

Проблема обоснования в формальном представлении знаний

Аннотация:

Понятие знание, которое используется в компьютерной науке, не включает в себя обычное для философской эпистемологии требование обоснования знания. Такое положение вещей имеет глубокие корни в истории философской логики в 20-м веке и приводит к негативным практическим последствиям. Исследования последних лет в области теоретико-доказательной семантики, гомотопической теории типов и формальной эпистемологии позволяют наметить теоретическую основу для решения этой проблемы.

Ключевые слова: представление знаний, обоснование, доказательство, теоретико-доказательная семантика, гомотопическая теория типов

1) Понятие знания в философской эпистемологии и в компьютерной науке

В центре современной эпистемологической дискуссии (по крайней мере в Аналитической философской традиции) находится восходящая к Платону так называемая JTB концепция знания, согласно которой знание - это обоснованное истинное мнение (Justified True Belief). Это определение расшифровывается так: эпистемический агент A знает, что p (где p это некоторое высказывание), если и только если выполнены следующие три условия:

- пропозиция p истинна (является истинным высказыванием)
- агент A верит, что p
- A имеет достаточное основание g , для того, чтобы верить, что p .

Если речь идет об общем знании, то агент A может специально не указываться: в этом случае говорят, что “ p известно”.

Заметим, что в рамках JTB-теории предполагается, что предмет знания в общем случае - это пропозиция. Такое знание называют пропозициональным знанием или знанием-что (knowledge that).

В связи с JTB-теорией широко обсуждаются так называемые “проблемы типа Гетье” [1], которые показывают, что понятие обоснования нуждается в дальнейшем уточнении. В частности, необходимы критерии, по которым можно было бы оценивать предложенное обоснование данного утверждения и признавать или не признавать его достаточным. Предпосылка JTB-теории о том, что всякое знание является пропозициональным, также широко обсуждается и оспаривается многими философами. Оставляя в стороне примеры другого рода, заметим что согласно JTB-теории, знание теоремы Пифагора требует умения

доказывать эту теорему, а не только веры в то, что утверждение этой теоремы истинно. Однако знание доказательства теоремы (в смысле умения ее доказывать) - это уже не пропозициональное, а процедурное знание (знание-как). Таким образом можно аргументировать, что понятие обоснования, которое используется в ЛТВ-теории, по крайней мере в некоторых важных случаях требует включения в рассмотрение понятия процедурного знания.

Компьютерная наука в отличие от философской эпистемологии не ставит своей целью построить общую теорию знаний и рассуждений. Тем не менее в некоторых монографиях по компьютерному представлению знаний делаются попытки неформально определить эти базовые понятия [2], [3], [4]. Ни одно из таких определений, из числа тех, которые нам удалось найти в компьютерной литературе, не включает в себя требование ЛТВ-теории о том, что знание должно включать в себя обоснование. Хотя представители компьютерных наук в последние годы много говорят и пишут о знаниях, как правило они не проводят никакого различия между знанием высказывания p и самим этим высказыванием.

Вместе с тем проблема обоснования знаний не является чисто теоретической и тем более чисто философской. Ненадежность распространяемой и получаемой через электронные средства коммуникации информации в последние годы стала социальной и даже политической проблемой. Некоторые философы и публицисты говорят в этой связи о наступлении “эпохи пост-правды” [5]. На наш взгляд, мы в данном случае имеем дело с недостатком существующих компьютерных и коммуникационных технологий, который может и должен быть исправлен. В этой статье мы не предлагаем технических решений, но анализируем истоки проблемы и показываем, что она имеет фундаментальный логический аспект. Далее мы описываем некоторые новые подходы в логической семантике, которые, на наш взгляд, могут послужить теоретической основой для будущих технических решений.

2) Логическое следование и логический вывод

В своей последней статье с говорящим названием “Пренебрежение эпистемологическими соображениями в логике” Горан Зундолм [6] показывает, что в мейнстриме развития философской логики 20-го века эпистемологические соображения систематически игнорировались. Опираясь на анализ Зундольма, мы показываем, что “пренебрежение” проблемой обоснования в теории и практике компьютерного представления знаний связано с тем, что создатели концепции искусственного интеллекта и их последователи до сих пор ориентировались на системы классической логики и логической семантики, в которых эпистемологические соображения являются второстепенными по сравнению с онтологическими соображениями.

Отождествление рассуждения с синтаксической процедурой будет бессодержательным, если не объяснить, как и почему данная синтаксическая процедура символическим образом выражает некоторое рассуждение. Семантическая теория, которая отвечает на этот вопрос, была предложена Альфредом Тарским в классической работе “О понятии логического следования”

1936-го года [7]. Тарский определяет это отношение так:

Формула B логически следует из совокупности формул A_1, \dots, A_n , если всякая интерпретация теории, при которой формулы A_1, \dots, A_n интерпретируются как истинные высказывания, также интерпретирует формулу B как истинное высказывание.

Тарский использует понятие логического следования в качестве эрзаца понятия логического вывода, не определяя это последнее понятие формально. Стандартные логические исчисления обладают свойством *корректности* (soundness): если формула B выводится (то есть порождается в согласии с синтаксическими правилами дедукции) из формул A_1, \dots, A_n , то B логически следует из и формул A_1, \dots, A_n в смысле приведенного выше определения. Такое свойство логического исчисления позволяет думать о синтаксическом выводе (то есть “механической” дедукции) как о логическом выводе, то есть процедуре, которая позволяет устанавливать логические следствия из данных предпосылок. Такую семантику логического вывода называют *теоретико-модельной*, чтобы отличить ее от *теоретико-доказательной* семантики, о которой мы скажем ниже.

Теоретико-модельная семантика логического вывода не несет никакой эпистемологической нагрузки. Отношение логического следования по Тарскому - это отношение между формулами, которое формулируется в терминах истинных предложений и их классов. Классическое понятие истины, которое здесь используется, предполагает, что истинность предложения никак не зависит от того, знает ли какой-либо эпистемический агент это предложение или нет. Таким образом данная логико-семантическая схема рассуждений не имеет никакого эпистемологического содержания.

Могут возразить, что в нашем кратком описании классической логической семантики мы просто забыли упомянуть о таком важном эпистемологически нагруженном понятии, как *доказательство*. Этим словом Гильберт и Бернайс [8] стали называть синтаксические выводы, то есть цепочки формул, которые включают в себя ряд предпосылок A_1, \dots, A_n (которые могут иметь статус *аксиом* теории) из которых в соответствии с правилами дедукции данной теории выводится некоторая искомая формула B : такую цепочку вслед за Гильбертом и Бернайсом часто называют “доказательством формулы B на основании предпосылок (аксиом) A_1, \dots, A_n ”. Однако, как убедительно аргументирует Правиц [9], такое понятие доказательства является чисто техническим и также не имеет эпистемологического содержания. Чтобы формальные доказательства выполняли эпистемические функции и работали как доказательства в обычном смысле слова, на них необходимо накладывать дополнительные требования. Сами по себе эти синтаксические конструкции позволяют лишь определенным образом структурировать классы высказываний.

Такая структуризация совершенно необходима при компьютерном представлении знаний в форме теорий, поскольку любой компьютер обладает конечными ресурсами памяти и в принципе не способен хранить в явной форме все предложения теории [10]. За уменьшение

расхода памяти приходится платить увеличением времени, требуемого для проверки истинности того или иного утверждения. Достижение баланса между расходом памяти и скоростью обработки относится к числу классических проблем теории алгоритмов и практики программирования, однако постановка и решение этой технической проблемы по-прежнему никак не учитывает эпистемическую функцию логического вывода.

3) Формальные онтологии

Теоретику-модельную логическую семантику в духе Тарского можно простым образом связать с онтологией, то есть теорией о том, какого рода сущности существуют (“бывают”) в мире вообще или по крайней мере в некоторой интересующей нас части мира. Для этого достаточно принять следующий тезис:

Для всякого истинного высказывания *существует* некоторая вещь (или вещи), которая делает это высказывание истинным.

Такую вещь по отношению к данному высказыванию в современной философской логике и аналитической метафизике принято называть *фактором истины* (truth-maker) этого высказывания, а приведенный выше тезис - тезисом *реализма факторов истины* (truth-maker realism, TMR).

TMR не решает вопроса о том, каков фактор или факторы истины каждого конкретного истинного предложения. Однако если речь идет о первопорядковой теории, то простейшим ответом на вопрос о факторах истины предложений этой теории будет онтологическое допущение о существовании индивидов (с их свойствами и отношениями), с помощью которых данная теория интерпретируется. Именно такой подход используется в *формальной онтологии*, которая в настоящее время является не только философской, но и компьютерной дисциплиной, имеющей приложения в искусственном интеллекте вообще и в первую очередь в области компьютерного представления знаний. Основным вариантом использования онтологии в инженерии компьютерных систем является проектирование структур информационного обеспечения [11]: баз данных, протоколов межпрограммного взаимодействия, форм пользовательского интерфейса и отчетов. Языки описания формальных онтологий предусматривают возможность задавать нетривиальные аксиомы и выводить из них утверждения, не записанные в структуре в явном виде.

Использование философских подходов наработок в теории и практике компьютерного представления знаний можно только приветствовать, но в данном случае вызывает недоумение тот факт, что для решения задач представления знаний компьютерная наука заимствует из философии именно онтологические, но не эпистемологические подходы.

4) Формальная эпистемология и конструктивная логика

В рамках описанной выше стандартной логической архитектуры компьютерного представления знаний истинностные значения приписываются высказываниям “с точки зрения Всезнающего

Высшего Существа” без учета того, каким образом то или иное истинностное значение стало известным. Как нужно изменить эту стандартную архитектуру чтобы учесть эпистемологическую сторону дела? Для решения этой задачи лучше подходит иная концепция логики, которую сегодня по историческим причинам принято называть *конструктивной*. Истина (предложения) понимается в конструктивной логике как существование доказательства (этого предложения), но при этом понятие доказательства понимается более широко, чем в классическом случае; в частности, доказательствами (или, как принято говорить в конструктивной логике, *свидетельствами*) при таком подходе могут считаться сами факторы истины. В последние годы были также построены новые семантики логического вывода, которые сегодня называют *теоретико-доказательными* [12]. Идея такой семантики состоит в том, чтобы синтаксическим процедурам и правилам вывода одних формул из других поставить в соответствие подробно эксплицированные эпистемологические процедуры. Сохранение истинности при всех возможных интерпретациях по прежнему мыслится как необходимое, но уже не как достаточное требование для того, чтобы данная синтаксическая процедура представляла логический вывод. С этим связано введение дополнительных (по сравнению со случаем теоретико-модельной семантики) синтаксических ограничений.

Представим себе информационную систему, которая не просто выдает пользователю в удобном виде некоторую полезную информацию по его запросу, осуществляя при этом поиск и обработку данных, но и предоставляет пользователю дополнительно соответствующее формальное *обоснование* в виде доступного для человеческого понимания и анализа описания выполненной процедуры. С помощью такого описания пользователь сможет (1) отождествить первоначальный источник нужной ему информации и (2) увидеть, по крайней мере в общих чертах, как именно информационная система обрабатывала исходные данные, чтобы ответить на его запрос [13]. Теоретико-доказательная семантика выводов может быть использована в такой ситуации для спецификации требований, при соблюдении которых предоставляемые компьютерной системой доказательства надежности информации будут валидными.

5) Гомотопическая теория типов как новая теоретическая модель для компьютерного представления знаний

Говоря о формальных выводах, мы до сих пор подразумевали, что формулы, представляющие посылки и заключения таких выводов, интерпретируются как высказывания. Сейчас мы кратко опишем формальную систему ГТТ (гомотопическая теория типов), которая имеет более широкие семантические возможности. Синтаксис ГТТ в первоначальном варианте этой теории совпадает с синтаксисом теории типов Мартина-Лефа (ТТМЛ) [14]. Базовые формулы этого исчисления имеют вид $a : A$, где A называется *типом*, а a - *термом* данного типа. Сами такие формулы называются *суждениями* и, согласно первоначальной идее Мартина-Лефа, всегда допускают любую из следующих неформальных интерпретаций:

- a является доказательством пропозиции A

- a является элементом множества A
- a является решением задачи A
- a является реализацией намерения A

Существенный прогресс в исследовании ТТМЛ, который заставил пересмотреть некоторые ключевые аспекты семантики этой теории, был достигнут в течении последнего десятилетия после того, как Владимир Воеводский установил связь между ТТМЛ и геометрической теорией гомотопий [15]. ТТМЛ и аналогичные теории с новой гомотопической семантикой сегодня принято называть *гомотопической теорией типов* (ГТТ). В рамках гомотопической семантики типы интерпретируются как абстрактные пространства (для которых определены основные понятия теории гомотопий), а термы - как точки таких пространств. ГТТ позволила установить, что “не все типы одинаковы”: только типы специального вида (гомотопического уровня), а именно типы содержащие самое большое единственный терм, можно отождествить с высказываниями; типы другого специального вида отождествляются с множествами; типы более высоких уровней мы оставим в стороне. Таким образом, первоначальная идея о том, что всякий тип допускает интерпретации как высказывание, множество и т.д. в контексте ГТТ не выглядит убедительной. Однако по всякому данному типу высшего порядка можно каноническим образом построить соответствующее высказывание, искусственно отождествляя все термы данного высшего типа. Такую процедуру в ГТТ называют *обрезанием* (truncation).

Покажем теперь, каким образом ГТТ может быть использована в качестве теоретической модели для систем представления знаний. Синтаксис ГТТ представляет собой систему правил, которые по отношению к пропозициональным типам (высказываниям) применяются как логические правила (правила вывода), а по отношению к типам более высоких гомотопических уровней - как правила построений непропозициональных конструкций, которые используются для верификации соответствующих высказываний. В настоящее время такой подход уже успешно используется для формализации и компьютерной проверки математических доказательств. Мы предполагаем, что такой подход может быть использован и за пределами чистой математики. В задачах представления технологических знаний предлагаемый подход открывает возможность не только описывать технологические процедуры с помощью программного кода, но и моделировать на компьютере процесс изготовления изделия и проверять на этой модели, будет ли спроектированное изделие обладать требуемыми характеристиками.

6) Заключение

В последние годы в компьютерном представлении знаний широко используются подходы анализа больших данных, машинного обучения и искусственных нейронных сетей. Существует даже мнение, что классические средства представления знания, основанные на явном символьном выражении фактов и законов, безнадежно устарели и скоро будут заменены нейронными сетями, хранящими знания в неявном распределенном виде. Однако, распределенное знание ненадежно и трудноверифицируемо, легко допускает ввод в

заблуждение [16]. Поэтому, на наш взгляд, нейронные сети и другие средства анализа больших данных могут значительно обогатить, но никоим образом не заменить собой логические подходы и инструменты в компьютерном представлении знания.

Для успешного использования логических подходов в компьютерном представлении знаний недостаточно пользоваться классической логической архитектурой и связанными с этой архитектурой онтологиями, которые, как показывает наш анализ, полностью лишены всякого эпистемологического содержания. Хотя методы формальной эпистемологии пока еще недостаточно развиты, чтобы уже сегодня найти применения в компьютерной науке и технологии, совместная работа в этой области логически ориентированных философов и представителей компьютерных наук открывает многообещающие перспективы как в теоретическом, так и в практическом плане.

Литература:

1. *Hetherington, S.*, Gettier Problems, Internet Encyclopedia of Philosophy, <https://www.iep.utm.edu/gettier/>
2. *Jakus, G., et al.*, Concepts, Ontologies and Knowledge Representation, Springer 2013, 67 P.
3. *Abraham, A., Grosan, C.*, Intelligent Systems : A Modern Approach, Springer 2011, 450 P.
4. *Lakemeyer, G., Nebel, B.* (Eds.) , Foundations of Knowledge Representation and Reasoning, Springer 1994, 355 P.
5. *Fuller, S.*, Post-Truth: Knowledge as a Power Game, CUP, 2018, 207 P.
6. *Sundholm, G.*, The Neglect of Epistemic Considerations in Logic: the Case of Epistemic Assumptions, forthcoming in Topoi
7. *Tarski, A.*, On the Concept of Logical Consequence, in: Logic, Semantics, Metamathematics, Hackett Publ., 1983, P. 409-420
8. *Гильберт, Д., Бернайс, П.*, Основания математики, т. 2 Теория доказательств, М. Наука 1979, 557 стр.
9. *Prawitz, D.*, On the Idea of the General Proof Theory, Synthese, 27 (1974), n. 1-2, P. 63-77
10. *Muggleton S., De Raedt L.*, Inductive Logic Programming: Theory and Methods // The Journal of Logic Programming. 1994. Vol. 19-20, P. 629-679
11. *Ковалёв, С.П.*, Применение онтологий при разработке распределенных автоматизированных информационно-измерительных систем // Автометрия. 2008. Т. 44, №2. с. 41–49
12. *Piesha, Th., and Schroeder-Heister, P.* (Eds.), Advances in Proof-Theoretic Semantics, Springer 2015
13. *Васильев, С.Н. и др.*, Интеллектуальное управление динамическими системами. М.: ФИЗМАТЛИТ, 2000. 352 с.
14. *Martin-Lof, P.*, Intuitionistic Type Theory, BIBLIOPOLIS, 1984, 91 P.
15. *Univalent Foundations Program*, Homotopy Type Theory, IAS Princeton, 2013, 473 P.
16. *Szegedy, C. et al.*, Intriguing Properties of Neural Networks. CoRR abs/1312.6199. 2013.

