

УДК 004.89

doi: 10.15622/rcai.2025.018

К ПРОБЛЕМЕ ИНТЕГРАЦИИ СТАТИСТИЧЕСКИХ И ДЕТЕРМИНИСТСКИХ МЕТОДОВ ИНТЕЛЛЕКТУАЛЬНОГО АНАЛИЗА ЭМПИРИЧЕСКИХ ДАННЫХ

М.И. Забежайло (*m.zabezhailo@yandex.ru*)

Федеральный исследовательский центр
«Информатика и управление» РАН, Москва

Представлены некоторые возможности интеграции статистических и детерминистских методов, используемых в интеллектуальном анализе эмпирических данных. Подход ориентирован на поиск неявным образом заданных причинно-следственных зависимостей, связывающих «причину» и «контекст ее актуальности» с «вызываемыми» ими «следствиями». Обсуждается пример использования этого подхода в области высокотехнологичной медицинской диагностики.

Ключевые слова: интеллектуальный анализ данных, тернарное отношение причинности, медицинская диагностика

Введение

Наряду с обширным перечнем решений в области стремящихся к автономности своего функционирования прикладных систем искусственного интеллекта (ИИ) – различных аппаратно-программных комплексов, располагающихся на борту того или иного подвижного – передвигающегося по земле или по воде, летающего и т.п. – устройства, а также завоевавших сегодня фантастическую популярность Больших Языковых Моделей (БЯМ), все более заметный интерес как разработчиков так и влиятельных пользователей наблюдается в настоящее время к ИИ-системам и решениям, ориентированным на оперативный анализ больших объемов данных и поддержку управленческих решений. Актуальные примеры таких потребностей не сложно найти, в частности, в области управления большими информационно-телекоммуникационными структурами (см., в том числе – [Verma et.al., 2015], [Tirmazi et.al., 2020], [Смирнов и др., 2024] и др.) обеспечения обороны и безопасности ([Entous, 2025a-b] и др.), противодействия мошенничествам в финансовой сфере [Грушо и др., 2021], а также ряде других значимых областей приложений компьютерных методов и технологий.

Проблемно-ориентированный анализ Big Data и поддержка принятия управленческих решений в режиме жестких ограничений по времени сегодня – это область востребованности достаточно «тонко» организованных математических моделей, методов и прикладных систем ИИ. Уже при поверхностном взгляде на рассматриваемую проблематику исследований и разработок, возникает несколько вопросов фундаментального характера, в частности:

- Как именно «бороться» (см. ограничения по времени анализа данных и поддержки принятия решений) с эффектом *Big*, а также связанным с ним эффектом *Open*? (например, как интегрировать «быстрые» заключения статистического анализа данных «в среднем» – т.е. в определенном смысле – приближенные, дающие результат с точностью до подкласса статистически неразличимых, воспринимаемых средствами статистического анализа как «однородные» – объектов) с точными, однако, как правило, требующими при их поиске большого перебора возможных вариантов) детерминистскими средствами?

- Как оценивать достаточность оснований для принятия предлагаемых системой ИИ выводов и заключений? (Своего решения здесь требует основополагающая для ИИ проблема доверия к результатам, формируемым системой искусственного интеллекта).

- Как обеспечить принимающих ответственные решения лиц (ЛПР) неформальной интерпретацией и объяснением формируемых системой ИИ рекомендаций. Необходимость такого «сервиса» обусловлена ответственностью за последствия решений, которую ЛПР предстоит принять на себя?

Один из вариантов использования возможностей интеллектуально-го анализа данных (ИАД) для ответа на эти вопросы и будет предметом представленного ниже обсуждения.

1. Компьютерный анализ эмпирических данных: от логики доказательства к логикам рассуждений

Проблематика компьютерного анализа данных эмпирической природы – накапливаемых результатов лабораторных экспериментальных исследований того или иного физического или технического эффекта, клинических данных о возникновении и развитии у пациентов заболеваний той или иной определенной нозологии, данных мониторинга возникновений и распространения сбоев в крупных ИТ-структурах и т.п. – имеет ряд специфических особенностей, недостаточное внимание к которым может привести к формированию неадекватных заключений и выводов.

Так в целом ряде случаев задачи обсуждаемого типа требуют аккуратно учитывать открытый характер анализируемой предметной области – присущего Big Data так называемого эффекта *Open*, обусловленного возможностями появления в некоторый момент времени таких новых данных, кото-

рые ранее еще не приходилось анализировать и о существовании которых на текущий момент еще не было накоплено какой-либо значимой информации. С этим явлением, с легкой руки Н.Талеба получившим название эффекта *Черного лебедя* [Талеб, 2015], нам пришлось столкнуться, в т.ч., в период пандемии COVID-19, когда практически каждая очередная мутация вируса требовала изменений или же вообще новых подходов в борьбе с ней – новых препаратов, новых медицинских протоколов лечения и т.п.

Необходимость оперировать в условиях эффекта *Open* указала на неадекватность применения в таких условиях некоторых традиционных методов анализа данных, причем – еще «до компьютера», т.е. на уровне математической модели и алгоритмов анализа данных. Так, например, применение традиционных методов статистического анализа данных основано на идее формирования так называемой *генеральной совокупности*, затем выделения из нее *репрезентативных выборок* и последующего «обучения» на таких выборках для того, чтобы в рамках математических моделей интерполяционно-экстраполяционного типа (интерполяция «обучающей» выборки эмпирическими зависимостями заданного типа, например, регрессиями того или иного вида, а затем «диагностика» вновь анализируемого объекта проверкой экстраполируемости на него какой-либо из уже найденных интерполяционных зависимостей) обосновать «аналогию» анализируемого нового прецедента с теми, которые собраны в репрезентативной выборке из *генеральной совокупности*. Однако, в открытых предметных областях (см. эффект *Open*) само понятие генеральной совокупности как коллекции данных, в которой отражены все варианты проявления изучаемого эффекта, оказывается, вообще говоря, не универсальным. Потенциально возможное появление соответствующего *Черного лебедя* (см. [Талеб, 2015] и др.) ставит под сомнение *генеральный* характер той или иной конкретной *совокупности* данных. И дело здесь вовсе не в использовании тех или иных компьютерных методов анализа данных, фундаментальная проблема – в *репрезентативности* используемых в каждом конкретном случае «обучающих» выборок из претендующей на универсальность конкретной совокупности эмпирических данных (ЭД).

Еще одной характерной особенностью компьютерного анализа ЭД является фокусировка внимания не на утверждениях универсального характера, справедливых для всей изучаемой предметной области или же ее «аналитически» выделяемого «целевого» фрагмента, а на специальном классе контекстно-определяемых утверждений – заключениях и выводах, принимаемых (т.е. оцениваемых как приемлемые, заслуживающие внимания) с точностью до релевантности заданному «контексту». Например, это могут быть заключения, которые релевантны набору эмпирических фактов (ЭФ), ранее уже «установленных» экспериментальным путем, – в частности, данным конкретных медицинских «тестов», результатам конкретных лабораторных «измерений» и т.п.

Таким образом, вполне естественным выглядит разделение процедурных «инструментов» – логико-математических моделей, методов и алгоритмов, ориентированных на формирование «универсальных» доказуемых утверждений (будем называть такие «инструментальные» средства анализа данных логиками доказательства), и тех, задача которых – сформировать «рациональные» следствия (причем – не только сугубо дедуктивного характера, но и те, которые используют формализованные варианты и других познавательных процедур – формирования индуктивных обобщений, рассуждений по аналогии, построения абдуктивных объяснений и др. [Финн, 2021, 2024]). Как следствие, актуальной оказывается задача порождения из накапливаемых ЭД не только «универсально» не оспариваемых утверждений (доказуемо корректных, как уже отмечалось выше, для всей анализируемой предметной области), но и тех – в некотором смысле «частных» (не универсальных) заключений, которые (как это удастся продемонстрировать) оказываются неоспариваемыми относительно заранее заданного «контекста» – например, набора определенных теоретических утверждений («априорных» гипотез) и набора некоторых ЭФ, зафиксированных объективными средствами в ходе экспериментов, которые выполнены с соблюдением определенных «правил». Примеры востребованности подобного рода заключений и выводов нетрудно найти в задачах диагностического типа [Забежайло и др., 2021] – в медицинской или технической диагностике, идентификации и противодействию мошенничествам в финансовой сфере, в задачах обеспечения кибербезопасности и др.

Задействованные при этом формализованные средства формирования «рациональных» заключений и выводов из накапливаемых ЭД представляется естественным рассматривать как логики рассуждений – логико-математический «инструментарий» систематического порождения *следствий* из анализируемого *контекста* в его текущем состоянии. Таким образом, логика рассуждений возникает как объединение двух этапов (и двух типов «инструментальных» средств анализа данных):

- *получения следствий* (заключений, причем – не обязательно строго дедуктивного характера) из имеющегося на текущий момент (и потенциально – расширяющегося – см. эффект *Open*) контекста, а также
- *оценки достаточности оснований для принятия* порожденных на первом этапе следствий (*оценки доверия* к таким следствиям).

При этом наряду с разработкой «инструментов» собственно порождения следствий из контекстов – математических моделей, методов и алгоритмов, позволяющих использовать формализованные средства рассуждения для формирования неоспариваемых на заданном контексте выводов и заключений, критически значимой оказывается разработка средств оценки доверия к полученным в процессе выполненных рассуждений результатам («инст-

рументов» оценки достаточности оснований для принятия выводов и заключений, полученных формализованными средствами). Таким образом, важно иметь соответствующие процедурно-«инструментальные» возможности не только рассчитать – целенаправленно «вывести» из анализируемого набора имеющихся ЭД с помощью процедур контролируемого «наследования» корректности – соответствующие «следствия», но и оценить возможности доверять им (оценить их неоспариваемость на уже накопленном контексте – имеющихся на данный момент ЭД, а также, разумеется, если это окажется возможным, – идентифицировать¹ неоспариваемость этих заключений на всей анализируемой предметной области, т.е. обосновать «универсально»-доказуемый характер полученных «следствий»).

Логико-математический «инструментарий» подобного типа – специальные средства логики рассуждений уже достаточно давно привлекают внимание специалистов. Так, например, еще в конце 70-х годов XX века чешские математики П. Гаек и Т. Гавранек предложили [Hajek et.al., 1978] оригинальный подход, который позволял автоматизированными компьютерными средствами выдвигать гипотезы на основе накапливаемых ЭД, а затем оценивать статистическими средствами достаточность оснований для их принятия. В работе [ван Бентем, 2011] Йохан ван Бентем сформулировал необходимость создания логики рассуждений, дополняющей логику доказательства средствами поддержки взаимоотношений между логическими и эмпирическими фактами. Обширное отражение в научной литературе получили исследования методов и средств порождения зависимостей на базе накапливаемых ЭД (см., например, [Agrawal et. al., 1996], [Hajek, 2001]).

Однако, обсуждаемая проблема разработки инструментов формирования и последующей оценки достаточности оснований для принятия результатов интеллектуального анализа данных (ИАД) в постоянно расширяемых новыми сведениями контекстах – коллекциях ЭФ – по-прежнему далека от общепризнанного решения. Дополнительные «краски» к сложившейся на текущий момент «картине мира» добавили представления о границах доказуемого и недоказуемого в условиях эффекта *Open* (см. в частности, естественные ограничения дедуктивной доказуемости или уже упоминавшиеся выше проблемы с формированием *генеральной совокупности* при применении методов статистического анализа данных.). В свою очередь, представления о доказуемости как о возможности приведения к неоспариваемости на уже накопленных ЭД потребовало уточнения требуемых границ такого контекста. Возможности пополнения текущего набора ЭД релевантной цели анализа новой информацией привели к необходимости задуматься над задачей проверки сохранения неоспариваемости ранее уже полученных выводов и заключений при направленном

¹ Или же, наоборот, – опровергнуть.

пополнении анализируемых данных новыми сведениями. Естественным следствием этому стал интерес к неформальной интерпретации и объяснению результатов ИАД. При этом, по-видимому, наиболее продуктивным оказался подход к объяснению, трактуемому как ответ на вопрос *ПОЧЕМУ?* ([Pearl, 1995] [Zabekhailo, 2021] и др.). Именно этим, в свою очередь, можно объяснить сегодняшний интерес сообщества исследователей ИИ к задачам разработки проблемно-ориентированных математических моделей, методов и алгоритмов восстановления причинно-следственных зависимостей, которые изначально скрыты в анализируемых ЭД ([Pearl, 1999, 2000] и др.).

2. Компьютерный анализ постоянно пополняемых эмпирических данных

Потребность оперировать `data_set`'ами ограниченного размера при решении задач компьютерного анализа данных не нова и достаточно хорошо знакома специалистам. Еще без малого 60 лет назад Ю.И.Журавлев предложил элегантную математическую конструкцию – так называемые корректные алгебры над множеством некорректных (эвристических) алгоритмов [Журавлев, 1977], которая позволила получить эффективные решения в целом ряде важных для страны прикладных задач². Однако, в тех задачах, как и в целом в классическом машинном обучении, речь шла о выборе наиболее «точного» (минимизирующего ошибки классификации) на заданном анализируемом `data_set`'е алгоритма в заданном семействе алгоритмов. В настоящее время ситуация принципиально иная: необходимо оперировать малыми³ по числу элементов `data_set`'ами, которые последовательно пополняются новыми данными той же «природы». В формализованном виде это требует разработки подходов, математических моделей и методов анализа данных, представляемых расширяющимися (по крайней мере – в части числа их строк – перечня объектов-прецедентов, входящих на текущий момент в анализируемый `data_set`) «плоскими»⁴ матрицами ОБЪЕКТЫ x ПРИЗНАКИ. Собственно, это и стало дифференцирующим фактором, позволяющим отделить классическое ML от ИАД, ориентированного, помимо прочего, и на учет динамики изменений в анализируемых данных. В свою очередь, работа с последовательностями динамически изменяемых (так называемых расширяющихся [Finn, 2019, 2023]) баз фактов средствами интерполяционно-экстраполяционных математических методов требует формирования проблемно-ориентированных «инструментов» контроля надежности ин-

² В т.ч., поиска промышленных месторождений золота [Журавлев, 2024].

³ Одним из формальных уточнений этого понятия может быть статистическая незначимость размеров таких наборов данных.

⁴ Где число строк, как правило, существенно меньше числа столбцов.

дуктивных (экстраполяционных) обобщений, формируемых на таких данных средствами ИАД. Результативным математическим «инструментарием», позволяющим выделять в множестве результатов ИАД подмножества таких, надежность которых не вызывает сомнений, стал поиск в анализируемых данных эмпирических зависимостей (ЭЗ) причинно-следственного типа – каузальных маркеров изучаемых эффектов.

Наиболее распространенным вариантом математических средств выявления в анализируемых данных эмпирических зависимостей – маркеров изучаемых эффектов – на сегодня являются методы статистического анализа данных (САД). Примером популярной математической техники САД может служить часто используемый в медицинских приложениях подход Каплана-Майера, в рамках которого результаты поиска эмпирических закономерностей в иерархии случайных величин (которыми в процессе выполняемого САД представлены клинические признаки) кладутся в основу прогнозирования исхода наблюдаемого заболевания ([Ардашев и др., 2024] и др.). Выявление взаимосвязей между группами клинических признаков (идентификация маркеров целевого эффекта), а также оценка меры их причинно-следственной взаимосвязанности позволяют целенаправленно формировать тактику персонализированных медицинских воздействий.

Однако, одним из уязвимых мест этой процедурной конструкции оказывается проблема «предсказательной силы» маркеров выбранного типа. Заключение, формируемые на базе частот встречаемости отдельных признаков в анализируемых *data_set*'ах, надежность (устойчивость) собираемых далее комбинаций таких признаков, оценки доверия и общая «прогностическая сила» финального заключения, к сожалению, далеко не всегда могут вести к выводам и рекомендациям неоспариваемого характера. Важным компонентом в множестве причин возникновения ситуаций этого типа оказывается то, что, хотя выделяемый в процессе САД фактор или комбинация факторов влияют на эффект «достаточно часто», тем не менее, есть прецеденты, в которых этот фактор\комбинация факторов наличествуют, а целевой эффект при этом отсутствует. Т.е. не смотря на то, что данная комбинация факторов «во *многих* случаях» позволяет указать на возможное присутствие изучаемого целевого эффекта у рассматриваемых объектов-прецедентов, она – эта комбинация факторов – в общем случае не годится для формирования на ее основе неоспариваемого заключения (например, неоспариваемого диагноза у конкретного пациента).

Что же – какие именно математические модели и методы САД – предложить в таких ситуациях для устранения некорректности в заключениях выполняемого ИАД (например, формирования неоспариваемого медицинского диагноза и т.п.). Как показали соответствующие исследования, одним из вариантов результативных действий может обеспечить ориентация на рабочую гипотезу вида:

помимо самого «базового» фактора влияния (или определенной комбинации таких факторов) для наличия целевого эффекта критически значим и определенный «контекст»⁵, во взаимодействии с которым «базовый» фактор обеспечивает появление целевого эффекта.

Как следствие, возникает задача: какими средствами идентифицировать такого рода контекст по имеющимся `data_set` у и целевому эффекту?

3. Предлагаемый вариант решения

В общем виде предлагаемый вариант решения, опирающийся на сформулированную выше рабочую гипотезу, характеризует следующая процедурная конструкция: будем использовать ДСМ-ИАД ([Финн, 2021, 2024], [Grusho et.al., 2024], [Грушо и др., 2021] и др.) на базе тернарного отношения причинности вида:

$$\begin{aligned} < [\text{статистически значимая}^6 \text{ комбинация} \\ &\text{факторов причинного влияния}]; [\text{контекст}] => \\ &=> [\text{целевой эффект}] >. \end{aligned}$$

Ключевые отличительные особенности предлагаемого подхода:

- неопровергаемость заключения ДСМ-ИАД на имеющихся на текущий момент эмпирических данных (в т.ч. – с учетом контрафактуальных ([Höfler, 2005], [Pearl, 1999, 2000] и др.) заключений, и выполнимости условия Запрета на КонтрПримеры ([Финн, 2021, 2024] и др.);
- неформальная интерпретируемость формируемых эмпирических зависимостей – маркеров целевого эффекта (и, как следствие, минимизация эффектов переобучения в процессе компьютерного анализа данных используемыми статистическими и детерминистскими средствами);
- объяснение результативности формируемых ЭЗ-маркеров целевого эффекта за счет использования (выявления) каузальных зависимостей, изначально скрытых в анализируемых эмпирических данных;
- интеграция статистических и детерминистских средств компьютерного анализа данных (идея Тьюки [Tukey, 1977]): сперва «быстрый» поиск базовых факторов влияния статистическим средствами, а затем – последующая идентификация «контекста» их «продуктивности» детерминистскими средствами (как оказалось, не требующая исчерпывающего перебора формируемых в процессе ДСМ-ИАД всех вариантов гипотез о причинах).

⁵ Некоторые определенные дополнительные факторы, принимающие некоторые специальные – взаимосвязанные со значениями «базового» фактора – значения.

⁶ Идентифицированная средствами той или иной версии стат.анализа данных.

4. Как это работает. Пример

В процессе изучения эффекта так называемой псевдопрогрессии (*ПсП*) опухолей головного мозга человека специалистами Национального медицинского исследовательского центра нейрохирургии имени академика Н.Н. Бурденко (НМИЦ НХ им. Н.Н. Бурденко) на ЭД 427 пациентов, накопленных почти за 20 лет клинических исследований, были идентифицированы 59 неоспариваемых прецедентов наличия эффекта *ПсП* и 368 случаями его отсутствия. При этом описание каждого из анализируемых прецедентов содержит значения более 150 отдельных параметров (клинических признаков⁷). У 48 из 59 пациентов с *ПсП* при этом зафиксировано наличие кисты в опухоли (45 имели чисто кистозную *ПсП*, а 3 – так называемую смешанную⁸). Таким образом, фактор наличие кисты в опухоли характеризовался высоким значением частоты встречаемости на прецедентах *ПсП* (что дает основания использовать его как указатель на соответствующую группу «риска»). Однако, у 78 из 368 пациентов без *ПсП* также наблюдалась киста в опухоли. Таким образом, только один этот клинический признак не представляется возможным использовать для надежного прогноза наличия/отсутствия эффекта *ПсП* в целом по рассматриваемому множеству пациентов.

В работе [Трунин, 2021] был предложен механизм (основанный на модели САД в духе метода Каплана-Майера ([Ардашев и др., 2024] и др.), позволяющий проводить дифференциальную диагностику эффектов *ПсП* и рецидива опухолей (*РО*). В его основе – полученный Ю.Ю. Труниным САД-критерий, интегрирующий три клинических параметра – *возраст пациента (младше 11 лет, 11 лет и старше)*, *локализация опухоли (супратенториальная, инфратенториальная)* и *наличие кисты в опухоли (да, нет)*. Риск возникновения эффекта *ПсП*, как показал Ю.Ю. Трунин, становится высоким (54%) уже через 6 месяцев после лучевой терапии (ЛТ), через 12 мес. после ЛТ он возрастает до 95.2%, и подходит к 100% через 48 мес. после ЛТ. В свою очередь, на имеющихся ЭД рецидивы опухоли возникают через 2 и более года (медиана 29 мес.) после ЛТ. Все это позволяет сформировать эффективные критерии отличия эффектов *ПсП* и *РО*.

Однако, открытым оставался ряд вопросов, связанных с *персонифицированной* диагностикой эффектов *ПсП* и *РО* (т.е. с возможностью формировать для конкретного пациента неоспариваемое⁹ заключение, учитывающее не только общие характеристики соответствующей целевой группы, но и его персональные особенности).

⁷ Мультимодальной природы количественных, качественных и др.

⁸ При которой имеются признаки как кистозного, так и солидного характера.

⁹ Например, на имеющемся на текущий момент массиве эмпирических данных.

Для решения этой задачи персонификации диагноза была использована представленная (см. Разделы II-III) математическая техника ИАД на основе тернарного отношения причинности. Было показано (см. [Забейко и др., 2024]), что совместное использование определенных значений дополнительных факторов – *гистология опухоли* и *локализация опухоли*, объединенных в виде сопутствующего контекста вместе с фактором *киста в опухоли*, позволяет получить неоспариваемый на имеющихся ЭД прогноз возникновения эффекта *ПсП*. При этом сформированный контекст «выдержал» проверки контрафактуальности по всем остальным имеющимся эмпирическим данным (находящимся за пределами множеств значений параметров, выделенных при формировании этого контекста).

Таким образом, дополнение статистически значимого фактора *киста в опухоли* идентифицируемым детерминистскими средствами контекстом позволило формировать неоспариваемые на имеющемся эмпирическом материале персонифицированные заключения относительно возникновения эффекта *ПсП*, в том числе – у отдельно изучаемого пациента.

Заключение

Актуальной для эффективного применения методов ИАД в практически значимых приложениях по-прежнему остается на сегодняшний день проблема «управления» размерностью множества параметров, характеризующих анализируемый *data_set*. Активно разрабатываются различные методы сокращения размерности, позволяющие «склеивать» «малосущественные» переменные без существенной потери в точности результата, формируемого на «усеченных» таким способом данных. Не теряет своей популярности классическая идея выделения *главных компонент* в анализируемых многомерных данных, а метод PCA [Pearson, 1901] – был и остается прекрасным примером такого математического инструментария для работы с данными метрического характера.

Представленная в данной работе версия ИАД на базе контекстного отношения сходства также может рассматриваться как пример использования аналогичного подхода (в основе которого – «склейки» некоторых параметров в описаниях объектов-прецедентов) при работе с неметрическими данными в решении задач диагностического типа.

Таким образом, отдельного внимания, по-видимому, заслуживают возможности, которые открывает интеграция этих двух подходов в задачах анализа больших наборов мультимодальных данных. Практически значимый пример – задачи поддержания так называемой ситуационной осведомленности ЛПР в ряде актуальных приложений ([Entous, 2025a-b] и др.).

Список литературы

- [Ардашев и др., 2024] Ардашев В.Н., Калёнова И.Е., Ляпкина Н.Б., Потехин Н.П., Фурсов А.Н. (Под ред. Проф. Бояринцева В.В.) Доказательная медицина: обзор современных математических методов анализа. – М.: УД Президента РФ, 2013. – 224 с. – <https://volynka.ru/Articles/Text/124?ysclid=m8lk721721581580264>.
- [ван Бентем, 2011] Бентем ван, Й. Логика и рассуждение: много ли значат факты? // Вопросы философии. – 2011. – №12. – С. 63-76.
- [Грушо и др., 2021] Грушо А.А., Забейайло М.И., Смирнов Д.В., Тимонина Е.Е. Интеллект. анализ пополняемых коллекций Big Data в режиме процессно-реального времени // Информатика и ее применения. – 2021. – Т. 15, № 2. – С. 36-43.
- [Журавлев, 1977] Журавлев Ю.И. Корректные алгебры над множествами некорректных (эвристических) алгоритмов. I-III // Кибернетика. – 1977. – № 4. – С. 5-17, 1977. № 6. С. 21-27, 1978. № 2. – С. 35-43.
- [Журавлев, 2024] Журавлев Ю.И. Распознавая образ / Колл. авт.; ред. А.Л. Семенов, Ю.В. Чехович и Е.О. Самойлова. – М.: Буки Веди, 2024. – 512 с.
- [Забейайло и др., 2021] Забейайло М.И., Грушо А.А., Грушо Н.А., Тимонина Е.Е. Поддержка решения задач диагностического типа // Системы и средства информатики. – 2021. – Т. 31, № 1. – С. 69-81.
- [Забейайло и др., 2024] Забейайло М.И., Михеенкова М.А., Трунин Ю.Ю. О небинарной версии отношения причинности в интеллектуальном анализе онкологических данных // НТИ, Сер. 2 «Информ. проц. и системы». – 2024. – №. 6. – С. 13-20.
- [Смирнов и др., 2024] Смирнов Д.В., Грушо А.А., Забейайло М.И. К задаче идентификации сбоев в информационно-технологической инфраструктуре путем мониторинга и анализа косвенных данных // Системы и средства информатики. – 2024. – Т. 34. – Вып. 3. – С. 14-22.
- [Талёб, 2015] Талёб Н.Н. Черный лебедь. Под знаком непредсказуемости. – М.: КоЛибри, 2015. – 36 С.
- [Трунин, 2021] Трунин Ю.Ю. Стереотаксическое облучение в комплексном лечении пациентов с пилоидными астроцитомами: дис. ... д-ра мед. наук: 14.01.18 – нейрохирургия и 14.01.13 – лучевая диагностика, лучевая терапия. – М.: НМИЦ НХ им. Н. Н. Бурденко, 2021. – 294 с.
- [Финн, 2021] Финн В.К. Искусственный интеллект: методология, применение, философия. – М.: ЛЕНАНД, 2021. – 468 с.
- [Финн, 2024] Финн В.К. ДСМ-метод автоматизированной поддержки исследований и аппарат понятий для искусственного интеллекта // Искусственные общества. – 2024. – Т. 19. – Вып. 2. – <https://artsoc.jes.su/s207751800030907-6-1/>.
- [Agrawal et. al., 1996] Agrawal R., Manilla H., Sukent R., Toivonen A., Verkamo A. Fast discovery of Association rules // In: Advance in Knowledge Discovery and Data Mining. – P. 307-328. AAAI, Menlo Park, 1996.
- [Entous, 2025a] Entous A. The Partnership: The Secret History of the War in Ukraine // The New York Times. – March 29, 2025. – <https://www.nytimes.com/interactive/2025/03/29/world/europe/us-ukraine-military-war-wiesbaden.html>.

- [**Entous, 2025b**] Entous A. Key Takeaways From America's Secret Military Partnership With Ukraine // The New York Times. – March 30, 2025. – https://www.nytimes.com/2025/03/30/world/europe/us-ukraine-military-war-takeaways.html?unlocked_article_code=1.8E4.nuPY.ksQThlmbAR9A&smid=nytcare-ios-share&referringSource=articleShare.
- [**Finn, 2019**] Finn V.K. On the Heuristics of JSM Research (Additions to Articles) // Autom. docum. and mathematical linguistics. – 2019. – Vol. 53, No. 5. – P. 250-282.
- [**Finn, 2023**] Finn V.K. On Empirical Regularities in the JSM Method of Automated Research Support // Automatic Documentation and Mathematical Linguistics. – 2023. – Vol. 57. – No. 6. – P. 362-381.
- [**Grusho et.al., 2024**] Grusho A., Grusho N., Zabezhailo M., Timonina E. (2024). On Some Possibil. of Using AI Methods in the Search for Cause-And-Effect Relat. in Accumul. Empirical Data. // In: Kovalev S., Kutenko I., Sukhanov A., Li Y., Li Y. (eds) // Proc. of the 8th Int. Scient. Conf. "Intell. Inform. Techn. for Industry" (IITI'24), Vol. 2. Lect. Not. in Netw. and Syst. V. 1210. – Springer, Cham. – P. 280-290.
- [**Hajek et.al., 1978**] Hájek P., Havránek T. Mechanizing Hypothesis Formation (Mathematical Foundations for a General Theory). – Springer, Heidelberg, 1978. – <https://link.springer.com/book/10.1007/978-3-642-66943-9>. – P. 396.
- [**Hajek, 2001**] Hájek P. The GUHA method and mining association rules // In: Proc. CIMA 2001, Bangor, Wales, 2001. – P. 533-539.
- [**Höfler, 2005**] Höfler M. Causal inference based on counterfactuals // BMC Med. Res. Methodol. – 2005. – 5, 28. – <https://bmcmmedresmethodol.biomedcentral.com/articles/10.1186/1471-2288-5-28>.
- [**Pearl, 1995**] Pearl J. Causal diagrams for emp. Research // Biometrika. – 1995. – 82. – P. 669-710.
- [**Pearl, 1999**] Pearl J. Probabilities of causation: three counterfactual interpretations and their identifications // Synthese. – 1999. – 121. – P. 93-149.
- [**Pearl, 2000**] Pearl J. Causality: models, reasoning, and inference. – N.-Y.: Cambr. Univ. Press, 2000. – 384 p.
- [**Pearson, 1901**] Pearson K. On lines and planes of closest fit to systems of points in space // Philos. Mag. – 1901. – 2(11). – P. 559-572.
- [**Tirmazi et.al., 2020**] Tirmazi M., Barker A., Deng N., Haque Md Em., Qin Z.G., Hand S., Harchol-Balter M., Wilkes J. Borg: the Next Generation // EuroSys '20: Proc. of the 15th Europ. Conf. on Computer Systems. – 2020. – Article No.: 30. – P. 1-14.
- [**Tukey, 1977**] Tukey J.W. Exploratory data analysis. – Reading, Mass.: Addison-Wesley Pub. Co., 1977. – 712 p.
- [**Verma et.al., 2015**] Verma A., Pedrosa L., Korupolu M., Oppenheimer D., Tune E., Wilkes J. Large-scale cluster management at Google with Borg // EuroSys '15: Proc. of the 10th Eur. Conf. on Computer Systems. – 2015. – Article No. 18. – P. 1-17.
- [**Zabezhailo, 2021**] Zabezhailo M.I. Models of Explanation in Intelligent Data Analysis // IMSC-2021: Integr. Models and Soft Comp. in AI Russian Advances in Fuzzy Systems and Soft Computing: Selected Contributions to the IMSC-2021", Kolomna, Russia, May 17-20, 2021 (CEUR Workshop Proc. – Vol. 2965. – P. 59-63.