

DOI: <https://doi.org/10.26896/1028-6861-2023-89-7-5-7>

ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ, НЕЙРОСЕТИ, БОЛЬШИЕ ДАННЫЕ И МАТЕМАТИЧЕСКИЕ МЕТОДЫ ИССЛЕДОВАНИЯ

© Александр Иванович Орлов

Московский государственный технический университет им. Н. Э. Баумана, Россия, 105005, Москва, 2-я Бауманская ул., д. 5; e-mail: prof-orlov@mail.ru

ARTIFICIAL INTELLIGENCE, NEURAL NETWORKS, BIG DATA, AND MATHEMATICAL RESEARCH METHODS

© Alexander I. Orlov

Bauman Moscow State Technical University, 5, 2-ya Baumanskaya ul., Moscow, 105005, Russia; e-mail: prof-orlov@mail.ru

В настоящее время часто используются такие термины, как искусственный интеллект, большие данные, нейросети, цифровая экономика. Полезно обсудить, что стоит за этими терминами, как они соответствуют тем или иным разделам математических методов исследования. Основной вывод состоит в том, что за этими распространенными терминами стоят давно разрабатываемые области, разница лишь в названиях. Согласно формулировке [1], «слова и их сочетания (термины) используются различные, а смысл, который стремились передать их интерпретаторы, сохраняется прежним». Безобидна ли игра с терминами? Академик РАН Д. А. Новиков обоснованно считает, что «вокруг искусственного интеллекта складывается очень тревожная структура знаний и компетенций» [2].

В научной области «математические методы исследования» происходят революционные преобразования. Они видны ведущим исследователям, в частности, авторам раздела «Математические методы исследования» нашего журнала, но пока мало заметны для тех ученых, кто применяет эти методы для решения актуальных задач конкретных прикладных областей.

Для анализа современной ситуации оказались весьма полезными основополагающие научные результаты, полученные основателем (вместе с Б. В. Гнеденко) и многолетним руководителем раздела «Математические методы исследования» журнала «Заводская лаборатория. Диагностика материалов» профессором Московского государственного университета Василием Васильевичем Налимовым. Он выпустил (совместно с З. М. Мульченко) первую в мире монографию по наукометрии [3], посвященную анализу развития науки как информационного процесса. Согласно его подходу вклад в науку исследовате-

лей и групп ученых следует оценивать по числу цитирований их работ в дальнейших научных публикациях. Кратко говоря, подход В. В. Налимова основан на простом соображении: раз работу цитируют, значит, она нужна, чем больше цитируют, тем больше вклад в развитие науки. Этот вклад можно количественно оценить с помощью Российского индекса научного цитирования (РИНЦ).

С удовлетворением можно констатировать, что работы авторов раздела «Математические методы исследования» вносят заметный вклад в развитие науки. Так, по данным РИНЦ автор настоящей работы — один из наиболее цитируемых отечественных математиков и экономистов. Этот факт дает основания полагать, что наши обсуждения ряда вопросов развития науки имеют надежные основания и будут интересны читателям.

В информационном процессе развития науки выявлена определяющая роль информационного барьера [4]. Речь идет о том, что интеллектуальные возможности конкретного исследователя ограничены тем, что воспринять он может лишь малую часть накопленного в науке. Покажем на упрощенном примере. Примем, что за неделю удастся изучить одну серьезную публикацию — книгу или статью. Тогда за год — 52 публикации, а за 100 лет — 5200. Сравним с общим числом публикаций, скажем, по математическим методам исследования. По нашей оценке, их более миллиона и каждый год добавляются десятки тысяч. Следовательно, каждый научный работник может непосредственно познакомиться лишь с 0,5 % публикаций по своей специальности. Представление обо всем массиве публикаций складывается на основе мнения окружающей научной общественности.

Информационный барьер мешает распространению новых идей. Поэтому нет ничего удивительного в том, что, например, распространенные в настоящее время учебники по теории вероятностей и математической статистике по своему научному уровню соответствуют середине XX в.

Для преодоления информационного барьера довольно часто используют введение новой терминологии. Новый термин привлекает внимание и дает преимущества при финансировании и организации новых научных структур [4]. Например, в конце XX в. заметного эффекта добились те, кто стал использовать новый термин «интеллектуальный анализ данных» вместо прежнего «прикладная статистика».

Разберем ряд новых терминов и выявим стоящие за ними давно известные области знаний.

Термин «искусственный интеллект» понимают по-разному. Будем исходить из определения, данного в «Национальной стратегии развития искусственного интеллекта на период до 2030 года»: «... искусственный интеллект — комплекс технологических решений, позволяющий имитировать когнитивные функции человека (включая самообучение и поиск решений без заранее заданного алгоритма) и получать при выполнении конкретных задач результаты, сопоставимые, как минимум, с результатами интеллектуальной деятельности человека. Комплекс технологических решений включает в себя информационно-коммуникационную инфраструктуру, программное обеспечение (в том числе в котором используются методы машинного обучения), процессы и сервисы по обработке данных и поиску решений» [5]. В этом определении не говорится прямо про научную основу «комплекса технологических решений». По нашему мнению, в качестве важной составляющей такой основы можно использовать математические методы исследования, включая высокие статистические технологии, в том числе нечисловую статистику, теорию и практику экспертных оценок, статистические методы анализа данных. Автор данной работы занимается проблемами искусственного интеллекта около полвека (первые статьи напечатаны в 1972 г.), хотя в XX в. и не использовал термин «искусственный интеллект». Ссылаясь на книгу 1974 г. сотрудников Института проблем управления В. Н. Вапника и А. Я. Червоненкиса «Теория распознавания образов», Д. А. Новиков констатирует, что «Институт проблем управления РАН — родина искусственного интеллекта, как минимум в России» [2], хотя в работах В. Н. Вапника и А. Я. Червоненкиса искусственный интеллект не упоминался.

При распространении новой терминологии информационный барьер отсекает от исследователей массив научных результатов, сформулиро-

ванных в прежних терминах [4]. Причина понятна — интеллектуальные возможности научных работников нацеливаются на изучение и развитие работ с использованием новомодной терминологии.

В целях преодоления информационного барьера автор настоящей статьи считал возможным изменить при переиздании название серии своих монографий-учебников «Нечисловая статистика» (2009), «Экспертные оценки» (2011) и «Статистические методы анализа данных» (2012). Первоначально серия книг называлась «Организационно-экономическое моделирование», при переиздании серии в 2022 г. название было изменено на «Искусственный интеллект».

Термин «нейросетевые методы» мы обсуждали в [6]: «Сейчас этот термин весьма популярен. Как мы установили в [4], под нейросетевыми методами понимают разнообразные математические методы (и созданные на их основе алгоритмы и программные продукты), построенные по аналогии с представлениями о работе сетей нервных клеток живых существ. Как известно, прототипы таких методов строили в середине прошлого века, опираясь на свойства процессов мышления, протекающих в мозге человека. Основоположники нейросетевых методов попытались смоделировать эти процессы. Естественно, что они исходили из знаний того времени. Отметим, что в настоящее время хорошо известно, что человеческий мозг работает иначе, чем предполагают энтузиасты нейросетей».

Для решения каких задач применяют нейросетевые методы? Их обычно используют для построения правил классификации (другими словами, для диагностики, дискриминации, распознавания образов). Исходные данные — обучающие выборки. Теория математической статистики позволяет сравнивать алгоритмы классификации по качеству. Для задач диагностики целесообразно проводить сравнение на основе прогностической силы алгоритма (см., например, [7]). Оказывается, что нейросетевые методы во многих случаях не дают оптимальных решений. Так, обсудим базовую задачу диагностики. Необходимо принять решение о том, к какому из двух классов причислить появившийся объект. Для каждого такого класса имеется обучающая выборка. Установлено (см., например, [7]), что следует принимать решение исходя из непараметрических оценок плотностей вероятностей, рассчитанных для классов с помощью анализа обучающих выборок. Этот метод является оптимальным при больших объемах обучающих выборок. Ни один нейросетевой метод не может дать лучшего результата.

Приходится констатировать, что мода на нейросети приводит к тому, что оптимальные методы и алгоритмы не используются просто потому,

что забыты. Их место среди интеллектуальных инструментов специалистов заняли нейросети. Одна из причин такой забывчивости — ограниченность возможностей мозга человека воспринимать, осмысливать и применять информацию. Очевидно, что указанное обстоятельство снижает эффективность технологических решений, разрабатываемых в области искусственного интеллекта.

На основе анализа конкретных методов заключаем [4], что нейросети, методы распознавания образов, и, например, генетические алгоритмы — другие названия ряда разделов прикладной статистики (статистических методов анализа данных), которые разрабатываются с давних времен, задолго до нынешней моды на указанные термины. Усилиями журналистов и публицистов, не очень разбирающихся в идеях и научных результатах прикладной статистики, новая терминология оказалась в центре внимания научной общественности. Это произошло по внеученым причинам, которые выявлены в заключительной части статьи [4].

Еще один популярный термин — большие данные. Современные информационные технологии позволяют анализировать все (в идеале) данные по определенному вопросу, имеющиеся в Интернете. Хорошим примером является РИНЦ, в электронной библиотеке которого собраны данные о миллионах научных публикаций (и часто тексты этих публикаций). Отметим, что всеобщий охват данных может позволить получить обоснованные выводы только тогда, когда данные не имеют систематических ошибок, как это часто бывает для социально-экономических данных.

Для работ по аналитической химии, физическим методам исследования и контроля, механике материалов характерно использование результатов измерений (наблюдений, анализов, опытов, испытаний, обследований), полученных в исследовательской лаборатории. Для изучения таких данных и предназначена прикладная статистика и другие математические методы исследования. При этом термин «большие данные» раскрывается с другой стороны. Полвека назад, при отсутствии мощных вычислительных средств, используемые методы основывались прежде всего на предельных теоремах теории вероятностей и математической статистики. Использование больших данных интерпретировалось при этом как переход к пределу при безграничном возрастании объемов выборок. Современные компьютеры с помощью соответствующих программных продуктов позволяют получать выводы для конкретных выборок, используя, например, метод статистических испытаний. Однако во многих случаях такие возможности будут

доступны только в перспективе. В настоящее время использование предельных теорем в прикладной статистике по-прежнему актуально.

Подводя итоги, констатируем, что появление и распространение новых терминов зачастую наносит вред развитию науки как информационного процесса [3]. Как исследователю находить необходимую информацию, в том числе в ранее выпущенных литературных источниках? Может помочь РИНЦ. Естественно проводить поиск публикаций и авторов по числу цитирований. Как правило, больше цитируют добротные источники [8].

ЛИТЕРАТУРА

1. Игры терминов и ловушки Фукидида (от редакции) / Онтология проектирования. 2022. № 3(45). С. 273 – 277. <https://cyberleninka.ru/article/n/igry-terminov-i-lovushki-fukidida> (дата обращения: 09.02.2023).
2. Новиков Д. А. Вокруг искусственного интеллекта складывается очень тревожная структура знаний и компетенций. <https://new.ras.ru/mir-nauky/news/vokrug-iskusstvennogo-intellekta-skladyvaetsya-ochen-trevozhnaya-struktura-znaniy-i-kompetentsiy-aka/?ysclid=ldx9taju6720668306> (дата обращения: 09.02.2023).
3. Налимов В. В., Мульченко З. М. Наукометрия. Изучение развития науки как информационного процесса. — М.: Наука, 1969. — 192 с.
4. Орлов А. И. Смена терминологии в развитии науки / Научный журнал КубГАУ. 2022. № 177. С. 232 – 246.
5. Указ Президента РФ от 10 октября 2019 г. № 490 «О развитии искусственного интеллекта в Российской Федерации».
6. Орлов А. И. О современных интеллектуальных инструментах экономики и менеджмента / Экономическая наука современной России. 2022. № 4(99). С. 30 – 38.
7. Орлов А. И. Искусственный интеллект: нечисловая статистика. — М.: Ай Пи Ар Медиа, 2022. — 446 с.
8. Лойко В. И., Луценко Е. В., Орлов А. И. Современные подходы в наукометрии. — Краснодар: КубГАУ, 2017. — 532 с.

REFERENCES

1. Games of terms and traps of Thucydides (editor's note) / Ontol. Proekt. 2022. N 3(45). P. 273 – 277. <https://cyberleninka.ru/article/n/igry-terminov-i-lovushki-fukidida> (accessed 09.02.2023) [in Russian].
2. Novikov D. A. A very disturbing structure of knowledge and competencies is emerging around artificial intelligence. <https://new.ras.ru/mir-nauky/news/vokrug-iskusstvennogo-intellekta-skladyvaetsya-ochen-trevozhnaya-struktura-znaniy-i-kompetentsiy-aka/?ysclid=ldx9taju6720668306> (accessed 09.02.2023) [in Russian].
3. Nalimov V. V., Mul'chenko Z. M. Scientometrics. Studying the development of science as an information process. — Moscow: Nauka, 1969. — 192 p. [in Russian].
4. Orlov A. I. Change of terminology in the development of science / Nauch. Zh. KubGAU. 2022. N 177. P. 232 – 246 [in Russian].
5. Decree of the President of the Russian Federation from 10 October 2019 N 490 “On the development of artificial intelligence in the Russian Federation” [in Russian].
6. Orlov A. I. About modern intellectual tools of economics and management / Ékon. Nauka Sovr. Rossii. 2022. N 4(99). P. 30 – 38 [in Russian].
7. Orlov A. I. Artificial intelligence: non-numeric statistics. — Moscow: Ai Pi Ar Media, 2022. — 446 p. [in Russian].
8. Loiko V. I., Lutsenko E. V., Orlov A. I. Modern approaches in scientometrics. — Krasnodar: KubGAU, 2017. — 532 p. [in Russian].