



В.К. Финн, М.А. Михеенкова

**Точная эпистемология,
искусственный интеллект и
интеллектуальный анализ данных в
науках о жизни и социальном
поведении**

Рекомендуемая форма библиографической ссылки

Финн В.К. Михеенкова М.А. Точная эпистемология, искусственный интеллект и интеллектуальный анализ данных в науках о жизни и социальном поведении // Проектирование будущего. Проблемы цифровой реальности: труды 3-й Международной конференции (6-7 февраля 2020 г., Москва). — М.: ИПМ им. М.В.Келдыша, 2020. — С. 218-226. — <https://keldysh.ru/future/2020/19.pdf> <https://doi.org/10.20948/future-2020-19>

Размещено также [видео выступления](#)

Точная эпистемология, искусственный интеллект и интеллектуальный анализ данных в науках о жизни и социальном поведении

В.К. Финн, М.А. Михеенкова

РГГУ, ФИЦ ИУ РАН

Аннотация. В работе рассматриваются проблемы точной эпистемологии в предметных областях, характеризующихся слабо формализованными знаниями и открытыми множествами эмпирических данных. Описаны подходы к решению этих проблем средствами ДСМ–метода автоматизированной поддержки исследований. Реализующие ДСМ–метод интеллектуальные системы (ИС–ДСМ) являются инструментом интеллектуального анализа данных в науках о жизни и социальном поведении.

Ключевые слова: точная эпистемология, интеллектуальный анализ данных, конструктивная эвристика, ДСМ–метод, интеллектуальные системы

Exact epistemology, artificial intelligence and knowledge discovery in the life and social sciences

V.K. Finn, M.A. Mikheyenkova

RSUH, FRC CSC RAS

Abstract. The paper considers the problems of exact epistemology in subject areas characterized by poorly formalized knowledge and open sets of empirical data. The approaches to solving these problems by means of the JSM Method of automated research support are described. Intelligent systems that implement the JSM Method (IS–JSM) are a tool for knowledge discovery in the life and social sciences.

Keywords: exact epistemology, knowledge discovery, constructive heuristics, JSM Method, intelligent systems

Введение

Исследования в области искусственного интеллекта (ИИ) направлены на создание компьютерных систем, реализующих

конструктивную имитацию и усиление познавательной деятельности человека [1:256-277]. Это требует разработки принципов эпистемологии – формирования системы знаний, необходимых для решения задач с использованием компьютерных систем, – и создания на основе этих принципов средств конструктивной эвристики для такого решения [2]. Практическая реализация развиваемых теоретических принципов в интеллектуальных системах (ИИ–системах) означает автоматизацию процесса получения нового знания в результате интеллектуального анализа эмпирических данных.

Современное информационное общество характеризуется повсеместным распространением и быстрым развитием компьютерных систем в управлении, принятии решений, коммуникациях, производстве, военной сфере, что приводит к восприятию ИИ как комплекса технологических решений. В этих условиях невнимание к теоретическим (эпистемологическим) основаниям ИИ влечёт лишь совершенствование имеющихся технологий – без возможности порождения новых подходов и создания реализующих их инструментов для использования в практических целях.

Критически необходимым решение указанных проблем оказывается для тех областей, где обнаружение нового знания (knowledge discovery) объективно затруднено отсутствием развитого формального аппарата и ясно определяемой структуры познавательного процесса в сочетании с обширным эмпирическим материалом, допускающим возможность структурирования. Это характерно для гуманитарных областей как таковых и, в частности, для наук о жизни и социальном поведении.

Основы точной эпистемологии

Идейным основанием искусственного интеллекта, направленного на создание компьютерных интеллектуальных систем – инструментов автоматизации познавательного процесса, прежде всего, в плохо формализованных областях, – является **точная эпистемология (ТЭ)** [3].

Точная эпистемология представляет собой исследование взаимодействия *познающего субъекта* и соответствующего ему *объекта* посредством эвристик и логик рассуждения, порождающих новое знание и его *принятие*. Содержанием ТЭ и её средствами являются: организация и методология исследований; преобразование идей в понятия; амплиативные выводы; логика рассуждений; верификация и фальсификация результатов исследования; средства порождения гипотез; средства обнаружения эмпирических закономерностей; несингулярные шкалы оценки качества рассуждений и гипотез; эвристики получения нового знания из исходных данных; изучение и обобщение опыта экспериментального применения эвристик. Под эвристикой, согласно Д. Пойа [4], понимается организация правил и методов для совершения открытий – получения нового знания –

посредством правдоподобных рассуждений (прежде всего, индукции и аналогии) и выдвижения гипотез.

Взаимодействие *познающего субъекта* (S) и соответствующего *объекта* познания (O) реализуется в интеллектуальном процессе (ИП) – S–ИП–O, – представляющем взаимодействие мыслительного процесса (МП) и управляемого им познавательного процесса (ПП) [5]. Интеллектуальный процесс характеризуется множеством интеллектуальных способностей естественного интеллекта (см. [1]), часть из которых относится к ПП, другие же составляют МП, включающий также некоторые высшие психические функции, невыразимые в формальном языке.

Конструктивное определение субъекта S обеспечивает возможность автоматизированной поддержки интеллектуального процесса ИП для порождения нового знания относительно исследуемого объекта O. С точки зрения ИИ таким субъектом является интеллектуальная система ИС – компьютерная система, реализующая в автоматическом и интерактивном режиме интеллектуальный процесс ИП. Функциональные возможности ИС отображаются в её архитектуре, представленной реализующим познавательные процедуры Решателем задач, подсистемой накопления знаний (базой фактов БФ и базой знаний БЗ) и развитыми инструментами восприятия и отображения результатов (интеллектуальным интерфейсом). Подсистемами Решателя интеллектуальной системы являются Рассуждатель, обеспечивающий реализацию содержащих амплиативные выводы рассуждений, Вычислитель для проведения вычислительных процедур и Синтезатор, осуществляющий взаимодействие Рассуждателя с Вычислителем и выбор стратегии решения задач.

Формальное представление универсального познавательного процесса «анализ данных – предсказание – объяснение» обеспечивает имитацию некоторых способностей естественного (рационального) интеллекта (рассуждения, аргументации, обучения, объяснения полученных результатов) и допускает воспроизведение в интеллектуальных компьютерных системах в автоматическом режиме. Однако слабо формализуемый мыслительный процесс, включающий установки, императивы, целеполагание, формирование открытых эмпирических теорий и их адаптацию в условиях коррекции данных и знаний, требует участия человека и может быть реализован в таких системах лишь в интерактивном режиме.

Разработка принципов точной эпистемологии и формирование на основе этих принципов эвристических средств решения задач ИИ является основой создания и развития технологических инструментов для практических применений. Таким образом, под **искусственным интеллектом (ИИ)** следует понимать *компьютерные реализации средств точной эпистемологии – языков представления знаний и логик рассуждений*. Технологическими продуктами ИИ являются системы ИИ –

компьютерные системы, применяющие отдельные методы ИИ, интеллектуальные системы ИС описанной выше архитектуры, обеспечивающие автоматизацию интеллектуального процесса, и интеллектуальные роботы, включающими ИС наряду с механическим и сенсорным модулями. Автоматизация интеллектуального процесса, обеспечивающего исследование и применение эвристик, включающих рассуждение с амплиативными выводами, позволяют выделить сферы применимости ИИ. Ими являются области, для которых характерна слабая формализация знаний при наличии допускающих структурное представление эмпирических данных. Иными словами, речь идёт о социальной и гуманитарной сферах – науках о жизни и социальном поведении, а также их приложениях, включая медицину и управление.

ДСМ–метод автоматизированной поддержки исследований

Научным и конструктивно реализованным аппаратом точной эпистемологии является ДСМ–метод автоматизированной поддержки исследований (ДСМ–метод АПИ – по первым буквам имени Д.С. Милля) [6,7]. ДСМ–метод является *методологией* создания интеллектуальных систем и *средством формализации, имитации и усиления* интеллектуальных процессов. Метод воспроизводит исследовательские эвристики, используя правдоподобные рассуждения (с амплиативными выводами, следствия которых не содержатся непосредственно в посылках) в открытом мире в виде синтеза неэлементарных познавательных процедур: эмпирической индукции (анализ), структурной аналогии (предсказание), абдуктивного объяснительного принятия гипотез. ДСМ–рассуждения – взаимодействие указанных амплиативных познавательных процедур – порождают новое знание. Соответственно, в интеллектуальных системах типа ДСМ (ИС-ДСМ) реализуется интеллектуальный анализ данных (ИАД), прежде всего – в науках о жизни и социальном поведении.

Целью ИАД является поддержка использующих эмпирические данные исследований, позволяющая выявлять эмпирические закономерности (законы и тенденции) [5]. Обнаруженные эмпирические закономерности, содержащиеся в базах знаний ИС, используются для формирования эмпирических теорий и позволяют сформулировать усиленный критерий демаркации, отделяющий завершённое научное исследование от пред-исследования [8]. Интеллектуальный анализ данных включает спецификацию и моделирование предметной области и предполагает использование синтеза познавательных процедур. Извлечение эмпирических закономерностей осуществляется при последовательном расширении баз эмпирических фактов, результатом чего является расширение уже баз знаний интеллектуальной системы.

Индукция в ДСМ–методе представлена формальными уточнениями и расширениями индуктивных методов Д.С. Милля [9] – метода сходства,

различия, остатков, сходства различия и сопутствующих изменений. ДСМ–метод АПИ включает семь компонент: условия применимости (которые могут быть формализованы); ДСМ–рассуждения (синтез познавательных процедур: индукция, аналогия, абдукция); открытые квазиаксиоматические теории (КАТ), являющиеся средствами представления знаний; упорядоченные стратегии ДСМ–рассуждений (дистрибутивные решетки индуктивных процедур); металогические средства исследований предметной области; эвристика обнаружения эмпирических закономерностей; ИС-ДСМ

ДСМ–метод оказывается эффективным инструментом ИАД при выполнении ряда существенных онтологических допущений относительно природы исследуемого «мира» $W^{(\pm)}$, в котором рассматриваемые явления должны носить детерминистский характер, а данные и знания могут пополняться. Должно быть формализуемо структурное – не метрическое и нестатистическое – сходство фактов, которое рассматривается как источник детерминаций. База фактов представляется множествами позитивных (+)- и негативных (-)-примеров изучаемых эффектов, включающих, соответственно, (+)- и (-)-причины (существенные влияния) проявления эффектов. Это позволяет автоматически порождать фальсификаторы индуктивных гипотез и может рассматриваться как основание для их абдуктивного принятия.

Первый (начальный) этап формализованной эвристики ДСМ–метода АПИ – применение ДСМ–рассуждений к начальной базе фактов БФ(0). Он включает индуктивное порождение гипотез о причинах, предсказание (с использованием этих гипотез) заранее не заданных эффектов у объектов исходной БФ с помощью вывода по аналогии и абдуктивное принятие порождённых гипотез. Целью второго этапа полноценного ДСМ–исследования является обнаружение средствами ДСМ–рассуждений эмпирических закономерностей ER – эмпирических законов (EL) и эмпирических тенденций (ET и WET, WET – слабые ET), $ER = EL \cup ET \cup WET$.

Эмпирические закономерности ER представляют собой регулярности сохранения соответствия причины эффекта и его предсказания в последовательностях расширяемых баз фактов $БФ(0) \subset \dots \subset БФ(s)$ и их вариациях. Возможная комбинаторика расширений БФ есть рассмотрение $(s+1)!$ вариаций БФ(0), $БФ(1) = БФ(0) \cup V_1, \dots, БФ(s) = БФ(0) \cup V_1 \cup \dots \cup V_s$, где $V_i \cap V_j = \emptyset, V_i \cap БФ(0) = \emptyset; i, j = 1, \dots, s$. Семантически обнаружение эмпирических законов EL означает наследование типов истинностных значений $v = 1$ и/или $v = -1$ во всех вариациях у индуктивных гипотез о (+)- и/или (-)-причинах и гипотез-предсказаний, полученных в выводе по аналогии с использованием причинных гипотез. Эмпирические тенденции обнаруживаются при сохранении типов истинностных значений $v \in \{1, -1\}$

лишь после q расширений: $\underbrace{T \dots T}_q \underbrace{V \dots V}_{s+1-q}$, ET – при условии $2q < s+1$, WET – при $2q \geq s+1$.

Эмпирические закономерности являются индуктивными операционально определяемыми (нестатистическими) обобщениями результатов синтеза познавательных процедур – индукции, аналогии и абдукции. Фальсификация гипотез с помощью условий, встроенных в реализующие ДСМ–рассуждения правила и предикаты, а также результатов исследования и обнаружение эмпирических закономерностей формируют усиленный критерий демаркации, характеризующий завершённое исследование.

4. Применение ДСМ–метода для ИАД в медицине социологии

Применимость ДСМ–метода для ИАД в медицине и социологии обусловлена сходными чертами этих областей: наличием обширного эмпирического материала в сочетании со слабой формализованностью знаний при потенциальной возможности структурирования данных (знаний). ДСМ–исследования в медицине имеют обширную библиографию (см., например, [1: 89-124; 5:387-407; 6:398-430]). Наиболее интересные результаты последнего времени – прогнозирование возникновения сахарного диабета и возникновения рака поджелудочной железы у больных хроническим панкреатитом [10]; предсказание ремиссии или рецидива после соответствующих видов лечения больных с опухолью меланомы (с использованием генетических данных), а также распознавание мутаций, ответственных за возникновение меланомы [11].

Наличие клинических данных в медицине позволяет говорить о некотором приближении к объективизации, чего лишена социальная сфера с доминированием носящих субъективный характер данных. Заметим, что распространённые здесь метрические подходы при отсутствии сформированной системы отношений также далеки от объективности. Имеющие дело с обезличенным субъектом и преимущественно количественными данными статистические методы не позволяют решать задачу изучения взаимодействий индивидов. При этом качественные методы, использующие формальные инструменты, чрезвычайно далеки от идеала объективности.

Задачу воспроизведения в компьютерных системах познавательной деятельности не решают ни инструменты CAQDAS (computer aided/assisted qualitative data analysis) [12], ни ориентированные на проблемы классификации/кластеризации методы data mining, ни достаточно распространённый в социологической практике качественный сравнительный анализ, Qualitative Comparative Analysis (QCA) [13], использующий инструмент булевой алгебры.

ДСМ–метод автоматизированной поддержки исследований осуществляет формальное воспроизведение и реализацию в интеллектуальных системах исследовательских эвристик действующих социологов. Это позволяет решать задачи качественного (неколичественного) анализа социологических данных (с распознаванием эмпирических закономерностей) средствами ИАД [5:409-491]. К числу таких задач относятся: исследование индивидуального поведения и обнаружение его детерминант с последующей типологизацией социума на их основе, учет влияния ситуации (контекста поведения), анализ и прогнозирование мнений.

Для использования ДСМ–рассуждений субъект социального взаимодействия задаётся термом \bar{X} (полный объект), $\bar{X} = \langle X, S, [\varphi] \rangle$. Здесь X – структурированное описание действующего субъекта, S – контекст (ситуация) поведения, φ – мнение субъекта (о ситуации, о возможном поведении и т.п.). Специфический вид поведения – мнение, личное восприятия индивидуумом различных сторон социальной действительности – формируется на основе оценки утверждений p_1, \dots, p_n , характеризующих ситуацию взаимодействия и аргументируют отношение к ней [14]. Мнение φ представляется в виде максимальной конъюнкции специально созданной многозначной логики JA_4 с аргументационной семантикой [14:312–338], $[\varphi]$ – множество элементов соответствующей максимальной конъюнкции. Такое представление позволяет решать отдельную проблему анализа рациональности мнений, рассматриваемой как аргументированное принятие решений, в том числе – оценивать степень рациональности социальной общности.

Сложная многопараметрическая структура социальных систем и разнообразные механизмы социальных взаимодействий требуют эпистемологически адекватного языка представления данных (в частности, их параметризации), выбора эффективных процедур и стратегий анализа, направленного формирования и пополнения массивов эмпирических фактов. В общем случае исходные данные представлены (+)-фактами $B\Phi^+ = \{ \langle \bar{X}, Y \rangle \mid J_{(1, 0)}(\bar{X} \Rightarrow_1 Y) \}$ («объект \bar{X} обладает множеством свойств Y »), (-)-фактами $B\Phi^- = \{ \langle \bar{X}, Y \rangle \mid J_{(-1, 0)}(\bar{X} \Rightarrow_1 Y) \}$ и фактами, описывающими объекты с заранее не заданными свойствами, $B\Phi^c = \{ \langle \bar{X}, Y \rangle \mid J_{(\tau, 0)}(\bar{X} \Rightarrow_1 Y) \}$. $B\Phi = B\Phi^+ \cup B\Phi^- \cup B\Phi^c$. Это позволяет варьировать реляционную структуру в зависимости от содержательной социологической модели [15;16].

Применение ДСМ–метода в социологии позволяет уточнить круг задач когнитивной социологии. Прежде всего, сюда должно быть включено изучение собственно познавательной деятельности социолога – познающего субъекта S – с последующей её имитацией и усилением средствами ИИ. Адекватная параметризация социологических фактов и логическое представление мнений позволяет решить и другую задачу

когнитивной социологии: анализ субъективных когнитивных эффектов, в частности, рациональности мнений и влияния познавательной активности на поведение на уровне социального взаимодействия.

5. Заключение

В заключение отметим, что распространившийся термин «цифровизация» (в том числе, «цифровая экономика») характеризует лишь цивилизационный (технологический) аспект развития, тогда как первичной составляющей современного информационного общества является явление культуры – интеллектуализация, проникающая в управление, коммуникации, медицину, научные исследования. Точная эпистемология является теоретическим основанием интеллектуализации, а её средствами – технологии ИИ системы ИИ, интеллектуальные системы, ИИ–роботы. Плодами интеллектуализации являются когнитивные исследования, в том числе – когнитивная социология, когнитивная психология, когнитивная лингвистика, evidence based medicine, интеллектуальное управление, когнитивная история и т.п.

ДСМ–метод автоматизированной поддержки исследований обеспечивает решение проблемы создания эпистемологических оснований и эвристических средств решения задач искусственного интеллекта. Интеллектуальные системы ИС-ДСМ представляют собой технологическое средство точной эпистемологии и являются партнёрскими человеко-машинными системами, эффективно реализующими генерацию нового знания, но при этом не заменяющими, а поддерживающими и усиливающими содержательную работу исследователя в различных предметных областях – от медицины до наук о человеке и обществе.

Работа выполнена при поддержке Российского фонда фундаментальных исследований, проект 18-29-03063-мк.

Литература

1. *Финн В.К.* Искусственный интеллект: методология, применения, философия. – М.: URSS, 2018. – 448 с.
2. *McCarthy J.* Epistemological problems of artificial intelligence // Fifth International Joint Conference on Artificial Intelligence IJCAI-77 (August 22 – 25, 1977, Cambridge, Massachusetts, USA). – Proceedings. – Vol. 2. – Pp. 1038-1044.
3. *Финн В.К.* Интеллект, информационное общество, гуманитарное знание и образование. – Москва: Издательство РГГУ, 2020 (в печати).
4. *Пойа Д.* Математика и правдоподобные рассуждения (под ред. С.А. Яновской). – М.: «Наука», Главная редакция физико-математической литературы, 1975 (изд. 2-е). – 464 с.

5. *Финн В.К.* Об эвристиках ДСМ–исследований (дополнения к статьям) // Научно-техническая информация. Серия 2. Информационные процессы и системы. – 2019. – №10, с. 1-34.
6. Автоматическое порождение гипотез в интеллектуальных системах. Под общ. ред. В.К. Финна. – М.: Книжный дом «ЛИБРОКОМ», 2020 (изд. 2-е стереотип.). – 528 с.
7. ДСМ–метод автоматического порождения гипотез: Логические и эпистемологические основания. Под общ. ред. О.М. Аншакова. – М.: Книжный дом «ЛИБРОКОМ», 2009. – 432 с.
8. *Финн В.К.* О классе ДСМ–рассуждений, использующих изоморфизм правил индуктивного вывода // Искусственный интеллект и принятие решений. – 2016. – № 3. – С. 48 – 61.
9. *Милль Дж.Ст.* Система логики силлогистической и индуктивной. Издание 5-ое. – М.: ЛЕНАНД, 2011. – 832 с.
10. *Шестерникова О.П., Финн В.К., Винокурова Л.В., Лесько К.А., Варварина Г.Г., Тюляева Е.Ю.* Интеллектуальная система диагностики заболеваний поджелудочной железы // Научно-техническая информация. Серия 2. – 2019. – №10. – С. 41-48.
11. *Чеванов Д.К., Михайлова И.Н.* Интеллектуальный анализ данных пациентов с меланомой для поиска маркеров заболевания и значимых генов // Научно-техническая информация. Серия 2. – 2019. – №10. – С. 35-40.
12. *Silver C., Lewins A.* Using Software in Qualitative Research: A Step by Step Guide (2nd ed.). – London: SAGE Publication, 2014. – 304 p.
13. *Marx A., Rihoux B., Ragin C.C.* The origins, development, and application of qualitative comparative analysis: The first 25 years // European Political Science Review. – 2014. – Vol. 6, No 1. – Pp. 115-142.
14. *Finn V.K., Mikheyenkova M.A.* Plausible reasoning for the problems of cognitive sociology // Logic and Logical Philosophy. – 2011. – Vol. 20. – Pp. 111-137.
15. *Климова С.Г., Михеенкова М.А.* Возможности ДСМ–метода для построения социологических гипотез (на примере анализа политического участия) // Социологический журнал. – 2017. – Том 23. – № 3. – С. 80-101.
16. *Михеенкова М.А., Климова С.Г.* Интеллектуальный анализ данных в социологических исследованиях // Научно-техническая информация. Серия 2. Информационные процессы и системы. – 2018. – № 12. – С. 12-24.