 2013660738 от 18.11.2013. – Режим доступа: http://lc.kubagro.ru/aidos/2013660738.jpg, 2 у.п.л.

8. Луценко Е.В. Сценарный АСК-анализ как метод разработки на основе эмпирических данных базисных функций и весовых коэффициентов для разложения в ряд функции состояния объекта или ситуации по теореме А.Н.Колмогорова (1957) / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2020. – №07(161). С. 76 – 120. – IDA [article ID]: 1612007009. – Режим доступа: http://ej.kubagro.ru/2020/07/pdf/09.pdf, 2,812 у.п.л.

9. Луценко Е.В. Детальный численный пример сценарного Автоматизированного системно-когнитивного анализа в интеллектуальной системе "Эйдос" / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2020. – №08(162). С. 273 – 355. – IDA [article ID]: 1622008020. – Режим доступа: http://ej.kubagro.ru/2020/08/pdf/20.pdf, 5,188 у.п.л.

10. Орлов А.И., Луценко Е.В. Системная нечеткая интервальная математика. Монография (научное издание). – Краснодар, КубГАУ. 2014. – 600 с. ISBN 978-5- 94672-757-0. http://elibrary.ru/item.asp?id=21358220/.

11. Луценко Е.В. Инвариантное относительно объемов данных нечеткое мультиклассовое обобщение F-меры достоверности моделей Ван Ризбергена в АСКанализе и системе «Эйдос» / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2017. – №02(126). С. 1 – 32. – IDA [article ID]: 1261702001. – Режим доступа: http://ej.kubagro.ru/2017/02/pdf/01.pdf 2 у.п.л.

12. Луценко Е.В. Количественный автоматизированный SWOT- и PEST-анализ средствами АСК-анализа и интеллектуальной системы «Эйдос-Х++» / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал

КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2014. – №07(101). С. 1367 – 1409. – IDA [article ID]: 1011407090. – Режим доступа: http://ej.kubagro.ru/2014/07/pdf/90.pdf 2,688 у.п.л.

13. Lutsenko E.V. Theoretical foundations, mathematical model and software tools for Automated system-cognitive analysis // July 2020, DOI: 10.13140/RG.2.2.21918.15685, License CC BY-SA 4.0, https://www.researchgate.net/publication/343057312

14. Луценко Е.В. Метод когнитивной кластеризации или кластеризация на основе знаний (кластеризация в системно-когнитивном анализе и интеллектуальной системе «Эйдос») / Е.В. Луценко, В.Е. Коржаков // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2011. – №07(071). С. 528 – 576. – Шифр Информрегистра: 0421100012\0253, IDA [article ID]: 0711107040. – Режим доступа: http://ej.kubagro.ru/2011/07/pdf/40.pdf 3,062 у.п.л.

15. Луценко Е.В. Системная теория информации и нелокальные интерпретируемые нейронные сети прямого счета / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2003. – №01(001). С. 79 – 91. – IDA [article ID]: 0010301011. – Режим доступа: http://ej.kubagro.ru/2003/01/pdf/11.pdf 0,812 у.п.л.

16. Луценко Е.В. Моделирование сложных многофакторных нелинейных объектов управления на основе фрагментированных зашумленных эмпирических данных большой размерности в системно-когнитивном анализе и интеллектуальной системе «Эйдос-X++» / Е.В. Луценко, В.Е. Коржаков // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2013. – №07(091). С. 164 – 188. – IDA [article ID]: 0911307012. – Режим доступа: http://ej.kubagro.ru/2013/07/pdf/12.pdf 1,562 у.п.л.

17. Луценко Е.В. Открытая масштабируемая интерактивная интеллектуальная online среда для обучения и научных исследований на базе АСК-анализа и системы «Эйдос» / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2017. – №06(130). С. 1 – 55. – IDA [article ID]: 1301706001. – Режим доступа: http://ej.kubagro.ru/2017/06/pdf/01.pdf, 3,438 у.п.л. http://lc.kubagro.ru/aidos/Presentation_Aidos-online.pdf

18. Луценко Е.В., Открытая масштабируемая интерактивная интеллектуальная online среда «Эйдос» («Эйдос-online»). Свид. РосПатента РФ на программу для ЭВМ, Заявка № 2017618053 от 07.08.2017, Гос.рег.№ 2017661153, зарегистр. 04.10.2017. – Режим доступа: http://lc.kubagro.ru/aidos/2017661153.jpg 2 у.п.л.

19. Сайт проф.Е.В.Луценко: http://lc.kubagro.ru/

20. Блог Е.В.Луценко в RG https://www.researchgate.net/profile/Eugene-Lutsenko

21. Луценко Е.В. Автоматизированный системно-когнитивный анализ в управлении активными объектами (системная теория информации и ее применение в исследовании экономических, социально-психологических, технологических и организационно-технических систем): Монография (научное издание). – Краснодар: КубГАУ. 2002. – 605 с. http://elibrary.ru/item.asp?id=18632909

22. Lutsenko E.V. Methods of writing scientific papers, logic and the manner in which scientific statements // February 2021, DOI: 10.13140/RG.2.2.23546.41920, License CC BY-SA 4.0, https://www.researchgate.net/publication/349039044

Литература к главе-14

1. Луценко Е.В. Системно-когнитивный анализ изображений (обобщение, абстрагирование, классификация и идентификация) / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2009. – №02(046). С. 146 – 164. – Шифр Информрегистра: 0420900012\0017, IDA [article ID]: 0460902010. – Режим доступа: http://ej.kubagro.ru/2009/02/pdf/10.pdf, 1,188 у.п.л.

2. Луценко Е.В. Автоматизированный системно-когнитивный анализ изображений по их внешним контурам (обобщение, абстрагирование, классификация и идентификация) / Е.В. Луценко, Д.К. Бандык // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2015. – №06(110). С. 138 – 167. – IDA [article ID]: 1101506009. – Режим доступа: http://ej.kubagro.ru/2015/06/pdf/09.pdf, 1,875 у.п.л.

3. Луценко Е.В. Автоматизированный системно-когнитивный анализ изображений по их пикселям (обобщение, абстрагирование, классификация и идентификация) / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного

университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2015. – №07(111). С. 334 – 362. – IDA [article ID]: 1111507019. – Режим доступа: http://ej.kubagro.ru/2015/07/pdf/19.pdf, 1,812 у.п.л.

4. Луценко Е.В. Решение задач ампелографии с применением АСК-анализа изображений листьев по их внешним контурам (обобщение, абстрагирование, классификация и идентификация) / Е.В. Луценко, Д.К. Бандык, Л.П. Трошин // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2015. – №08(112). С. 862 – 910. – IDA [article ID]: 1121508064. – Режим доступа: http://ej.kubagro.ru/2015/08/pdf/64.pdf, 3,062 у.п.л.

5. Луценко Е.В. Идентификация видов жуков-жужелиц (Coleoptera, Carabidae) путем АСК-анализа их изображений по внешним контурам (обобщение, абстрагирование, классификация и идентификация) / Е.В. Луценко, В.Ю. Сердюк // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2016. – №05(119). С. 1 – 30. – IDA [article ID]: 1191605001. – Режим доступа: http://ej.kubagro.ru/2016/05/pdf/01.pdf, 1,875 у.п.л.

6. Луценко Е.В. Классификация жуков-жужелиц (Coleoptera, Carabidae) по видам и родам путем АСК-анализа их изображений / Е.В. Луценко, В.Ю. Сердюк // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2016. – №07(121). С. 166 – 201. – IDA [article ID]: 1211607004. – Режим доступа: http://ej.kubagro.ru/2016/07/pdf/04.pdf, 2,25 у.п.л.

7. Сердюк В.Ю. Создание обобщенных изображений родов жуков-жужелиц (Coleoptera, Carabidae) на основе изображений входящих в них видов, методом АСК-анализа / В.Ю. Сердюк, Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2016. – №09(123). С. 30 – 66. – IDA [article ID]: 1231609002. – Режим доступа: http://ej.kubagro.ru/2016/09/pdf/02.pdf, 2,312 у.п.л.

8. Луценко Е.В. Идентификация типов и моделей самолетов путем АСК-анализа их силуэтов (контуров) (обобщение, абстрагирование, классификация и идентификация) / Е.В. Луценко, Д.К. Бандык // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2015. – №10(114). С. 1316 – 1367. – IDA [article ID]: 1141510099. – Режим доступа: http://ej.kubagro.ru/2015/10/pdf/99.pdf, 3,25 у.п.л.

9. Луценко Е.В. Решение задачи классификации боеприпасов по типам стрелкового нарезного оружия методом АСК-анализа / Е.В. Луценко, С.В. Швец, Д.К. Бандык // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2016. – №03(117). С. 838 – 872. – IDA [article ID]: 1171603055. – Режим доступа: http://ej.kubagro.ru/2016/03/pdf/55.pdf, 2,188 у.п.л.

10. Луценко Е.В. Определение типа и модели стрелкового нарезного оружия по боеприпасам методом АСК-анализа / Е.В. Луценко, С.В. Швец // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2016. – №04(118). С. 1 – 40. – IDA [article ID]: 1181604001. – Режим доступа: http://ej.kubagro.ru/2016/04/pdf/01.pdf, 2,5 у.п.л.

11. Симанков В.С., Луценко Е.В., Лаптев В.Н. Системный анализ в адаптивном управлении: Монография (научное издание). /Под науч. ред. В.С.Симанкова. – Краснодар: ИСТЭК КубГТУ, 2001. – 258с. http://elibrary.ru/item.asp?id=21747625

12. Луценко Е.В. Автоматизированный системно-когнитивный анализ в управлении активными объектами (системная теория информации и ее применение в исследовании экономических, социально-психологических, технологических и организационно-технических систем): Монография (научное издание). – Краснодар: КубГАУ. 2002. – 605 с. http://elibrary.ru/item.asp?id=18632909

13. Луценко Е.В. Универсальная автоматизированная система распознавания образов "Эйдос" (версия 4.1).-Краснодар: КЮИ МВД РФ, 1995.- 76с. http://elibrary.ru/item.asp?id=18630282

14. Луценко Е.В. Теоретические основы и технология адаптивного семантического анализа в поддержке принятия решений (на примере универсальной автоматизированной системы распознавания образов "ЭЙДОС-5.1"). - Краснодар: КЮИ МВД РФ, 1996. - 280с. http://elibrary.ru/item.asp?id=21745340

15. Симанков В.С., Луценко Е.В. Адаптивное управление сложными системами на основе теории распознавания образов. Монография (научное издание). – Краснодар: ТУ КубГТУ, 1999. - 318с. http://elibrary.ru/item.asp?id=18828433

16. Луценко Е.В. Интеллектуальные информационные системы: Учебное пособие для студентов специальности 351400 "Прикладная информатика (по отраслям)". – Краснодар: КубГАУ. 2004. – 633 с. http://elibrary.ru/item.asp?id=18632737

17. Луценко Е.В., Лойко В.И., Семантические информационные модели управления агропромышленным комплексом. Монография (научное издание). – Краснодар: КубГАУ. 2005. – 480 с. http://elibrary.ru/item.asp?id=21720635

18. Луценко Е.В. Интеллектуальные информационные системы: Учебное пособие для студентов специальности "Прикладная информатика (по областям)" и другим экономическим специальностям. 2-е изд., перераб. и доп.– Краснодар: КубГАУ, 2006. – 615 с. http://elibrary.ru/item.asp?id=18632602

19. Луценко Е.В. Лабораторный практикум по интеллектуальным информационным системам: Учебное пособие для студентов специальности "Прикладная информатика (по областям)" и другим экономическим специальностям. 2-е изд.,перераб. и доп. – Краснодар: КубГАУ, 2006. – 318с. http://elibrary.ru/item.asp?id=21683721

20. Наприев И.Л., Луценко Е.В., Чистилин А.Н. Образ-Я и стилевые особенности деятельности сотрудников органов внутренних дел в экстремальных условиях. Монография (научное издание). – Краснодар: КубГАУ. 2008. – 262 с. http://elibrary.ru/item.asp?id=21683724

21. Луценко Е. В., Лойко В.И., Великанова Л.О. Прогнозирование и принятие решений в растениеводстве с применением технологий искусственного интеллекта: Монография (научное издание). – Краснодар: КубГАУ, 2008. – 257 с. http://elibrary.ru/item.asp?id=21683725

22. Трунев А.П., Луценко Е.В. Астросоциотипология: Монография (научное издание). – Краснодар: КубГАУ, 2008. – 264 с. http://elibrary.ru/item.asp?id=21683727

23. Луценко Е.В., Коржаков В.Е., Лаптев В.Н. Теоретические основы и технология применения системно-когнитивного анализа в автоматизированных системах обработки информации и управления (АСОИУ) (на примере АСУ вузом): Под науч. ред.д.э.н., проф. Е.В.Луценко. Монография (научное издание). – Майкоп: АГУ. 2009. – 536 с. http://elibrary.ru/item.asp?id=18633313

24. Луценко Е.В., Коржаков В.Е., Ермоленко В.В. Интеллектуальные системы в контроллинге и менеджменте средних и малых фирм: Под науч. ред. д.э.н., проф. Е.В.Луценко. Монография (научное издание). – Майкоп: АГУ. 2011. – 392 с. http://elibrary.ru/item.asp?id=21683734

25. Наприев И.Л., Луценко Е.В. Образ-Я и стилевые особенности личности в экстремальных условиях: Монография (научное издание). – Saarbrucken, Germany: LAP Lambert Academic Publishing GmbH & Co. KG,. 2012. – 262 с. Номер проекта: 39475, ISBN: 978-3-8473-3424-8.

26. Трунев А.П., Луценко Е.В. Автоматизированный системно-когнитивный анализ влияния факторов космической среды на ноосферу, магнитосферу и литосферу Земли: Под науч. ред. д.т.н., проф. В.И.Лойко. Монография (научное издание). – Краснодар, КубГАУ. 2012. – 480 с. ISBN 978-5-94672-519-4. http://elibrary.ru/item.asp?id=21683737

27. Трубилин А.И., Барановская Т.П., Лойко В.И., Луценко Е.В. Модели и методы управления экономикой АПК региона. Монография (научное издание). – Краснодар: КубГАУ. 2012. – 528 с. ISBN 978-5-94672-584-2. http://elibrary.ru/item.asp?id=21683702

28. Горпинченко К.Н., Луценко Е.В. Прогнозирование и принятие решений по выбору агротехнологий в зерновом производстве с применением методов искусственного интеллекта (на примере СК-анализа). Монография (научное издание). – Краснодар,КубГАУ. 2013. – 168 с. ISBN 978-5-94672-644-3. http://elibrary.ru/item.asp?id=20213254

29. Орлов А.И., Луценко Е.В. Системная нечеткая интервальная математика. Монография (научное издание). – Краснодар, КубГАУ. 2014. – 600 с. ISBN 978-5-94672-757-0. http://elibrary.ru/item.asp?id=21358220

30. Луценко Е.В. Универсальная когнитивная аналитическая система «Эйдос". Монография (научное издание). – Краснодар, КубГАУ. 2014. – 600 с. ISBN 978-5-94672-830-0. http://elibrary.ru/item.asp?id=22401787

31. Орлов А.И., Луценко Е.В., Лойко В.И. Перспективные математические и инструментальные методы контроллинга. Под научной ред. проф.С.Г.Фалько. Монография (научное издание). – Краснодар, КубГАУ. 2015. – 600 с. ISBN 978-5-94672-923-9. http://elibrary.ru/item.asp?id=23209923

32. Орлов А.И., Луценко Е.В., Лойко В.И. Организационно-экономическое, математическое и программное обеспечение контроллинга, инноваций и менеджмента: монография / А. И. Орлов, Е. В. Луценко, В. И. Лойко ; под общ. ред. С. Г. Фалько. – Краснодар : КубГАУ, 2016. – 600 с. ISBN 978-5-00097-154-3. http://elibrary.ru/item.asp?id=26667522

33. Лаптев В. Н., Меретуков Г. М., Луценко Е. В., Третьяк В. Г., Наприев И. Л.. : Автоматизированный системно-когнитивный анализ и система «Эйдос» в правоохранительной сфере:

монография / В. Н. Лаптев, Г. М. Меретуков, Е. В. Луценко, В. Г. Третьяк, И. Л. Наприев; под научной редакцией проф. Е. В. Луценко. – Краснодар: КубГАУ, 2017. – 634 с. ISBN 978-5-00097-226-7. http://elibrary.ru/item.asp?id=28135358

34. Луценко Е.В. Проблема референтного класса и ее концептуальное, математическое и инструментальное решение в системно-когнитивном анализе / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2008. – №09(043). С. 1 – 47. – Шифр Информрегистра: 0420800012\0130, IDA [article ID]: 0430809001. – Режим доступа: http://ej.kubagro.ru/2008/09/pdf/01.pdf, 2,938 у.п.л.

35. Рузавин Г. И., Абдукция как метод поиска и обоснования объяснительных гипотез // Теория и практика аргументации. М., 2001. с. 44.

36. Луценко Е.В. Системно-когнитивный анализ как развитие концепции смысла Шенка – Абельсона / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2004. – №03(005). С. 65 – 86. – IDA [article ID]: 0050403004. – Режим доступа: http://ej.kubagro.ru/2004/03/pdf/04.pdf, 1,375 у.п.л.

37. Луценко Е.В. Методологические аспекты выявления, представления и использования знаний в АСК-анализе и интеллектуальной системе «Эйдос» / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2011. – №06(070). С. 233 – 280. – Шифр Информрегистра: 0421100012\0197, IDA [article ID]: 0701106018. – Режим доступа: http://ej.kubagro.ru/2011/06/pdf/18.pdf, 3 у.п.л.

38. Луценко Е.В. Подборка публикаций по вопросам выявления, представления и использования знаний. Сайт: http://www.twirpx.com/file/793311/

39. Луценко Е.В. Проблемы и перспективы теории и методологии научного познания и автоматизированный системно-когнитивный анализ как автоматизированный метод научного познания, обеспечивающий содержательное феноменологическое моделирование / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2017. – №03(127). С. 1 – 60. – IDA [article ID]: 1271703001. – Режим доступа: http://ej.kubagro.ru/2017/03/pdf/01.pdf, 3,75 у.п.л.

40. Луценко Е.В. Инвариантное относительно объемов данных нечеткое мультиклассовое обобщение F-меры достоверности моделей Ван Ризбергена в АСК-анализе и системе «Эйдос» / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2017. – №02(126). С. 1 – 32. – IDA [article ID]: 1261702001. – Режим доступа: http://ej.kubagro.ru/2017/02/pdf/01.pdf, 2 у.п.л.

41. Луценко Е.В. Метризация измерительных шкал различных типов и совместная сопоставимая количественная обработка разнородных факторов в системно-когнитивном анализе и системе «Эйдос» / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2013. – №08(092). С. 859 – 883. – IDA [article ID]: 0921308058. – Режим доступа: http://ej.kubagro.ru/2013/08/pdf/58.pdf, 1,562 у.п.л.

42. Луценко Е.В. Количественный автоматизированный SWOT- и PEST-анализ средствами АСК-анализа и интеллектуальной системы «Эйдос-X++» / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2014. – №07(101). С. 1367 – 1409. – IDA [article ID]: 1011407090. – Режим доступа: http://ej.kubagro.ru/2014/07/pdf/90.pdf, 2,688 у.п.л.

43. Луценко Е.В. Системная теория информации и нелокальные интерпретируемые нейронные сети прямого счета / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2003. – №01(001). С. 79 – 91. – IDA [article ID]: 0010301011. – Режим доступа: http://ej.kubagro.ru/2003/01/pdf/11.pdf, 0,812 у.п.л.

44. Луценко Е.В. Синтез адаптивных интеллектуальных измерительных систем с применением АСК-анализа и системы «Эйдос» и системная идентификация в эконометрике, биометрии, экологии, педагогике, психологии и медицине / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2016. – №02(116). С. 1 – 60. – IDA [article ID]: 1161602001. – Режим доступа: http://ej.kubagro.ru/2016/02/pdf/01.pdf, 3,75 у.п.л.

45. Астапчук И.Л. Возбудитель сетчатой пятнистости листьев ячменя: биология, этиология, вирулентность, устойчивость растения – хозяина (краткий обзор) / И.Л. Астапчук // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2017. – №03(127). С. 604 – 627. – IDA [article ID]: 1271703041. – Режим доступа: http://ej.kubagro.ru/2017/03/pdf/41.pdf, 1,5 у.п.л.

46. Луценко Е.В. Количественный автоматизированный SWOT- и PEST-анализ средствами АСК-анализа и интеллектуальной системы «Эйдос-Х++» / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2014. – №07(101). С. 1367 – 1409. – IDA [article ID]: 1011407090. – Режим доступа: http://ej.kubagro.ru/2014/07/pdf/90.pdf 2,688 у.п.л.

47. Луценко Е.В. Метод когнитивной кластеризации или кластеризация на основе знаний (кластеризация в системно-когнитивном анализе и интеллектуальной системе «Эйдос») / Е.В. Луценко, В.Е. Коржаков // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2011. – №07(071). С. 528 – 576. – Шифр Информрегистра: 0421100012\0253, IDA [article ID]: 0711107040. – Режим доступа: http://ej.kubagro.ru/2011/07/pdf/40.pdf 3,062 у.п.л.

Литература к главе-15

1. Луценко Е.В. Синтез семантических ядер научных специальностей ВАК РФ и автоматическая классификации статей по научным специальностям с применением АСК-анализа и интеллектуальной системы «Эйдос» (на примере Научного журнала КубГАУ и его научных специальностей: механизации, агрономии и ветеринарии) / Е.В. Луценко, Н.В. Андрафанова, Н.В. Потапова // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2019. – №01(145). С. 31 – 102. – IDA [article ID]: 1451901033. – Режим доступа: http://ej.kubagro.ru/2019/01/pdf/33.pdf, 4,5 у.п.л.

2. Луценко Е.В. Формирование семантического ядра ветеринарии путем Автоматизированного системно-когнитивного анализа паспортов научных специальностей ВАК РФ и автоматическая классификация текстов по направлениям науки / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2018. – №10(144). С. 44 – 102. – IDA [article ID]: 1441810033. – Режим доступа: http://ej.kubagro.ru/2018/10/pdf/33.pdf, 3,688 у.п.л.

3. Луценко Е.В. Интеллектуальная привязка некорректных ссылок к литературным источникам в библиографических базах данных с применением АСК-анализа и системы «Эйдос» (на примере Российского индекса научного цитирования – РИНЦ) / Е.В. Луценко, В.А. Глухов // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2017. – №01(125). С. 1 – 65. – IDA [article ID]: 1251701001. – Режим доступа: http://ej.kubagro.ru/2017/01/pdf/01.pdf, 4,062 у.п.л.

4. Луценко Е.В. Применение АСК-анализа и интеллектуальной системы "Эйдос" для решения в общем виде задачи идентификации литературных источников и авторов по стандартным, нестандартным и некорректным библиографическим описаниям / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2014. – №09(103). С. 498 – 544. – IDA [article ID]: 1031409032. – Режим доступа: http://ej.kubagro.ru/2014/09/pdf/32.pdf, 2,938 у.п.л.

5. Луценко Е.В. АСК-анализ проблематики статей Научного журнала КубГАУ в динамике / Е.В. Луценко, В.И. Лойко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2014. – №06(100). С. 109 – 145. – IDA [article ID]: 1001406007. – Режим доступа: http://ej.kubagro.ru/2014/06/pdf/07.pdf, 2,312 у.п.л.

6. Луценко Е.В. Атрибуция анонимных и псевдонимных текстов в системно-когнитивном анализе / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2004. – №03(005). С. 44 – 64. – IDA [article ID]: 0050403003. – Режим доступа: http://ej.kubagro.ru/2004/03/pdf/03.pdf, 1,312 у.п.л.

7. Луценко Е.В. Атрибуция текстов, как обобщенная задача идентификации и прогнозирования / Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2003. – №02(002). С. 146 – 164. – IDA [article ID]: 0020302013. – Режим доступа: http://ej.kubagro.ru/2003/02/pdf/13.pdf, 1,188 у.п.л.

8. Луценко Д.С., Луценко Е.В. Интеллектуальная датировка текста, определение авторства и жанра на примере русской литературы XIX и XX веков, 2020 // Статья в открытом архиве. 38 с. – DOI: 10.13140/RG.2.2.28824.01281, https://www.elibrary.ru/item.asp?id=43796415

9. Lutsenko D.S., Lutsenko E.V. Intellectual attribution of literary texts (finding the dates of the text, det ermining authorship and genre on the example of russian literature of the XIX and XX centuries), 2020 // Статья в открытом архиве. 9 p. – DOI: 10.13140/RG.2.2.15349.81122, https://www.elibrary.ru/item.asp?id=43794562

10. Ссылки на работы по применению АСК-анализу для интеллекутального анализа текстов: http://lc.kubagro.ru/aidos/Works_on_ASK-analysis_of_texts.htm

Научное издание

Орлов Александр Иванович
Луценко Евгений Вениаминович

АНАЛИЗ ДАННЫХ, ИНФОРМАЦИИ И ЗНАНИЙ В СИСТЕМНОЙ НЕЧЕТКОЙ ИНТЕРВАЛЬНОЙ МАТЕМАТИКЕ

Монография

В авторской редакции
Компьютерная верстка – Е. В. Луценко
Обложка – Е. В. Луценко

Подписано в печать 22.02.2022. Формат 60 × 84 $^{1}/_{16}$.
Усл. печ. л. – 20,0. Уч.-изд. л. – 13,75.
Тираж 500 экз. Заказ №312-50 экз.

Кубанский государственный аграрный университет.
350044, г. Краснодар, ул. Калинина, 13