УДК 519.2

Прогностическая сила как показатель качества алгоритма диагностики

А.И. Орлов

Московский государственный технический университет им. Н.Э. Баумана, Институт высоких статистических технологий и эконометрики, Москва, Россия,
prof-orlov@mail.ru; http://orlovs.pp.ru; +7(495)6994282; 123104, Москва, Сытинский пер., д.7/14, кв.14

Аннотация. *Показана нецелесообразность использования вероятности правильной диагностики в качестве показателя качества алгоритма диагностики. Предложен и изучен новый показатель – прогностическая сила, основанная на расстоянии Махаланобиса между классами. Найдено асимптотическое распределение прогностической силы, указан способ проверки адекватности ее применения. В задаче проверки двух простых гипотез установлена связь прогностической силы с расстоянием Хеллингера.*

Ключевые слова: алгоритм диагностики, показатель качества, вероятность правильной диагностики, расстояние Махаланобиса, прогностическая сила, предельные теоремы, проверка гипотез, расстояние Хеллингера.

Математические методы классификации [1] – часть прикладной статистики [2]. В одной и той же задаче классификации можно применять различные алгоритмы. Чтобы обоснованно выбрать алгоритм, необходимо уметь оценивать качество его работы. В триаде "построение классификации" – "изучение классификации" – "использование классификации" рассмотрим последнюю составляющую. Ее разные авторы называют дискриминацией (дискриминантным анализом), диагностикой, распознаванием образов с учителем, автоматической классификацией с учителем, статистической классификацией.... В соответствии с [3] эту научную область будем называть диагностикой. Для простоты обсуждения ограничимся случаем двух областей. Примеры в терминах прикладных исследований: дефектные и годные единицы продукции, налаженные и разлаженные состояния технологического процесса, пациенты с инфарктом миокарда и без него, пробы с допустимым и недопустимым содержанием примесей и т.д. Часто используют такой показатель качества алгоритма диагностики, как "вероятность (или доля) правильной классификации (диагностики)"[4, 5]

– чем этот показатель больше, тем алгоритм лучше. Цель статьи – показать нецелесообразность повсеместного применения этого показателя и обосновать другой – "прогностическую силу", найденную путем пересчета на модель линейного дискриминантного анализа.

Вероятностно-статистическая модель диагностики

Пусть классифицируемые объекты описываются переменными x, лежащими в некотором пространстве X, два класса – это два распределения вероятностей с плотностями $f(x)$ и $g(x)$, $x \in X$, соответственно (если X дискретно, то под $f(x)$ и $g(x)$ понимаем не плотности, а вероятности попадания в точку x). Типичная схема разработки конкретного математического метода диагностики такова. С помощью специалистов соответствующей прикладной области составляют две обучающие выборки – объема m_0 из первого класса и объема n_0 из второго класса. На их основе определяют решающее правило $A : X \to \{1, 2\}$, ставящее в соответствие результату наблюдения $x \in X$ номер класса, к которому его следует отнести. Качество работы алгоритма проверяют по контрольной выборке, состоящей из m элементов первого класса и n элементов второго. Результаты проверки удобно представить в виде табл.1.

Таблица 1

Результаты работы алгоритма диагностики

	Всего	Отнесено к первому классу	Отнесено ко второму классу
Элементы первого класса	m	a	b
Элементы второго класса	n	c	d

Ясно, что о качестве алгоритма диагностики, т.е. решающего правила A, надо судить на основе данных, приведенных в табл.1. Естественно использовать

$\kappa = a/m$ – долю правильной диагностики в первом классе;

$\lambda = d/n$ – долю правильной диагностики во втором классе.

Доля правильной диагностики μ равна

$$\mu = \frac{a+d}{m+n} = \pi_1\kappa + \pi_2\lambda,$$

где $\pi_1 = \dfrac{m}{m+n}$ – априорная доля первого класса, $\pi_2 = \dfrac{n}{m+n}$ – априорная доля второго класса. Очевидно, $\pi_1 + \pi_2 = 1$. В вероятностной модели речь идет об априорных вероятностях классов. Удобно и в статистической постановке (табл.1), и в вероятностной использовать единый термин "доля", поскольку это не приводит к недоразумениям.

Рассмотрим сначала случай, когда $\min(m_0, n_0) \to \infty$, т.е. f и g можно считать известными. При применении теории статистических решений к задачам диагностики [5] считаются известными также априорные вероятности классов и потери $C(j|i)$ от ошибочной диагностики – от отнесения объекта i-го класса к классу с номером j. Доказано (при любом числе классов), что в случае, когда все $C(j|i)$ равны между собой, минимизация суммарных потерь эквивалентна максимизации вероятности правильной диагностики [5], состоятельной оценкой которой служит μ. *Видимо, этот факт, вместе с вычислительной простотой и наглядностью μ, является причиной широкого использования доли (вероятности) правильной классификации μ как показателя качества алгоритма диагностики.*

Оптимальное решающее правило при совпадающих потерях $C(1|2) = C(2|1)$ имеет вид: если

$$\frac{f(x)}{g(x)} > \frac{\pi_2(\infty)}{\pi_1(\infty)}, \tag{1}$$

то отнести x к первому классу, в противном случае – ко второму; здесь $\pi_1(\infty)$ и $\pi_2(\infty)$ – априорные вероятности первого и второго классов соответственно. Лемма Неймана-Пирсона, исходя из другой оптимизационной постановки (из минимизации ошибки второго рода при ограничении на уровень значимости), также дает правило, основанное на отношении плотностей вероятностей. В большинстве прикладных задач плотности вероятности неизвестны. Однако их при $\min(m_0, n_0) \to \infty$ можно заменить состоятельными оценками, например, непараметрическими ядерными оценками плотности [6], и получить *универсальное асимптотически оптимальное правило диагностики,* с которым необходимо сравнивать все другие правила диагностики [1]. Выбор других правил диагностики для решения конкретных прикладных задач должен быть обоснован на основе соответствующих задаче критериев, например, быстродействия алгоритмов и объемов имеющейся информации [7].

Обратим внимание, что в случае, когда априорная вероятность одного из классов существенно больше априорной вероятности другого, правило (1) может любое наблюдение относить к классу с наибольшей априорной вероятностью.

Разберем ситуацию подробнее. Пусть имеется некоторый алгоритм диагностики на два класса с долями правильной диагностики κ – в первом классе и λ – во втором. Сравним его с двумя тривиальными алгоритмами диагностики. Первый тривиальный алгоритм относит все классифицируемые объекты к первому классу, для него $\kappa = 1$ и $\lambda = 0$, следовательно, $\mu = \pi_1$. Второй тривиальный алгоритм относит все классифицируемые объекты ко второму классу, для него $\kappa = 0$ и $\lambda = 1$, следовательно, $\mu = \pi_2$.

В качестве показателя качества алгоритма диагностики будем использовать долю правильной диагностики μ. Когда первый тривиальный алгоритм лучше исходного? В том случае, когда $\pi_1 > \kappa\pi_1 + \lambda\pi_2$, т.е.

$$\frac{\lambda}{1 - \kappa + \lambda} < \pi_1$$

(с учетом того, что $\pi_1 + \pi_2 = 1$). Когда второй тривиальный алгоритм лучше исходного? В том случае, когда $\pi_2 > \kappa\pi_1 + \lambda\pi_2$, т.е.

$$\pi_1 < \frac{1 - \lambda}{1 + \kappa - \lambda}.$$

Таким образом, для любого заданного алгоритма диагностики существуют границы $d(1)$ и $d(2)$ для доли первого класса π_1 в объединенной контрольной выборке такие, что при $\pi_1 < d(1)$ рассматриваемый алгоритм хуже второго тривиального алгоритма, а при $\pi_1 > d(2)$ он хуже первого тривиального алгоритма.

Разобранная ситуация встречается на практике. В конкретной прикладной медицинской задаче величина μ оказалась больше для тривиального прогноза, согласно которому у всех больных течение заболевания (инфаркта миокарда) будет благоприятно. Тривиальный прогноз сравнивался с алгоритмом выделения больных с прогнозируемым тяжелым течением заболевания. Он был разработан группой математиков и кардиологов под руководством И.М. Гельфанда. Применение этого алгоритма с медицинской точки зрения вполне оправдано [8 - 10]. Итак, по доле правильной классификации μ алгоритм группы И.М. Гельфанда оказался хуже тривиального – объявить всех больных легкими, т.е. не требующими специального наблюдения. Этот вывод очевидно нелеп. И причина появления нелепости вполне понятна. Хотя доля тяжелых больных невелика, но смертельные исходы сосредоточены именно в этой группе больных. Поэтому целесообразна гипердиагностика – рациональнее часть легких больных объявить тяжелыми, чем сделать ошибку в противоположную сторону.

Поэтому мы полагаем, что использовать в качестве показателя качества алгоритма диагностики долю правильной диагностики μ нецелесообразно.

Работы группы И.М. Гельфанда [8 - 10] показывают также, что теория статистических решений не может быть основой для выбора показателя качества диагностики. Применение этой теории требует знания потерь от ошибочной диагностики, а в большинстве научно-технических и экономических задач определить также потери, как уже отмечалось, сложно. В частности, из-за необходимости оценивать человеческую жизнь в денежных единицах. По этическим соображениям это, на наш взгляд, недопустимо. Сказанное не означает отрицания пользы страхования, но, очевидно, страховые выплаты следует рассматривать лишь как способ первоначального смягчения потерь от утраты близких. Следовательно, применение теории статистических решений в рассматриваемой постановке вряд ли возможно, поскольку оценить количественно потери от смерти больного нельзя по этическим соображениям.

Прогностическая сила как показатель качества классификации

С целью поиска приемлемого показателя качества диагностики рассмотрим восходящую к Р. Фишеру [11] широко известную параметрическую вероятностную модель (модель линейного дискриминантного анализа), в которой X – конечномерное пространство, $f(x)$ и $g(x)$ - многомерные нормальные плотности с математическими ожиданиями m_1 и m_2 соответственно и совпадающими ковариационными матрицами Σ. Тогда при произвольных априорных вероятностях и потерях оптимальное решающее правило определяется плоскостью

$$H(x) = \left(x - \frac{m_1 + m_2}{2}\right)^{\top} \Sigma^{-1}(m_1 - m_2) = C,\, x \in X,$$

где константа C зависит от априорных вероятностей классов здесь $\pi_1(\infty)$ и $\pi_2(\infty)$, а также от потерь $C(1|2)$ и $C(2|1)$ [12, с.186]. Основной параметр модели - расстояние Махаланобиса между классами

$$d = \sqrt{(m_1 - m_2)^{\top}\Sigma^{-1}(m_1 - m_2)}.$$

(Величину d^2 нельзя называть "расстоянием", как это делается в [12, с.187], поскольку для d^2 не выполнено неравенство треугольника. Всем аксиомам, задающим метрику, удовлетворяет d).

Через d и C/d выражаются вероятности правильной диагностики $\kappa(\infty)$ и $\lambda(\infty)$, являющиеся пределами при $\min(m, n) \to \infty$ ранее введенных долей κ и λ. Если X – случайная величина с описанной выше плотностью $f(x)$, то случайная величина $H(X)$ имеет нормальное распределение с математическим ожиданием ${d^2}/{2}$ и дисперсией d^2. Аналогично для случайной величины Y с плотностью $g(x)$ случайная величина $H(Y)$ имеет нормальное распределение с математическим ожиданием $\left(-{d^2}/{2}\right)$ и дисперсией d^2

[12, с.187]. Поэтому

$$\kappa(\infty) = P(H(X) > C) = \Phi\left(\frac{d}{2} - \frac{C}{d}\right),$$

$$\lambda(\infty) = P(H(Y) \le C) = \Phi\left(\frac{d}{2} + \frac{C}{d}\right), \qquad (2)$$

где $\Phi(x)$ – функция нормального распределения с математическим ожиданием 0 и дисперсией 1. Из (2) вытекает, что при любых потерях $C(1|2)$ и $C(2|1)$, априорных вероятностях π_1 и π_3 при использовании оптимального решающего правила величина

$$d_0 = \Phi^{-1}(\kappa(\infty)) + \Phi^{-1}(\lambda(\infty))$$

постоянна и равна d, где $\Phi^{-1}(y)$ – функция, обратная к $\Phi(x)$. Следовательно, именно расстояние Махаланобиса d следует рассматривать как меру различия между классами, заданными плотностями рассматриваемого вида. Чтобы выразить меру различия в тех же единицах, что и вероятности правильной диагностики, введем "прогностическую силу"

$$\delta = \Phi\left({}^{d}\!/_{2}\right),$$

при $C = 0$ равную совпадающим значениям правильной диагностики $\kappa(\infty)$ и $\lambda(\infty)$ из (2).

Мы предлагаем в качестве показателя качества произвольного алгоритма диагностики использовать эмпирическую прогностическую силу

$$\delta_* = \Phi\left({}^{d_*}\!/_{2}\right), \qquad (3)$$

где

$$d_* = \Phi^{-1}(\kappa) + \Phi^{-1}(\lambda), \qquad (4)$$

а доли κ и λ правильной диагностики в классах определены по данным табл.1. Таким образом, вместо взвешенного по априорным долям классов среднего арифметического μ долей правильной диагностики в классах κ и λ предлагаем использовать их среднее по Колмогорову с весовой функцией Φ^{-1}.

Если классы описываются выборками из многомерных нормальных совокупностей с одинаковыми матрицами ковариаций, а для классификации применяется классический линейный дискриминантный анализ Р.Фишера [11, 12], то величина d_* представляет собой состоятельную статистическую оценку расстояния Махаланобиса между двумя рассматриваемыми совокупностями, причем независимо от порогового значения, определяющего конкретное решающее правило. В общем случае показатель d_* вводится как эвристический.

Пример 1. Если $\kappa = 0,90$ и $\lambda = 0,80$, то $\Phi^{-1}(k) = 1,28$ и $\Phi^{-1}(\lambda) = 0,84$, откуда $d_* = 2,12$ и эмпирическая прогностическая сила $\delta_* = \Phi^{-1}(1,06) = 0,86$. При этом доля правильной диагностики μ может принимать любые значения между 0,80 и 0,90, в зависимости от долей классов в объединенной совокупности $\pi_i, i = 1, 2$; $\pi_1 + \pi_2 = 1$.

Асимптотические свойства выборочной прогностической силы

С какой точностью можно определить прогностическую силу по выборочным данным? Другими словами, каково распределение δ_*? Рассмотрим сначала случай, когда решающее правило не зависит от элементов контрольной выборки. Тогда случайные величины $m\kappa$ и $n\lambda$ независимы и имеют биномиальные распределения с параметрами $(m, \kappa(\infty))$ и $(n, \lambda(\infty))$ соответственно.

Пусть алгоритм диагностики применялся к совокупности, состоящей из m объектов первого класса и n объектов второго класса.

Теорема 1. *Пусть* m, . *Тогда для всех*

$$P\left\{\frac{\delta_* - \delta}{A(\kappa, \lambda)} < x\right\} \to \Phi(x), \tag{5}$$

где δ *– истинная "прогностическая сила"алгоритма диагностики (соответствующая* $\kappa(\infty)$ *и* $\lambda(\infty)$*);* δ_* *– ее эмпирическая оценка, полученная по формулам (3) и (4),*

$$A^2(\kappa, \lambda) = \frac{1}{4}\left\{\left[\frac{\varphi(d_*/2)}{\varphi(\Phi^{-1}(\kappa))}\right]^2 \frac{\kappa(1-\kappa)}{m} + \left[\frac{\varphi(d_*/2)}{\varphi(\Phi^{-1}(\lambda))}\right]^2 \frac{\lambda(1-\lambda)}{m}\right\}. \tag{6}$$

Здесь $\varphi(x) = \Phi'(x)$ *– плотность стандартного нормального распределения вероятностей с математическим ожиданием 0 и дисперсией 1.*

Для доказательства теоремы 1 используются следующие факты и методы (описанные, например, в [2, гл.4]):

(1) асимптотическая нормальность вектора (κ, λ), вытекающая из теоремы Муавра-Лапласа;

(2) результаты о наследуемости сходимости, в соответствии с которыми, в частности, асимптотическое распределение δ_* при $\min(m, n) \to \infty$ совпадает с распределением, которое получится, если в формулах (3) - (4) пронормированный и отцентрированный вектор (κ, λ) заменить на соответствующий нормально распределенный вектор;

(3) метод линеаризации функции в малой окрестности точки ($\kappa(\infty)$, $\lambda(\infty)$).

Из теоремы 1 вытекает способ определения нижней и верхней доверительных границ для прогностической силы δ, а именно: доверительной вероятности γ соответствуют асимптотические (при $\min(m, n) \to \infty$) двусторонние доверительные границы. (7)

$$\delta_H = \delta_* - u(\gamma)A(\kappa, \lambda), \quad \delta_B = \delta_* + u(\gamma)A(\kappa, \lambda), \quad u(\gamma)\Phi^{-1}\left(\frac{1+\gamma}{2}\right) \qquad (7)$$

Пример 2. В условиях примера 1 при $m = n = 100$ найдем асимптотическое среднее квадратическое отклонение $(0,90; 0,80)$. Поскольку $\varphi(\Phi^{-1}(\kappa)) = \varphi(1,28) = 0,176, \varphi(\Phi^{-1}(\lambda)) = \varphi(0,84) = 0,280, \varphi(d_*/2) = \varphi(1,06) = 0,227$, то подставляя в выражение для ^{2} численные значения, получаем, что

$$A^2(0,90;0,80) = \frac{0,0372}{m} + \frac{0,0265}{n}.$$

При $m = n = 100$ имеем $(0,90; 0,80) = 0,0252$. При доверительной вероятности $\gamma = 0,95$ имеем $u(0,95) = \Phi^{-1}(0,975) = 1,96$, а потому нижняя доверительная граница для прогностической силы δ есть $\delta_H = 0,86 - 1,96 \times 0,0252 = 0,81$, при этом верхняя доверительная граница такова: $\delta_B = 0,86 + 1,96 \times 0,0252 = 0,91$. Аналогичный расчет при $m = n = 1000$ дает $\delta_H = 0,845, \delta_H = 0,875$.

Иногда отступают от описанной выше схемы и в качестве контрольной используют обучающую выборку (обычно это связано с трудоемкостью получения данных). Тогда доли правильной диагностики κ и λ зависимы как случайные величины. Однако, в случае, когда решающее правило основано на использовании дискриминантной поверхности, κ и λ асимптотически (при $\min(m, n) \to \infty$) независимы [13], что позволяет использовать формулы (5) - (7) и в этом случае.

Проверка возможности применения прогностической силы

В ряде прикладных задач диагностики вычисляют значение некоторого прогностического индекса (фактора, переменной) и решение принимают на основе его сравнении с некоторым заданным порогом . Объект относят к первому классу, если $<$, ко второму, если $>$. Прогностический индекс – это обычно линейная функция от характеристик рассматриваемых объектов. Другими словами, от координат векторов, описывающих объекты. Возьмем два значения порога c_1 и c_2. Если классы описываются выборками из многомерных нормальных совокупностей с одинаковыми матрицами ковариаций, а для построения прогностического индекса применяется классический линейный дискриминантный анализ Р.Фишера, другими словами, если пересчет на модель линейного дискриминантного анализа обоснован, то, как можно показать, "прогностические силы"для обоих правил совпадают: $\delta(c_1) = \delta(c_2)$. Выполнение этого равенства можно проверить как статистическую гипотезу. Если эта гипотеза принимается, то целесообразность использования прогностической силы подтверждается, значение $\delta^*(c_1) \approx \delta_*(c_2)$ есть основания рассматривать как объективную оценку качества алгоритма диагностики. Если же рассматриваемая гипотеза отклоняется, т.е. значения $\delta_*(c_1)$ и $\delta_*(c_2)$ сильно различаются, то пересчет на модель линейного дискриминантного анализа и использование расстояния Махаланобиса для измерения различия классов и прогностической силы как показателя качества диагностики некорректны.

Укажем способ проверки, т.е. опишем соответствующий критерий проверки статистической гипотезы. Пусть κ_1 – доля объектов первого класса, для которых $y < c_1$, а κ_2 – доля объектов первого класса, для которых $c_1 < y < c_2$. Аналогично пусть λ_2 – доля объектов второго класса, для которых $c_1 < y < c_2$, а λ_3 – доля объектов второго класса, для которых $> {}_2$. Тогда можно рассчитать две оценки одного и того же расстояния Махаланобиса. Они имеют вид:

$$d_*(c_1) = \Phi^{-1}(\kappa_1) + \Phi^{-1}(\lambda_2 + \lambda_3),\ \ d_*(c_2) = \Phi^{-1}(\kappa_1 + \kappa_2) + \Phi^{-1}(\lambda_3).$$

Теорема 2. *Если истинные прогностические силы двух правил диагностики совпадают, т.е.* $\delta(c_1) = \delta(c_2)$, *то при* $\min(m, n) \to \infty$ *при всех*

$$P\left\{\frac{d_*(c_1) - d_*(c_2)}{B} < x\right\} \to \Phi(x),$$

где

$$B^2 = \frac{1}{m}T(\kappa_1, \kappa_2) + \frac{1}{n}T(\lambda_3, \lambda_2),$$

$$T(x, y) = \frac{x(1-x)}{\varphi^2(\Phi^{-1}(x))} + \frac{(x+y)(1-x-y)}{\varphi^2(\Phi^{-1}(x+y))} - \frac{2x(1-x-y)}{\varphi(\Phi^{-1}(x))\varphi(\Phi^{-1}(x+y))}.$$

Из теоремы 2 вытекает метод проверки рассматриваемой гипотезы: при выполнении неравенства

$$\left|\frac{d_*(c_1) - d_*(c_2)}{B}\right| \le \Phi^{-1}\left(1 - \frac{\alpha}{2}\right)$$

она принимается на уровне значимости, асимптотически равном α, в противном случае – отвергается.

Пример 3. Пусть данные примеров 1 и 2 соответствуют порогу c_1, а порогу c_2 соответствуют $\kappa' = 0,95$ и $\lambda' = 0,70$. Тогда в обозначениях теоремы 2 $\kappa_1 = 0,90, \kappa_2 = 0,05, \lambda_2 = 0,10, \lambda_3 = 0,70$, Далее $d_*(c_1) = 2,12$ (см. пример 1), $d_*(c_2) = 2,17$, $T(\kappa_1, \kappa_2) = 2,22, T(\lambda_3, \lambda_2) = 0,89$. Гипотеза $\delta(c_1) = \delta(c_2)$ о совпадении прогностических сил, соответствующих двум порогам, принимается на уровне значимости $\alpha = 0,05$ тогда и только тогда, когда

$$\left(\frac{0,05^2}{\frac{2,22}{m} + \frac{0,89}{n}}\right) \leq 1,96^2,$$

т.е. когда

$$\frac{2,22}{m} + \frac{0,89}{n} \geq 0,00065.$$

Так гипотеза принимается при $m = n = 1000$ и отвергается при $m = n = 5000$.

Асимптотика сближающихся гипотез

Для выбора наилучшего алгоритма диагностики используют, кроме "вероятности правильной диагностики"μ, подход, основанный на векторной оптимизации

$$(\kappa, \lambda) \to \max.$$

При этом вводят ограничения на вероятности ошибочной классификации. Классическим примером является задача

$$\kappa \to \max, \ \ \lambda \leq \lambda_0 = const, \tag{8}$$

решение которой дается леммой Неймана-Пирсона. Дальнейшие результаты в этом направлении получены Л.Н. Большевым [14], Ю. Круописом [15] и применительно к диагностике А.С. Фоминым [16]. Однако во всех этих постановках используются априорные константы, задающие ограничения, типа λ_0 в (8). Вопрос прикладника: "Откуда берется уровень значимости?"[17] вполне оправдан. Поэтому обсудим возможность переноса на общую задачу проверки двух простых гипотез подхода, развитого выше в случае задачи диагностики.

Пусть $X_1, X_2, ..., X_n$ – выборка из распределения с плотностью f(x). Рассмотрим согласно [18] задачу проверки двух простых гипотез

$$H_1 : f(x) \equiv f_1(x),$$

$$H_2 : f(x) \equiv f_2(x)$$

(отметим, что понятия нулевой и альтернативной гипотез в [18] не используются). Как показано в [18], оптимальное решающее правило имеет вид: если

$$Y = \sum_{1 \leq i \leq n} \ln \frac{f_2(X_i)}{f_1(X_i)} < \ln D,$$

то принимается гипотеза H_1, если же $Y \geq \ln D$, то принимается гипотеза H_2, где D – априорная константа, зависящая от величин типа C и d в формуле (2) и λ_0 в формуле (8).

Рассмотрим асимптотику сближающихся гипотез [18]: плотность $f(x)$ входит в некоторое параметрическое семейство, описываемое одномерным параметром θ, удовлетворяющее введенным в [18] условиям регулярности (RR):

$$H_1 : \theta = \theta_1,$$
$$H_2 : \theta = \theta_2, \;\; \theta_2 = \theta_1 + \frac{v}{\sqrt{n}},$$

где величины θ_1 и v считаются фиксированными. Пусть, как и ранее, κ – частота принятия гипотезы H_1, когда она верна, и λ – частота принятия гипотезы H_2, когда она верна. Воспользуемся асимптотической нормальностью Y. Согласно [18] при $\min(m,n) \to \infty$

$$\kappa \to \Phi\left(\frac{\frac{v^2}{2}I(\theta_1) + \ln D}{|v|\sqrt{I(\theta_1)}}\right), \lambda \to \Phi\left(\frac{\frac{v^2}{2}I(\theta_1) - \ln D}{|v|\sqrt{I(\theta_1)}}\right), \qquad (9)$$

где $I(\theta_1)$ – информация Фишера. Из формулы (9) следует, что

$$d_* = \Phi^{-1}(\kappa) + \Phi^{-1}(\lambda) \to |v|\sqrt{I(\theta_1)}.$$

Как и в задаче диагностики, величина d_* (асимптотически) не зависит от априорной константы D. (В этом абзаце и далее предполагаем, что проведено две серии опытов: m опытов, когда верна гипотеза H_1, и n опытов, когда верна гипотеза H_2, и именно по этим двум сериям, как и в начале статьи, подсчитаны частоты κ и λ; для изучения $\kappa(\infty)$ и $\lambda(\infty)$ достаточно брать вероятности относительно соответствующих вероятностных мер [18]).

Величина d_* характеризует саму постановку задачи проверки гипотез, т.е. различие между f_1 и f_2. Действительно, из известной [18] связи между информацией Фишера и расстоянием Хеллингера следует, что при $\min(m,n) \to \infty$

$$d_*^2 \approx 4\int_{-\infty}^{\infty} \left(\sqrt{f_1(x)} - \sqrt{f_2(x)}\right)^2 dx.$$

"Прогностическая сила"δ_*, заданная формулой (3), которую в рассматриваемом контексте лучше назвать "разрешающей силой", дополняет до 1 совпадающие между собой значения ошибок первого и второго рода асимптотически минимаксного критерия [18]. Таким образом, при наличии нескольких критериев в одной задаче проверки гипотез мы предлагаем использовать "разрешающую силу"δ_* как показатель качества статистического критерия. Теорема 1 дает асимптотическое распределение этого показателя. Теорема 2 позволяет выявлять критерии, "поведение"которых отличается от "поведения"оптимальных критериев (типа Неймана-Пирсона). Роль порогов c_1 и c_2 играют уровни значимости α_1 и α_2.

Библиографический список

1. *Орлов А.И.* О развитии математических методов теории классификации // Заводская лаборатория. Диагностика материалов. - 2009. - Т.75. є 7. - С.51-63.

2. *Орлов А.И.* Прикладная статистика: учеб. для вузов. - М.: Экзамен, 2006. - 672 с.

3. *Орлов А.И.* Математические методы исследования и диагностика материалов // Заводская лаборатория. Диагностика материалов. - 2003. - Т.69. є 3. - С.53-64.

4. Алгоритмы и программы восстановления зависимостей / Под ред. *Вапника В.Я.* - М.: Наука, 1984. - 816 с.

5. *Горелик А.Л., Скрипкин В.А.* Методы распознавания: учеб. для вузов. - М.: Высшая школа, 1984. - 208 с.

6. *Орлов А.И.* Ядерные оценки плотности в пространствах произвольной природы // Статистические методы оценивания и проверки гипотез. Межвузовский сборник научных трудов. - Пермь: Пермский госуниверситет, 1996. - С.68-75.

7. *Толчеев В.О.* Модифицированный и обобщенный метод ближайшего соседа для классификации библиографических текстовых документов // Заводская лаборатория. Диагностика материалов. - 2009. - Т.75. є 7. - С.63-70.

8. *Алексеевская М.А., Гельфанд И.М., Губерман Ш.А., Мартынов И.В., Ротвайн И.М., Саблин В.М.* Прогнозирование исхода мелкоочагового инфаркта миокарда с помощью программы узнавания // Кардиология. - 1977. - Т.17, є 7. - С.26-71.

9. *Гельфанд И.М., Губерман Ш.А., Сыркин А.Л., Головня Л.Д., Извекова М.Л., Алексеевская М.А.* Прогнозирование исхода инфаркта миокарда с помощью программы "Кора-3"// Кардиология. - 1977. - Т.17, є 6. - С.19-23.

10. *Гельфанд И.М., Розенфельд Б.И., Шифрин М.А.* Очерки о совместной работе математиков и врачей (2-е, дополненное издание). - М. УРСС, 2004. - 320 с.

11. *Фишер Р.Э.* Использование множественных измерений в задачах таксономии // Современные проблемы кибернетики. - М.: Знание, 1979. С. 6 - 20. (Fisher R.A. The use of multiple measurements in taxonomic problems. - Ann. Eugenics, 1936, September, v.7, 179 - 188.)

12. *Андерсон Т.* Введение в многомерный статистический анализ. - М.: ГИФМЛ, 1963. - 500 с.

13. *Орлов А.И.* Некоторые вероятностные вопросы теории классификации // Прикладная статистика. Ученые записки по статистике, т.45. - М.: Наука, 1983. - С.166-179.

14. *Большев Л.Н.* Избранные труды. Теория вероятностей и математическая статистика. - М.: Наука, 1987. - 286 с.

15. *Круопис Ю.* О проверке многих простых гипотез // Тезисы докладов Второй Вильнюсской международной конференции по теории вероятностей и математической статистике: Том 1. - Вильнюс, 1977. - С.207-209.

16. *Фомин А.С.* Векторный критерий качества распознавания // Автоматика и телемеханика. - 1978. є3. - С.131-136.

17. *Алимов Ю.И.* Альтернатива методу математической статистики. - М.: Знание, 1980. - 62 с.

18. *Боровков А.А.* Математическая статистика. - М.: Лань, 2010. - 704 с.

Prognostic strength as an quality indicator of diagnostic algorithm

A.I. Orlov

Bauman Moscow State Technical University, Institute of high statistical technologies and econometrics, Russia, Moscow; prof-orlov@mail.ru; http://orlovs.pp.ru; +7(495) 6994282; 123104, Moscow, the Sytinsky lane, house 7/14, apartment 14

Abstract. Inexpediency of use of probability of correct diagnostics as an quality indicator of diagnostic algorithm is shown. The new indicator - the prognostic strength based on Mahalanobis distance between classes is offered and studied. It is found asymptotic distribution the prognostic strength, the way of testing of adequacy of its application is specified. In a problem of testing of two simple hypotheses the prognostic strength connection is established with Hellinger distance.

Key words: diagnostic algorithm, a quality indicator, probability of correct diagnostics, Mahalanobis distance, prognostic strength, limiting theorems, testing of hypotheses, Hellinger distance.