

1. НАУЧНЫЕ СТАТЬИ

УДК: 004.82, 621.13

1.1. Экономика данных, обучение нейросетей и многомерная геометрия

Козырев А. Н., ЦЭМИ РАН, г. Москва, Россия

Главное в науке – уметь посмотреть с одной точки зрения на разные предметы и на один предмет – с разных точек зрения».

*Леонид Витальевич Канторович
своим ученикам*

В статье на конкретных примерах показано, как математика высокого класса иногда помогает решать проблемы в области экономики и информационных технологий, для решения которых она изначально не предназначалась. Особо подчеркивается роль выдающихся отечественных математиков относительно недалекого прошлого.

1. Введение

Эта статья была задумана и написана как продолжение темы цифровизации, заявленной ранее в публикациях [Козырев, 2017, 2024], где много внимания уделено мемам «цифровая экономика» и «экономика данных», истории их появления, а также людям, событиям и коммерческим интересам, стоящим за этими двумя мемами. Такой подход представляется оправданным по ряду причин, одна из которых – очевидное наличие общественного запроса на профессиональное, основанное на фактах, а не мифах и мнениях, разъяснение смысла этих двух словосочетаний, ставших мемами, но не терминами. Разумеется, наивно было бы думать, что написание и публикация статьи в профильном журнале удовлетворит этот запрос. По законам жанра неизбежно образуется «культурный слой» из статей с пересказом авторами мнений друг друга и постепенной утратой какого-либо содержания. Тем не менее, профессиональный ответ на этот запрос должен быть дан, ровно для этого создавался наш журнал, выпускаемый группой энтузиастов. В этом наш профессиональный долг, его надо выполнять и перейти, развивая тему, к более деликатным и сложным вопросам, связанным с развитием технологий и математикой.

Главная цель этой статьи – показать опасность недооценки математики и, прежде всего, чистой математики в работе с данными и абсурдность противопоставления этих двух сфер деятельности. К сожалению, недооценка чистой математики и даже её противопоставление работе с большими данными – явление довольно обычное, свойственное не только «эффективным менеджерам», но и многим молодым людям, связывающим свое будущее с анализом больших данных, искусственным интеллектом и опосредованными компьютером транзакциями. В частности, это можно видеть на примере студентов МФТИ, где отношение студентов к математике становилось все более безразличным, как минимум, на протяжении последних пятнадцати лет. И это при том, что выпускники МФТИ весьма успешны в области применения информационных технологий (ИТ). Разумеется, столь категоричное мнение не может не быть субъективным, но именно потому, что оно основано на личных наблюдениях (своих и близких по духу коллег), то есть на своего рода первичных данных, не улучшенных в угоду хорошей отчетности. Кроме того, нам есть с чем сравнивать, мы тоже когда-то учились, программировали, потом стали учить.

Успехи в области применения ИТ не так значительны, как это принято представлять, хотя прогресс неоспорим. Вычислительные ресурсы тратятся крайне неэффективно, количество совершаемых ошибок растет, а коммерческие интересы фирм далеко не всегда совпадают с интересами пользователей. Особенно ярко это проявляется в области искусственного интеллекта (ИИ), где сам термин слегка отдает рекламой, но не наукой. Если же говорить о коммерческих интересах, то наиболее ярко они представлены в публикациях авторов, аффилированных с парой Google и Alphabet. В частности, они более открыты, чем представители других ИТ-фирм, регулярно пишут о своих достижениях и целях, как правило, задним числом. Но бывают и исключения. Так, совсем недавно в открытом доступе опубликован препринт arXiv:2401.09718, где группа лиц, аффилированных с Google и Alphabet, излагает свои взгляды на прошлое, настоящее и будущее ИИ, его колоссальное положительное влияние на экономику и науку, а также некоторые опасения по поводу возможных злоупотреблений и последствий [Ben-Ishai et al, 2024].

Несколько раньше, команда специалистов, объединенных псевдонимом Google Brain, опубликовала статью [Gilmer et al. 2018], где используется термин «многомерная геометрия», а используется он в контексте обсуждения не решенных до настоящего времени проблем безопасности при обучении нейросетей. Разумеется, 2018 год – это по сегодняшним меркам далекое прошлое, технологии обучения нейросетей за прошедшее время ушли далеко вперед. Но, судя по свежим публикациям, например

[Tanner et al, 2024], проблема не решена до сих пор и, можно быть уверенным, не будет решена, пока ее пытаются решать привычными для современной ИТ-индустрии методами, то есть без обращения к более совершенной математике или техническим средствам. Но той математике, которая может стать новым инструментом, не учат будущих специалистов по ИТ. Беда в том, что заранее трудно сказать, что именно из математики здесь сработает, но это не может быть что-то совсем банальное. Если же говорить о технических средствах, то невольно вспоминаешь об аппаратной защите данных, которая появилась в СССР во времена еще ламповых ЭВМ, но до сих пор используется в некоторых типах процессоров. Разумеется, отсюда не следует, что для решения проблемы, связанной с «отравлением данных» и появлением состязательных примеров, о которых говорится в цитируемых публикациях, необходимо применение аппаратного контроля. Но решение таких проблем может находиться далеко за пределами традиционных для ИТ-индустрии инструментов, в том числе это могут быть неожиданные на первый взгляд математические результаты. Интересный пример – публикация [Liu et al, 2024], где рассматриваются возможности нейросетей на основе результатов, полученных двумя выдающимися советскими математиками А.Н. Колмогоровым и В.И. Арнольдом [Колмогоров, 1956, 1957; Арнольд, 1957, 1959] в ходе дискуссии о решении тринадцатой проблемы Гильберта [Hilbert, 1935]. Более старый, но не менее интересный сюжет – применение тропической геометрии для обучения нейросетей. Он интересен и тем, что тропическая математика – это очень молодое и быстро развивающееся направление, и тем, что она создает своего рода мост между дискретной математикой и геометрией выпуклых многогранников. Основа этого моста – идея двойственности – одна из самых глубоких идей, пронизывающая многие математические дисциплины, а с некоторых пор и экономическую теорию (в её части «для умных»).

Далее изложение строится следующим образом. Сразу за настоящим введением следует раздел о междисциплинарном, или (точнее) мультидисциплинарном подходе и о необходимости понимать друг друга, как минимум, на уровне терминов. Следом за ним идут разделы о состязательных примерах, о тропической геометрии, о сетях Колмогорова-Арнольда и, наконец, о том, что могут полезного сделать экономисты. Этому посвящен эпилог, поскольку называть заключение этот раздел жалко.

2. О междисциплинарном подходе и многозначности слова «данные»

Как уже отмечалось в [Козырев, 2024], тема цифровизации, данных и ИИ касается слишком большого круга вопросов, в решении которых задействованы представители разных специальностей и разных научных дисциплин, использующие одинаковые по написанию и звучанию термины, означающие подчас очень разные предметы и явления. Вторжение в чужую область со своим пониманием терминов или пониманием терминов на бытовом уровне приводит к нелепым ситуациям. Здесь же мы собираемся говорить об экономической реальности, но о такой её части, где одинаковое понимание слов – необходимое, хотя и не всегда достаточное. В этом смысле очень примечателен популярный лозунг «Данные – это новая нефть». Сходство с нефтью здесь ровно в том, что необработанные данные также мало

пригодны к употреблению, как сырая нефть. В остальном данные радикально отличаются от нефти, что особенно ярко видно на их сложении. Оно идемпотентно, а вычитание для них не определено. На экономическом языке это называется неконкурентностью в потреблении. Именно по этой причине возможно создание открытых баз данных для обучения ИИ. Но этим различия не исчерпываются, а мантра про «новую нефть» только уводит от сути дела.

На бытовом уровне, в экономике и информатике слово «данные» понимается очень по-разному в зависимости от того, о чем идет речь. Но, коли речь заходит об экономике данных, имеет смысл ограничить применение этого термина и провести границы между данными, информацией и знаниями. Для этого в информатике применяется концепция «пирамиды данных». На рисунке 1 представлен вид пирамиды данных в том виде, как это понимают и декларируют специалисты из

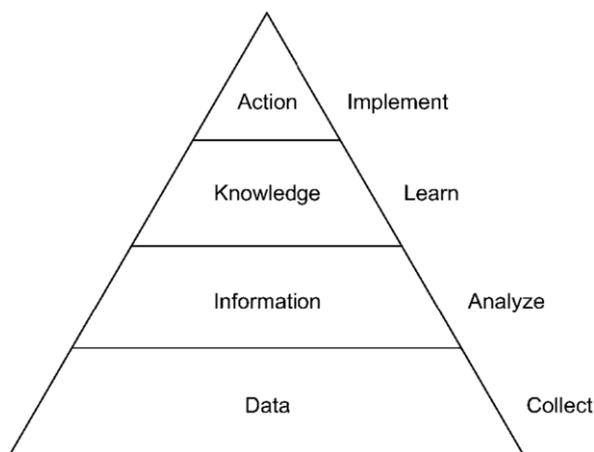


Рис. 1. Пирамида информации.
Источник [Varian, 2019]

Google [Varian, 2019]. Данные в этой пирамиде составляют нижний слой, это те первичные данные в нечитаемом человеком виде, которые собирает какое-то устройство. Например, это может быть видеорегистратор, сканер или что-то еще. Вариантов достаточно много даже в случае, когда речь идет только об изображениях. В целом их много больше, но общий признак – предназначенность для чтения и обработки машиной, а не человеком. Следующий слой – информация, предназначенная для восприятия человеком – получается из первичных данных путем их обработки. Слои, обозначенный термином «знания» (Knowledge), состоит из усвоенной человеком информации, на основе которой он принимает решения. Возможно, такая интерпретация покажется слишком узкой, но речь идет о технике, а не о философии или маркетинге. Есть нюансы и с принятием решений.

В классическом варианте схемы работы с данными какая-то система должна собирать необработанные данные, а затем систематизировать и анализировать их, чтобы превратить в информацию - например, в изображение текстового документа, которое может быть понято человеком. Пиксели на изображении превращаются в понятные человеку надписи. Это понимание, полученное из информации, затем может быть преобразовано в знания, которыми обладают люди. Мы можем представить себе данные, хранящиеся в битах, информацию, хранящуюся в документах, и знания, хранящиеся в головах людей. Существуют хорошо развитые рынки и нормативно-правовая база для информации в таком понимании (книг, статей, веб-страниц, музыки, видео) и знаний (консультационные услуги, обучение). Все это или почти все может быть алгоритмизировано и передано машине, как и делается в традиционном машинном обучении (ML). Проблемы здесь тоже есть, но не о них речь.

Сегодняшние достижения ИИ и соразмерные им проблемы связаны с глубинным обучением, они начинаются там, где машина совершает весь путь от сбора первичных данных к действию. Классический пример – управление ИИ автомобилем без какого-либо участия человека. В этом случае нет нужды в получении информации, воспринимаемой человеком, как и в последующем превращении ее в знания, за которыми последует то или иное действие. Такое спрямление пути дает огромное преимущество по сравнению с традиционным пошаговым движением от данных к информации, знаниям и от них к действию. Но этот путь сопряжен с опасностями, связанными с «отравлением данных», составительскими примерами и другими проблемами, которыми занимаются специалисты по информационной безопасности с переменным успехом. Одна из проблем состоит в том, что средства нападения здесь обходятся на порядки дешевле, чем средства обеспечения информационной безопасности. А это – экономика, и решать ее должны экономисты, но не любые, а понимающие техническую сторону вопроса.

Идеальный пример современного экономиста, владеющего смежными специальностями, это Хэл Вэриан. Стоит напомнить, что по магистерской специальности Вэриан математик, по диссертации PhD – специалист по информационным технологиям, а известен он в основном как экономист и автор знаменитого учебника по экономикс промежуточного уровня, то есть с математикой, но без ухода за пределы понимания для студентов экономистов. А потому его мнение по данному вопросу особенно интересно.

Искусственный интеллект и большие данные у всех на слуху благодаря постоянному повторению этих словосочетаний в средствах массовой информации, как правило, без понимания стоящей за ними реальности, включая достигнутый уровень технологий, возможности сбоев и злоупотреблений. Об этих достижениях и проблемах пишут в специальной литературе. О них можно почитать дискуссии на профессиональных сетевых ресурсах. А совсем недавно группа лиц, аффилированных с Google и Alphabet, опубликовала своего рода манифест о влиянии ИИ на разные сферы жизни [Ben-Ishai et al, 2024] с интригующим названием «Искусственный интеллект и возможности для общего процветания: Уроки из истории технологий и экономики». Среди шести авторов этого манифеста чьи вклады в его содержание декларированы как равные, есть и Хэл Вэриан. Тем не менее, весь этот манифест написан в позитивном ключе и без использования формул, вероятно для того, чтобы его прочло максимальное число людей, интересующихся экономикой и влиянием на нее ИИ. Однако с его идеями можно поспорить.

Основная идея манифеста, если отбросить идеологические заявления, состоит в том, что ИИ – технология общего назначения (general-purpose technology), причем в том ряду, куда входят еще паровой двигатель, электричество и персональный компьютер [Bresnahan & Trajtenberg, 1995], ИИ – не один из,



Рисунок 2. Слайд из презентации автора 14.07.2018

а нечто большее по масштабу и глубине. Здесь видится целый ряд подмен. Самая очевидная из них – использование термина «искусственный интеллект», не являющегося термином в общепринятом смысле этого слова. Его наполнение смыслами постоянно меняется. Реально речь идет о новых нейросетевых технологиях, получивших в последнее время действительно широкое распространение, что и позволяет говорить о них в целом как о технологии общего назначения. Но под шапкой ИИ помимо них могут скрываться и какие-то другие технические решения, как, собственно говоря, и было на протяжении примерно семидесяти лет. Популярность ИИ и объем финансирования изменяется от одной «зимы ИИ» к другой с периодами относительного благополучия между «зимами». Каждый раз «зиме» предшествует ажиотаж, связанный с каким-то успехом и очень громкими обещаниями. Потом выясняется, что обещания были сильно завышены или несколько преждевременны, начинается новая «зима» с сокращением финансирования и внимания прессы. Так, в знаменитом матче между AlphaZero и Stockfish в 2017 году программа Stockfish играла с отключенными базами по дебютам и эндшпилям, чтобы «выровнять условия». Но при этом AlphaZero использовала тензорный процессор, а Stockfish – обычный, но ничего тут выравнивать не стали, поскольку «технически невозможно». К

счастью для Google эти детали не получили тогда широкой огласки, хотя внутри шахматного сообщества они обсуждались. Нечто подобное было и в знаменитом матче, когда в 1997 году Deep Blue, созданная командой IBM, обыграла тогдашнего чемпиона мира по шахматам Гарри Каспарова. Ему не давали пользоваться электронным дебютным справочником, то есть он играл так, как принято в матчах между людьми, но его «противник» пользовался таким справочником.

Разумеется, через относительно небольшой период времени машины стали играть так, что человеку в принципе невозможно с ними соперничать, это связано с тем, что шахматы – матричная игра. Каждая стратегия любой из сторон – набор ответов (ходов) для каждой позиции с ходом этой стороны. Их совокупность – полный набор стратегий для каждой из сторон – конечное, но очень большое множество. Однако нет никакой необходимости рассматривать все стратегии. Надо рассматривать их классы. Например, если можно поставить мат в один ход, то нет надобности рассматривать другие ходы. При этом совершенно неважно, как стоят остальные фигуры. Мы имеем дело с одним классом. То же самое с некоторыми нюансами имеет место практически на каждой стадии игры. Множество стратегий разбивается на классы эквивалентности, которых намного меньше, чем стратегий в исходном смысле. И тут оказывается, что самообучение машины с какого-то момента оказывается более быстрым. На момент матча между AlphaZero и Stockfish человек мог обучить программу идеально играть в эндшпилье и очень хорошо в дебюте. Именно эти возможности и отключили Stockfish. В другие подробности вдаваться не будем, речь не об этом, а о том, что погоня за сенсациями и увеличением финансирования постоянно толкает на мелкий или не очень мелкий подлог. Иногда это сходит с рук, иногда нет.

Особенность сегодняшней ситуации в том, что очередная «зима» не наступила в положенный ей срок, а наоборот, возник ажиотаж, связанный с успехами генеративных сетей. Но будет ли от них больше пользы, чем вреда – большой вопрос. Беда в том, что ответ мы узнаем, когда будет уже поздно.

Опосредованные компьютером транзакции, сбор данных и составительские примеры

Об опосредованных компьютером транзакциях говорят и пишут много меньше, чем об ИИ и о данных как «новой нефти», хотя для понимания современной экономики важен и сам этот термин, и стоящая за ним реальность. О них многое было сказано в работах Хэлла Вэриана [Varian, 2010, 2013], но там речь шла о том, что происходит «за пределами Больших Данных» (Beyond Big Data), как и называлась одна из его статей, то есть об информации, полученной в виде следов от разного рода транзакций и представленной в форме, пригодной для восприятия человеком. В частности, речь шла о фотографиях с ценниками на свинину в разных магазинах Шанхая, об информации с датчиков скорости на автомобилях, сдаваемых в аренду или на условиях каршеринга, а также о том, как эту информацию можно использовать для создания новых форм договоров и форм бизнеса. Его правоту в этих вопросах подтвердил стремительный взлет Google после того, как в 2002 году Вэриан пришел туда работать сначала консультантом, а с 2007 – главным экономистом. Но сейчас речь не об этом, а о первичных данных, используемых для обучения нейросетей, минуя превращение их информации, понимаемую людьми.

Очень часто, открывая какой-то ресурс, приходится подтверждать, что Вы не робот. Иногда для этого нужно просто поставить галочку в подходящем поле. Но достаточно часто при попытке скачать текст научной статьи или книги, находящейся в открытом доступе, вдруг возникает необходимость пометить все фотографии, на которых есть светофор или дорожный знак. В этом случае можно с уверенностью сказать, что Вас используют для создания обучающей подборки меченых данных. В случае со светофорами или дорожными знаками это может быть подборка для обучения ИИ управлению автомобилем, хотя и не обязательно. Альтернативные варианты получения такой подборки – это, например, нанять некоторое количество туповатых, но вменяемых фрилансеров, которые бы на платной основе метили фото с теми самыми объектами. В какие суммы это обойдется – вопрос к экономистам. Эксплуатировать тех, кто ищет материалы для статей или совершает покупки в сети, точно дешевле. Еще один вариант – обратиться к открытым базам. Тут снова можно сослаться на цитируемые выше статьи коллег из Google, они указывают множество таких источников. У тех, кто выкладывает огромные массивы меченных данных для обучения чужих сетей, вероятно, есть понимание мотивов своих действий, но они не обязательно те же, что у поэта, написавшего «Граждане, послушайте меня».

Со сбором меченых данных, в том числе из открытых источников, связаны специфические риски, справиться с которыми помогает правильное понимание многомерной геометрии [Dube, 2018, 2020].

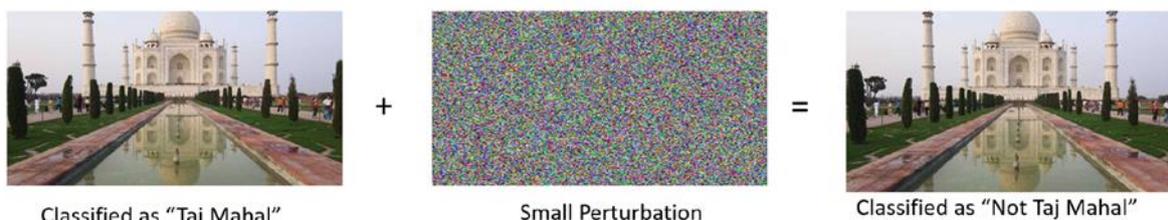


Рисунок 3. Небольшое возмущение приводит к тому, что сеть изменяет результат. Источник [Dube, 2018].

На рисунке 3 приведен пример из работы [Dube, 2018], показывающий резкое изменение ответа сети, настроенной на распознавание знаменитого на весь мир объекта. Там же приводится пример

фейковых неузнаваемых изображений, которые обученная сеть распознает как Тадж Махал. Такие примеры показывают уязвимость сети и возможность опасных ошибок. Например, это может привести к аварии, если сеть неправильно распознает дорожный знак. Примечательно то, что смысл знака может быть совершенно ясен человеку, выделяющему информацию в указанном выше смысле из массива необработанных пикселей, а потом действующего на основании знаний правил дорожного движения и увиденной информации. А машина с обученной сетью перейдет от массива считанных пикселей к управлению более коротким путем, зато она может увидеть то, чего нет, а есть только сделанный кем-то фейк.

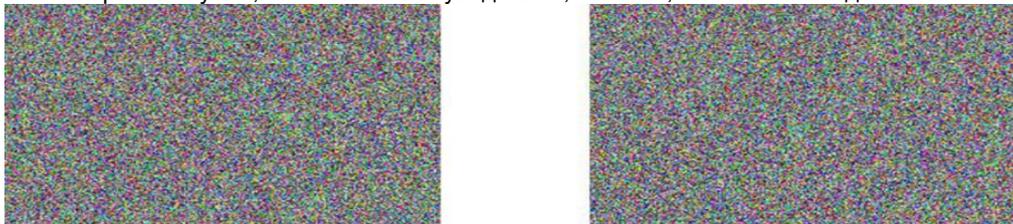


Рисунок 4. Примеры фейковых изображений, распознаваемых сетью как Тадж Махал. Источник [Dube, 2018].

На рисунке 4 как раз изображены два таких фейка. Они безобидны. Но при наличии недружественных намерений можно сделать фейк, по которому будет нанесен удар самой предполагаемой жертвой по себе самой. На эту тему можно фантазировать, но пафос цитируемой публикации не в этом.

В статье [Dube, 2018] глубокое обучение анализируется с математической точки зрения, исследована работа глубоких нейронных сетей и первопричины возникновения противоречивых примеров. Затем обсуждаются возможности улучшить их в будущем и устранить противоречивые примеры. В статье не только получены новые результаты, но и указаны ошибки аргументации в ранее опубликованных статьях по теме. А примечательно то, что все это делается с применением математических свойств многомерных пространств, использована геометрическая и топологическая аргументация. Далее рассматриваются оптимизационные ландшафты глубоких нейронных сетей и количество седловых точек относительно числа локальных минимумов. Там же показано, как природа изображений с множественным разрешением объясняет противоречивые примеры, основанные на возмущениях. Все это преподносится в виде математических формул и утверждений. Наконец, используя гипотезу множественного обучения по частям для естественных изображений, автор исследует работу глубоких нейронных сетей и первопричины возникновения противоречивых примеров, а также обсуждает, как можно улучшить сети в будущем и как можно устранить противоречивые примеры. Были ли эти идеи воплощены в жизнь за последующие годы, не до конца ясно. Но подход Дюбе интересен, в нем чувствуется настоящая математика.

Тропическая геометрия двойственности

Тропическая геометрия получила свое название в честь бразильского математика венгерского происхождения Имре Симона в связи с публикацией [Simon, 1988], оно прижилось как интригующее и в целом более удачное, чем исходное название соответствующей ветви математики – идемпотентного анализа [Маслов, Колокольцев, 1994]. Как и другие ветви математики, идемпотентная математика успешно развивалась в СССР [Воробьев, 1963] и позже в постсоветской России [Кривулин, 2009], хотя первородство при желании можно оспорить. Главное отличие идемпотентной математики от обычной состоит в том, что обычное сложение заменяется идемпотентной бинарной операцией. Чаще всего это операция максимума или минимума. Если говорить конкретно о тропической геометрии, то сложение здесь заменяется операцией минимума, умножение – операцией обычного сложения, а в качестве нуля используется бесконечность, обычно обозначаемая символом ∞ . Получаемая в результате таких замен алгебраическая система представляет из себя полуполе, поскольку операция сложения в ней необратима. В остальном она достаточно похожа на обычную арифметику. Чтобы не возникало путаницы, имеет смысл использовать символ \oplus для идемпотентного сложения и символ \otimes для умножения. Символ ∞ спутать с чем-то еще достаточно сложно. Получаемая алгебраическая система с коммутативными сложением и умножением, но без деления является полуполем. Над этим полуполем строится полноценная (за небольшими исключениями) математика. Там есть свой аналог линейной алгебры и своя геометрия.

Решениями уравнений в этой математике считаются значения переменных, при которых совпадают значения каких-то одночленов. Так, решениями уравнения $ax \oplus by = c$ считаются значения переменных x и y , при которых выполняется хотя бы одно из равенств $ax = bx$, $ax = c$ или $bx = c$. Соответственно, прямая линия в такой геометрии состоит из трех лучей, соединяющихся в точке $ax = by = c$ и разделяющих плоскость на три части. Как это выглядит в конкретном случае, показано на рисунке 5. Уравнение $x^2 \oplus y^2 = 1$ имеет решения $x^2 = 1$, $y^2 = 1$ и $x^2 = y^2$. Его тоже любопытно нарисовать.

С увеличением числа переменных растет размерность пространства. Его уже трудно нарисовать. Но даже простейший случай с разделением плоскости на три части имеет нетривиальные практические приложения. Например, в 1998 году банк Англии раздавал кредиты на 100 миллиардов фунтов стерлингов (в сумме) по низким ставкам с надежным обеспечением, по более высоким ставкам с менее надежным обеспечением, причем без ограничения числа заявок от каждого просителя. Все 100 миллиардов фунтов стерлингов были розданы, причем все получили кредит на условиях более благоприятных, чем они готовы

были взяты. Кто-то кредит не получил. Это были те, кто просил кредит на лучших условиях, чем «прямая» отсечения. Пикантность ситуации как раз в том, что эта «прямая» делила плоскость на 3, а не на 2 части.

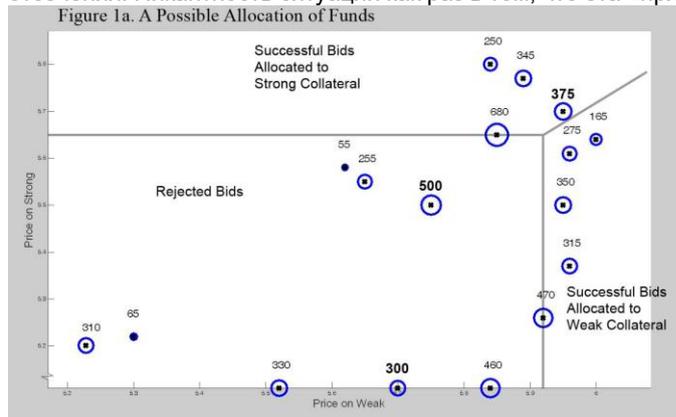


Рисунок 5. Тропическая прямая при раздаче кредитов 1998 г. Источник [Klemperer, 2008]

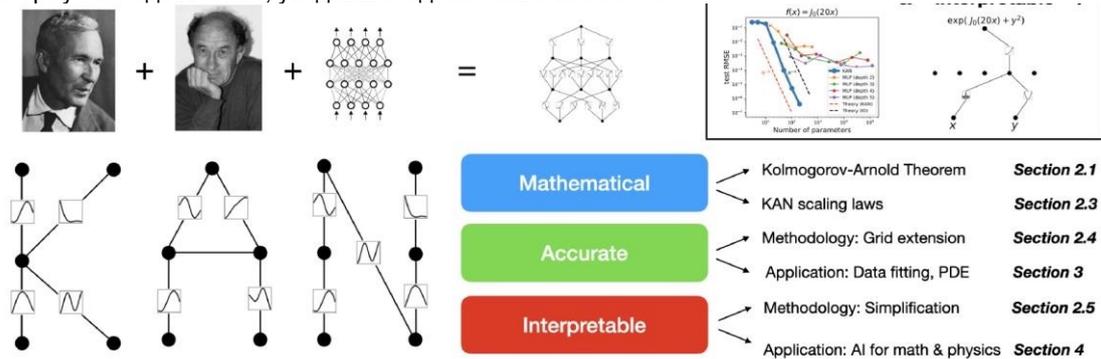
При всем при том применение тропической геометрии в экономике пока делает лишь первые шаги. Гораздо удачнее обстоит дело с её использованием при обучении нейронных сетей. С этим же отчасти связано и само её название в честь бразильского математика, ставшего пионером данного направления. И снова здесь большие данные, но это не те данные, или, точнее, не только те данные, которые получаются из цифровых следов. Очень большое количество данных для обучения нейронных сетей приходится собирать вручную. Это стоит дорого. Но и тут есть надежда на улучшение ситуации. Она тоже связана с чистой (до поры) математикой. Тропическая геометрия несколько превышает тот уровень, который Вэриан счел доступным

для экономистов, когда писал свои учебники по экономикс промежуточного уровня, но в данном случае это не столь уж важно. Достаточно понимать, что она есть и востребована не только в экономике, но и в работе с данными. Она очень удобна при работе с задачами на графах, а высокую эффективность её применения обеспечивает связь с геометрией выпуклых многогранников [Baldwin, Klemperer, 2012]. Эта связь возникает через двойственность и многогранник Ньютона, то есть она обязана своим наличием достаточно давним достижениям математики. Примечательно здесь то, что двойственная задача для исходной целочисленной задачи, сформулированной на языке тропической геометрии, оказывается непрерывной и потому проще решаемой. Тропическая геометрия нашла довольно широкое применение в обучении нейросетей [Bhatia et al., 2024]. Ей учат в ряде российских вузов [Кормаков, 2002], в частности, на ВМК МГУ и в СПбГЭТУ «ЛЭТИ».

Сети Колмогорова-Арнольда (KAN)

Совсем недавно появилось сообщение [Liu et al, 2024] о работах по созданию нового вида нейросетей на основе теоремы о представлении, которая появилась в результате дискуссии между двумя выдающимися советскими математиками А.Н. Колмогоровым и В.И. Арнольдом о 13-й проблеме Гильберта, откуда возникло название такого рода сетей. Стоит отметить, что речь идет об очень глубоком математическом результате, состоящем из двух частей. Сначала Колмогоров [Колмогоров, 1956] показал возможность представления непрерывных функций $n \geq 3$ переменных в виде суперпозиции функций трех переменных. Потом Арнольд свел это к суперпозиции двух переменных [Арнольд, 1957, 1958].

И только совсем недавно этот результат нашел применение в области ИИ, что вызвало большой восторг у исследователей, увидевших здесь новые возможности.



Our proposed Kolmogorov-Arnold networks are in honor of two great late mathematicians, Andrey Kolmogorov and Vladimir Arnold. KANs are mathematically sound, accurate and interpretable.

Рисунок 6. Источник [Liu et al, 2024].

Сети Колмогорова-Арнольда (KAN) предлагаются в цитируемой работе [Liu et al, 2024] как многообещающая альтернатива популярным сегодня многослойным перцептронам (MLP). С точки зрения математика её очевидными преимуществами надо считать относительную компактность, красоту и понятность. В то время как MLP имеют фиксированные функции активации на узлах (нейронах), KAN имеют обучаемые функции активации на ребрах (весах). KAN вообще не имеют линейных весов – каждый

весовой параметр заменяется одномерной функцией, параметризованной как сплайн. В цитируемой работе показано, что это, казалось бы, простое изменение позволяет KAN превосходить MLP с точки зрения точности и интерпретируемости в небольших задачах, связанных с искусственным интеллектом и естественными науками. Что касается точности, то небольшие KAN могут достичь сопоставимой или лучшей точности, чем более крупные MLP, в задачах подбора функций. Теоретически и эмпирически KAN обладают более быстрыми законами нейронного масштабирования, чем MLP. Для удобства интерпретации KAN могут быть интуитивно визуализированы и легко взаимодействовать с пользователями-людьми. С их приходом в область искусственного интеллекта (ИИ) наступает конец бравированию набившим оскомину «мы сами не понимаем, как она (сеть) это делает».

Бравировать непониманием – занятие, строго говоря, не для ученых. Тем не менее, в сфере ИИ это непонимание часто преподносится чуть ли не как свидетельство высокого уровня ИИ (еще не общего ИИ, но почти). Похоже, этой «прекрасной эпохе» приходит конец.

Там же [Liu et al, 2024] на двух примерах из области математики и физики показано, что KAN могут быть полезны ученым, помогая (заново) открывать математические формулы и физические законы. Таким образом, KAN – многообещающая альтернатива MLP, она открывает возможности для дальнейшего совершенствования современных моделей глубокого обучения, основанных вплоть до настоящего времени преимущественно на MLP. Но смысл этой истории глубже, она вызывает ассоциации.

В далеком уже 1972 году к д.ф.-м.н. Г.Ш. Рубинштейну в Институт математики СО АН СССР пришли физики из Института ядерной физики СО АН СССР и попросили помочь в создании математической модели для оптимального размещения нового оборудования в одном из помещений. Бывший аспирант Л.В. Канторовича и автор весьма любопытного приложения к его книге [Канторович, 1960] предложил нарисовать контуры агрегатов на бумаге, вырезать и подвигать по нарисованному же полу помещения, сохранив пропорции. А потом рассказал автору этой статьи (в тот момент студенту-дипломнику) про решение задачи о проводах и изоляции, решаемой с применением бильярдных шаров и ремня (рис 7). С точки зрения математики здесь важно только сечение, пропорции между наибольшим сечением бильярдного шара и сечением провода в идеале такие же, как между сечением ремня и изоляционной трубки. И в этом соль.



Рисунок 7. Одинаковое сечение – ключ к решению

Эпилог

Приведенные выше примеры использования при работе с данными красивых математических идей, относимых до последнего времени к чистой математике, показывают фантастическую продуктивность этой науки. В самом деле, все перечисленные выше достижения в этой области (до их применения) получены людьми очень высокой квалификации, но с зарплатой просто несоизмеримой с зарплатой кинозвезд или топ-менеджеров. При этом не тратилась энергия, как при физических экспериментах, не велись раскопки и так далее. Но на их основе созданы более совершенные математические методы. Более эффективные вычислительные процедуры, используемые во множестве компьютеров много-много раз, приносят колоссальную экономию. Впрочем, чистая математика прекрасна и без этого.

Подводя итог сказанному выше, можно заметить, что экономистам есть чем заняться в экономике данных, если есть желание занять конструктивную позицию. Осмысленные вопросы, как минимум, отчасти относящиеся к компетенции экономистов представлены ниже в разбивке на 4 группы.

1. Как можно использовать во благо «следы» опосредованных компьютерами транзакций?
 - a. Как предложение адаптируется к реальному спросу, а как это может вести к завышению цен и играть против пользователя (покупателя)?
 - b. Так ли страшна ценовая дискриминация (дифференциация цен) или без этого и двух-компонентных цен экономика с растущей отдачей на масштаб не живет?
 - c. Какие формы аукционов становятся возможны, а какие отпадают?
1. Как накопление информации порождает надзорный капитализм?
 - a. Кто владеет и/или распоряжается Большими Данными?
 - b. Уживутся ли приватность, удобство и безопасность вместе?
 - c. «Новая нефть» – это Большие Данные или эксплуатируемое большинство?
2. Где экономически оправдано применение генеративных сетей, если сопоставить затраты на обучение и реальный эффект?
 - a. Какие профессии могут быть реально заменены ИИ?
 - b. Где это экономически нецелесообразно?
 - c. А как это коснется самих экономистов?
3. Каковы экономические перспективы промышленного майнинга?
 - a. Как растут затраты энергии?
 - b. Каково соотношение между пользой (с учетом мультипликативного эффекта) и затрат?
 - c. Чем может быть оправдан практически очевидный отрицательный результат?

4. Зачем люди играют в эту игру под названием «промышленный майнинг»?
 - a. Как растёт объём хранимой информации в системах блокчейн с ростом числа участников?
 - b. Как решается проблема слишком быстрого роста?
 - c. Как при этом происходит отказ от изначально декларируемой концепции и её целей?

На часть из перечисленных выше вопросов уже есть ответы, например в книге [Zuboff, 2019], уже переведённой на русский язык и широко обсуждаемой гуманитариями. Но совсем не очевидно, что эти ответы устроят тех, кто сегодня принимает решения,

Литература

1. Арнольд В. И. (1957) О функциях трех переменных, ДАН СССР, т. 114, № 4 (1957), 679—681.
2. Арнольд В. И. (1959) О представлении непрерывных функций трех переменных суперпозициями непрерывных функций двух переменных, Матем. сб., 1959, том 90, номер 1, 3—74
3. Канторович Л.В. (1960) Экономический расчет наилучшего использования ресурсов, М.: Издательство АН СССР, 1960.
4. Воробьев, Н. Н. (1963) Экстремальная алгебра матриц, Докл. АН СССР, 1963, том 152, « 1, 24—27
5. Козырев А.Н. (2018), Цифровая экономика и цифровизация в исторической перспективе // Цифровая экономика, № 1, 2018, с.5-19, DOI: [10.34706/DE-2018-01-01](https://doi.org/10.34706/DE-2018-01-01)
6. Козырев А.Н. (2024) Цифровая экономика и экономика данных // Цифровая экономика № 2(28), 2024 – с. 5–13. DOI: [10.33276/de-2024-02-01](https://doi.org/10.33276/de-2024-02-01)
7. Колмогоров А.Н. (1965) Три подхода к определению понятия «количество информации» // Проблемы передачи информации. – 1965. – Т.1 – Вып. 1. – с. 3–115.
8. Колмогоров А. Н. (1956), О представлении непрерывных функций нескольких переменных суперпозициями непрерывных функций меньшего числа переменных, ДАН СССР, т. 108, № 2 (1956), 179—182.
9. Колмогоров А.Н. (1957), О представлении непрерывных функций нескольких переменных в виде суперпозиций непрерывных функций одного переменного и сложения, ДАН СССР, т. 114, № 5 (1957), 953—956.).
10. Кормаков, Г. В. (2002) Интерпретация процесса переобучения искусственных нейронных сетей с использованием тропической геометрии / Г. В. Кормаков // Прикладная математика и информатика. – Москва: ООО "МАКС Пресс", 2022. – С. 86-105. – EDN ZRDXMF.
11. Кривулин Н.К. (2009), Методы идемпотентной алгебры в задачах моделирования сложных систем. – СПб: Изд-во С.-Петербур. Ун-та, 2009. – 256 с.
12. Маслов В.П., Колокольцев (1994), Идемпотентный анализ и его применение в оптимальном управлении. М.: Физматлит, 1994, ISBN 5-02-014886-5. 144 с.
13. Baldwin E., Klemperer P. (2012): "Tropical Geometry to Analyse Demand," Working Paper, Nuffield College, [PDF Tropical Geometry to Analyse Demand \(researchgate.net\)](https://www.researchgate.net/publication/312211111)
14. Ben-Ishai, G., Dean, J., Manyika, J., Porat, R., Varian, H., Walker, K. (2024), AI and the Opportunity for Shared Prosperity: Lessons from the History of Technology and the Economy arXiv preprint arXiv:2401.09718, 2024, Equal contributions, the authors are affiliated with Google and Alphabet.
15. Bhatia, S., Cao, Y., Lezeau, P., Monod, A. (2024) Tropical Expressivity of Neural Networks, arXiv:2405.20174v1 [cs.LG] 30 May 2024
16. Bresnahan T. F., Trajtenberg, M. (1995) General purpose technologies 'Engines of growth'? Journal of Econometrics, Volume 65, Issue 1, January 1995, Pages 83-108
17. Gilmer, J., Metz, L., Faghri, F., Schoenholz, S., Raghu, M., Wattenberg, M. & Goodfellow, I.n (2018) The Relationship Between High-Dimensional Geometry and Adversarial Examples, arXiv:1801.02774v3 [cs.CV] 10 Sep 2018
18. Ben-Ishai, G., Dean, J., Manyika, J., Porat, R., Varian, H., Walker, K. (2024), AI and the Opportunity for Shared Prosperity: Lessons from the History of Technology and the Economy arXiv preprint arXiv:2401.09718, 2024, Equal contributions, the authors are affiliated with Google and Alphabet.
19. Dube S. (2018) High Dimensional Spaces, Deep Learning and Adversarial Examples, arXiv:1801.00634v5 [cs.CV] 15 Apr 2018.
20. Dube S. (2021) An Intuitive Exploration of Artificial Intelligence: Theory and Applications of Deep Learning". Publisher: Springer Nature.
21. Klemperer, P. (2008) A New Auction for Substitutes: Central Bank Liquidity Auctions, the U.S. TARP, and Variable Product-Mix Auctions. <http://www.pauklempere.org>
22. D. Hilbert, Gesammette Abhandlungen Vol. 3, No. 17, Berlin, 1935.
23. Klemperer P. (2010) The product-mix auction: A new auction design for differentiated goods. Journal of the European Economic Association, 8(2-3):526-536, 2010.
24. Ziming Liu, Yixuan Wang, Sachin Vaidya, Fabian Ruehle, James Halverson, Marin Soljac'ic', Thomas Y. Hou, Max Tegmark, (2024) KAN: Kolmogorov–Arnold Networks, <https://arxiv.org/abs/2404.19756>
25. Simon I. (1988), *Recognizable sets with multiplicities in the tropical semiring*, Mathematical foundations of computer science, 1988 (Carlsbad, 1988), Lecture Notes in Comput.Sci., vol. 324, Springer, Berlin, 1988, pp. 107-120. 1

26. Varian, H. (2021) Economics at Google: The first ten years. *Bus Econ* 56, 195–199 (2021). <https://doi.org/10.1057/s11369-021-00243-2>
27. Zuboff, Sh. (2019) *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*. New York, Public Affairs, 2019, 704 pp.
28. Varian H (2019), *Artificial Intelligence, Economics, and Industrial Organization* // National Bureau of Economic Research, Volume Title: *The Economics of Artificial Intelligence*, p. 399 – 419
29. Varian H (2018), *Use and Abuse of Network Effects*". *Toward a Just Society: Joseph Stiglitz and Twenty-First Century Economics*, edited by Martin Guzman, New York Chichester, West Sussex: Columbia University Press, 2018, pp. 227-239. <https://doi.org/10.7312/guzm18672-013>
30. Varian, H. (2013) Beyond Big Data. *Bus Econ* 49, 27–31 (2014). <https://doi.org/10.1057/be.2014.1>
31. Varian, Hal R. (2010). "Computer Mediated Transactions," *American Economic Review*, VOL. 100, NO. 2, MAY 2010, pp. 1–10.

References in Cyrillics

1. Arnol'd V. I. (1957) O funkciyax trex peremenny'x, *DAN SSSR*, t. 114, № 4 (1957), 679—681.
2. Arnol'd V. I. (1959) O predstavlenii nepreryvny'x funkciy trex peremenny'x superpoziciyami nepreryvny'x funkciy dvux peremenny'x, *Matem. sb.*, 1959, tom 90, nomer 1, 3–74
3. Vorob'ev, N. N. (1963) E'kstremalnaya algebra matricz, *Dokl. AN SSSR*, 1963, tom 152, № 1, 24–27
4. Kantorovich L.V. (1960) E'konomicheskiy raschet nailuchshego ispol'zovaniya resursov, M.: Izdatel'stvo AN SSSR, 1960.
5. Kozyrev A.N. (2018), *Cifrovaya e'konomika i cifrovizaciya v istoricheskoy perspektive* // *Cifrovaya e'konomika*, № 1, 2018, s.5-19, DOI: 10.34706/DE-2018-01-01
6. Kozyrev A.N. (2024) *Cifrovaya e'konomika i e'konomika danny'x* // *Cifrovaya e'konomika* № 2(28), 2024 – s. 5–13. DOI: 10.33276/de-2024-02-01
7. Kolmogorov A.N. (1965) Tri podxoda k opredeleniyu ponyatiya «kolichestvo informacii» // *Problemy peredachi informacii*. – 1965. – T.1 – Vy'p. 1. – s. 3–11.
8. Kolmogorov A. N. (1956), O predstavlenii nepreryvny'x funkciy neskol'kix peremenny'x super-poziciyami nepreryvny'x funkciy men'shego chisla peremenny'x, *DAN SSSR*, t. 108, № 2 (1956), 179-182.
9. Kolmogorov A.N. (1957), O predstavlenii nepreryvny'x funkciy neskol'kix peremenny'x v vide superpozicij nepreryvny'x funkciy odnogo peremennogo i slozheniya, *DAN SSSR*, t. 114, № 5 (1957), 953-956.).
10. Kormakov, G. V. (2002) Interpretaciya processa pereobucheniya iskusstvenny'x nejronny'x setej s ispol'zovaniem tropicheskoj geometrii / G. V. Kormakov // *Prikladnaya matematika i informatika*. – Moskva: OOO "MAKS Press", 2022. – S. 86-105. – EDN ZRDXXMF.
11. Krivulin, N.K. (2009), *Metody idempotentnoj algebrы v zadachax modelirovaniya slozhny'x sistem*. – SP: Izd-vo S.-Peterb. Un-ta, 2009. – 256 s.
12. Maslov V.P., Kolokol'cev (1994), *Idempotentny'j analiz i ego primenenie v optimal'nom upravlenii*. M.: Fizmatlit, 1994, ISBN 5-02-014886-5. 144 s.

Ключевые слова

Большие данные, компьютерно-опосредованные транзакции, генеративные сети, глубинное обучение, обучение с подкреплением, тропическая математика

Козырев Анатолий Николаевич, к.ф.-м.н., д.э.н
Центральный экономико-математический институт РАН
 ORCID 0000-0003-3879-5745,
kozyrevan@yandex.ru

Anatoly Kozyrev, Data Economy, neural network learning and multidimensional geometry

Keywords

Big Data, computer-mediated transactions, generative networks, deep learning, reinforcement learning, tropical mathematics.

DOI: 10.34706/DE-2024-03-01

JEL classification C8 Методология сбора и оценки данных; компьютерные программы; O33 – Научно-технический прогресс: этапы и последствия; процесс распространения

Abstract

The article uses specific examples to show how high-class mathematics sometimes helps to solve problems in the field of economics and information technology, for which it was not originally intended. The role of outstanding Russian mathematicians of the relatively recent past is particularly emphasized. interpreter.