

МАТЕМАТИКА

MATHEMATICS

УДК 004.055

doi: 10.25206/2311-4908-2020-7-2-4-9

ЗАДАЧНЫЙ ПОДХОД В ИСКУССТВЕННОМ ИНТЕЛЛЕКТЕ

© С. С. Гончаров¹, Е. Е. Витяев², Д. И. Свириденко³

^{1,2,3}Институт математики им. С. Л. Соболева СО РАН, Новосибирск, Россия

^{1,2,3}Новосибирский государственный университет, Новосибирск, Россия

¹S.S.Goncharov@math.nsc.ru

²evgenii.vityaev@math.nsc.ru

³dsviridenko47@gmail.com

Аннотация. В работе рассматривается задачный подход к искусственному интеллекту. Показывается что, с одной стороны, он обобщает такие интеграционные подходы как агентный и общий искусственный интеллект, а, с другой стороны, достаточно точно отражает когнитивные процессы и целенаправленную деятельность, описанную в физиологической теории функциональных систем работы мозга.

Ключевые слова: задача, моделирование, искусственный интеллект, когнитивная архитектура.

PROBLEM-BASED APPROACH IN ARTIFICIAL INTELLIGENCE

© S. S. Goncharov¹, D. I. Sviridenko², E. E. Vityaev³

^{1,2,3}S. L. Sobolev Institute of Mathematics SB RAS, Novosibirsk, Russia

^{1,2,3}Novosibirsk state university, Novosibirsk, Russia

¹S.S.Goncharov@math.nsc.ru

²evgenii.vityaev@math.nsc.ru

³dsviridenko47@gmail.com

Abstract. The paper considers the task approach to artificial intelligence. It is shown that, on the one hand, it generalizes such approaches as the agent-based approach and general artificial intelligence, and, on the other hand, accurately reflects the cognitive processes and purposeful behavior described in the physiological theory of functional brain systems.

Keywords: task, modeling, Artificial Intelligence, cognitive architecture.

1. Введение

В классическом современном искусственном интеллекте можно выделить четыре основных направления его развития – глубокое обучение, агентный подход, объяснимый (XAI) и общий (AGI) искусственный интеллект. Охарактеризуем кратко эти подходы.

1.1 Глубокое обучение

Эмпирическую основу глубокого обучения составляют нейронные сети. При всех достоинствах нейронных сетей им присущи и серьезные недостатки. Во-первых, нейронные сети являются «черным ящиком» и не дают возможность объяснить причины принятия тех или иных решений при их использовании. Это существенно ограничивает и в отдельных случаях делает даже невозможным применение нейросетей в таких областях как медицина,

финансы, военные приложения, логистика, где цена ошибки либо слишком высока, либо объяснение решения необходимо по юридическим причинам. Например, отказ нейронной сетью в выдаче кредита или рекомендация в необходимости совершения опасной хирургической операции должно быть аргументировано юридически.

Во-вторых, нейронные сети обладают слабой способностью к генерализации. Например, нейросеть, натренированная распознавать слонов и китов, в случае предъявления кита, выброшенного на берег, будет видеть слона, а купающегося в прибое слона будет видеть как кита.

В-третьих, нейросети запоминают отдельные, зачастую случайные детали предъявленных в ходе обучения образцов и принимают дальнейшие решения на основе именно этих деталей, а не на основе полноценного обобщенного предмета. Например, внесение в изображение незначительного шума или замена изображения шумом может приводить к распознаванию несуществующего предмета, а замена только одного пиксела в изображении – к распознаванию предмета, отличного от предъявленного.

1.2. Агентный подход

Этот подход, подробно изложенный в монографии [1], выступает в определенном смысле как альтернативный подход к нейросетевому глубокому обучению, поскольку его явным достоинством является отчетливый интеграционный характер. Данное обстоятельство является следствием удачно выбранной онтологической базы, где в качестве основных понятий предлагаются понятия «рациональный агент» и «внешняя среда», а различные задачи искусственного интеллекта рассматриваются как задачи взаимодействия «рационального агента» с «внешней средой». В работе [1] проведена детальная классификация агентов и сред, а также решаемых агентами задач. Обратим внимание на последнее замечание, поскольку оно во-многом объясняет последующий важный вывод о том, что агентный подход фактически следует задачному подходу, но только без явного определения понятия «задача» и всех присущих этому понятию компонент.

1.3. Объяснимый искусственный интеллект (eXplainable Artificial Intelligence - XAI)

Данный подход к искусственному интеллекту требует от методов, используемых при решении задач, обязательного присутствия компонент, помогающих понять, как и почему (например, в глубоком нейросетевом обучении), принимаются те или иные решения. Одними из основных методов XAI являются следующие:

- метод, показывающий вклад каждого признака в полученный прогноз; при этом в сопутствующем прогнозу объяснении показывается прогнозируемый класс вместе, например, с пикселями наиболее сильно повлиявшими на результат прогноза;
- метод, выделяющий признаки, специфичные для данного прогноза; заметим, что для того, чтобы выделить признаки, влияющие на прогноз объектов некоторого класса, необходимо объединить признаки по объектам класса.

При всех достоинствах данного подхода, указанные методы, тем не менее, не снимают проблему «черного ящика», свойственную нейронным сетям.

1.4. Общий искусственный интеллект (Artificial General Intelligence – AGI)

Глубокое обучение показало, что моделирование когнитивных процессов человека не обязательно является условием решения когнитивных (интеллектуальных) задач той предметной области, к которой относится данная задача. Соответственно, возникает понятие «Общего Искусственного Интеллекта» (ОИИ или AGI - Artificial General Intelligence), которым, естественно в той или иной степени, потенциально может обладать, как человек или живой организм с высокоразвитой центральной нервной системой, так и абстрактная робототехническая система. Ведущие разработчики ОИИ (Бен Гёрцел, Шейн Легге, Пей Ванг) дают следующее определение ОИИ: «Это способность решать когнитивные задачи в целом,

действуя целенаправленно, адаптируясь к условиям среды через обучение, минимизируя риски и оптимизируя потери на достижение своих целей». В настоящее время ОИИ, представляя, согласно определению, весьма многообещающее и перспективное направление, носит пока, в основном, чисто методологический и теоретический характер. Тем не менее, отметим, что ОИИ, определенный выше, носит явный интеграционный характер и во многом созвучен агентному подходу. Но и как агентному подходу ОИИ свойственно отсутствие точного определения понятия задачи, что, на взгляд авторов, и является одной из причин слабых практических достижений.

Цель настоящей статьи - показать, что предлагаемый нами задачный подход к искусственному интеллекту способен не только вместить в себя все указанные выше направления, удовлетворяя их определениям, но и, кроме того, достаточно точно моделировать целенаправленную деятельность человека в соответствии с физиологической теорией функциональных систем (ТФС) работы мозга [2].

2. Задачный подход

Начнем изложение задачного подхода с наиболее точного описания задачи, возникающего в основаниях математики [3].

2.1. Понятие «задача» в основаниях математики [3]

Анализ понятия задачи начинается со следующих простых рассуждений, сформулированных д.ф.н. К.Ф. Самохваловым: «Я хочу пить» – что это значит? Нет, конечно, никакой ошибки полагать, что слова «я хочу пить» означают просто вот это, где это – определенное состояние сознания, которое я переживаю сейчас и которое я именую жаждой. Но тогда возникает новый вопрос: как ощущение жажды (хотения) связано с фактическим питьем (удовлетворением хотения)? Откуда я знаю, что удовлетворить жажду можно питьем? Сохранится ли в самом переживании жажды сознание того, чем эту жажду можно удовлетворить? Знать желание не означает знать желаемое, а означает способность узнать желаемое, т.е. иметь критерий удовлетворения желания.

Таким образом, задача определена (осмыслена) тогда и только тогда, когда у нас есть критерий решенности задачи – критерий проверки действительно ли предъявленное решение является решением задачи. В математических теориях таким критерием обычно считается наличие доказательства решения задачи. Но этот критерий применим только тогда, когда в рамках самой формальной системы мы имеем как доказательство решения задачи, так и возможность убедиться средствами самой же системы, что данное доказательство действительно является решением задачи. В [3] было доказано, что только в «слабых» формальных системах (для которых не проходит теорема Гёделя) мы можем средствами самой формальной системы определить, является ли некоторый текст доказательством решения задачи или нет.

В результате программа Гильберта обоснования математики может быть сформулирована иначе: не нужно для всей математики доказывать её непротиворечивость, как это было провозглашено в программе Гильберта – это невозможно и ненужно. Надо формулировать и решать задачи в рамках соответствующих им слабых формальных систем.

2.2. Понятие «задача» в когнитивных науках [4]

Обобщением понятия задача в когнитивных науках является понятие *Цели*. Цель нельзя достичь, не имея критерия её достижения, иначе всегда можно считать, что она уже достигнута. Поэтому формулировка Цели всегда должна включать критерий достижения цели. Достижение Цели дает определенный Результат.

Единственной физиологической теорией, в которой достижение Цели и получение Результата рассматривается как решение мозгом ЗАДАЧИ по удовлетворению некоторой потребности, является Теория Функциональных Систем П.К. Анохина [2]. Эта теория также

выявляет физиологические механизмы достижения цели и решения этой задачи мозгом. П.К. Анохин писал: «Пожалуй, одним из самых драматических моментов в истории изучения мозга как интегративного образования является фиксация внимания на самом действии, а не на его результатах ... мы можем считать, что результатом «хватательного рефлекса» будет не само хватание как действие, а та совокупность афферентных раздражений, которая соответствует признакам «схваченного» предмета». «Совокупность афферентных раздражений» и есть критерий достижения цели в ТФС. Целью в ТФС является удовлетворение некоторой потребности: «Каждая потребность, даже при незначительном отклонении жизненно важной функции от оптимального для метаболизма уровня (в чём и проявляется потребность), немедленно воспринимается специальными рецепторными аппаратами» (критерием достижения цели).

Согласно П. К. Анохину, центральные механизмы функциональных систем, обеспечивающих целенаправленные поведенческие акты, имеют однотипную архитектуру.

2.3. Афферентный синтез

Начальную стадию поведенческого акта любой степени сложности составляет афферентный синтез, включающий в себя синтез *мотивационного возбуждения, памяти, обстановочной и пусковой афферентации*.

Мотивационное возбуждение. Постановка цели осуществляется возникшей потребностью, которая трансформируется в мотивационное возбуждение.

Память. Мотивационное возбуждение «извлекает из памяти» все возможные способы достижения цели, а также всю последовательность и иерархию результатов, которые должны быть получены для достижения цели некоторым конкретным способом.

Пусковая афферентация. Пусковая афферентация также является обстановочной афферентацией, только связанной со временем и местом достижения результата.

Пусковая афферентация отвечает на вопрос, *где и когда* можно достичь результат.

2.4. Принятие решений

На стадии афферентного синтеза мотивационным возбуждением может быть извлечено из памяти (в данной обстановке) несколько способов достижения цели. На стадии принятия решения выбирается только один из этих способов – *конкретный план действий*. «Вытягивая» из памяти весь накопленный опыт, мотивационное возбуждение преобразуется в *конкретную цель*, определяющую способ своего достижения. Конкретная цель называется в ТФС *«высшей мотивацией»*.

2.5. Акцептор результатов действия

Мотивационное возбуждение «извлекает из памяти» также всю последовательность и иерархию результатов, которые должны быть получены для выполнения плана действий. Эта последовательность называется в ТФС *акцептором результатов действия*. Акцептор результатов действия представляет собой доминирующую потребность (Цель) организма, трансформированную в форме опережающего возбуждения мозга, как бы в своеобразный *комплексный рецептор* будущего подкрепления, являющийся *критерием достижения конкретной цели*.

2.6. Подкрепление. Санкционирующая стадия

Если в результате выполнения конкретного плана действий цель будет достигнута (потребность удовлетворена) и все результаты акцептора результатов действия получены, то возникает последняя санкционирующая стадия, в которой осуществляется удовлетворение потребности и занесение выполненного конкретного плана действий в память.

2.7. Понятие «задача» в семантическом моделировании

Ранее было показано, что задачный подход достаточно точно отражает когнитивные процессы мозга, направленные на удовлетворение некоторой потребности специально организованным целенаправленным поведением. Расширим теперь задачный подход, ставя своей целью рассмотрение максимально широкого класса задач. Для этого, естественно, нам следует уточнить понятие «задача». Далее будем считать, что некая задача определена в том и только в том случае, когда в ее формулировке присутствуют:

- указание предметной области и знания о предметной области, зафиксированные в виде ее модели, включая описание сигнатуры и структуры языка описания предметной области, набора терминов и понятий, исходные данные, факты и гипотезы;
- на какой запрос (вопрос), сформулированный в задаче, относящийся к предметной области, мы должны получить ответ;
- критерий удовлетворения запроса – в каком случае можно считать, что ответ на запрос (вопрос) получен;
- в каком контексте следует искать ответ на запрос (вопрос) – какую цель мы преследуем, решая задачу, т. е. что мы ожидаем от полученного результата и каковы его последствия и что делать, если ответ окажется отрицательным.

Предлагаемый нами задачный подход предполагает, что истинным назначением ИИ является автоматизация решения задач, понимая термин «автоматизация» в самом широком смысле и считая, что для решения задачи, соответствующие ей компоненты такие как описание предметной области и запросы, должны формулироваться в терминах исполнимых спецификаций. В качестве концепции базовой модели вычислений предлагается взять концепцию Σ -определимости вычислений [4], дополнив ее процедурой проверки истинности Σ -формул на конструктивной модели M , рассматриваемой совместно с ее списочной надстройкой $NW(M)$ [5, 6]. Определим теперь в рамках данной концепции понятие «задача».

Итак, предполагается, что в нашем распоряжении имеется многосортная конструктивная модель M вместе со своей списочной надстройкой $NW(M)$, выступающая как некий базовый вычислитель. Тогда модель предметной области рассматриваемой задачи может быть сформулирована в сигнатуре языка исчисления предикатов этой базовой конструктивной модели M вместе с ее списочной надстройкой $NW(M)$ как набор Σ -определений, т.е. Σ -формул и Σ -термов этого языка. При этом допускаются рекурсивные схемы Σ -определений с некоторыми ограничениями на вхождения в них определяемых предикатов и термов. Запрос же к модели предметной области мы определим также как Σ -формулу, в записи которой могут использоваться как сигнатурные конструкции базовой конструктивной модели, так и определяемые предикаты, и термы предметной области.

Под решением так сформулированной задачи будем понимать набор констант, делающий Σ -формулу запроса при означивании ее переменных константами истинной на модели предметной области. Важно подчеркнуть, что истинность Σ -формулы запроса, получаемой подстановкой констант вместо переменных, и есть критерий решенности задачи. Заметим, что наборов констант, делающих Σ -формулу запроса истинной, может оказаться несколько и потому есть возможность выбора в каком-то смысле наилучшего решения с учетом контекста решения задачи. В этом случае критерий решенности задачи должен содержать и критерий выбора наилучшего ответа на запрос. Так понимаемый подход к формулировке и решению задач носит название *семантического моделирования*.

3. Заключение

Следует отметить, что с практической точки зрения для спецификации задач вполне достаточно использовать не весь Σ -язык, а его Δ_0 -фрагмент, куда относятся Σ -формулы, содержащие в своей записи только ограниченные кванторы всеобщности и существования. Этот Δ_0 -фрагмент обладает тем достоинством, что при определенных условиях можно га-

рантировать полиномиальную вычислимость специфицируемых в этом языке предикатов и функций [7-9].

Работа выполнена при финансовой поддержке Российского научного фонда (проект № 17-11-01176).

Библиографический список

1. Рассел С., Норвиг П. Искусственный интеллект: современный подход. Москва: Вильямс, 2006. 1409 с.
2. Анохин П. К. Избранные труды. Философские аспекты теории функциональной системы. Москва: Наука, 1978. 400 с.
3. Ершов Ю. Л., Самохвалов К. Ф. Современная философия математики: недомогания и лечение. Новосибирск: Издательство Параллель, 2007. 142 с.
4. Vityaev Evgenii E. Purposefulness as a Principle of Brain Activity // Anticipation: Learning from the Past, (ed.) M. Nadin. Cognitive Systems Monographs, Vol. 25, Ch. 13. Springer, 2015, P. 231-254.
5. Ершов Ю. Л. Определимость и вычислимость. Новосибирск : Научная книга, 1996. 300 с.
6. Гончаров С. С., Свириденко Д. И. Семантическое моделирование и искусственный интеллект // Сибирский философский журнал. 2018. Т. 16, № 4. С. 5–25.
7. Goncharov S. S. Conditional terms in semantic programming // Siberian Mathematical Journal. 2017. Т. 58. № 5. С. 794-800.
8. Гончаров С. С., Свириденко Д. И. Логический язык описания полиномиальной вычислимости // Доклады РАН, 2018. Т. 485, № 1. С. 11–14.
9. Goncharov S., Ospichev S., Ponomaryov D., Sviridenko D. The expressiveness of looping terms in the Semantic Programming // Сибирские электронные математические известия. 2020. С. 380-394.

Сведения об авторах / Information about authors

Гончаров Сергей Савостьянович, д-р физ.-мат. наук, академик РАН, директор, Институт математики им. С. Л. Соболева, пр. ак. Коптюга, 4, 630090, г. Новосибирск, Россия. / Goncharov Sergei, doctor of science, academician RAS, director, Sobolev institute of mathematics, 4 Acad. Koptyug avenue, 630090 Novosibirsk Russia,

Витяев Евгений Евгеньевич, д-р физ.-мат. наук, в.н.с., Институт математики им. С. Л. Соболева, пр. ак. Коптюга, 4, 630090, г. Новосибирск, Россия. / Vityaev Evgenii, doctor of science, leader scientist, Sobolev institute of mathematics, 4 Acad. Koptyug avenue, 630090 Novosibirsk Russia.

Свириденко Дмитрий Иванович, д-р физ.-мат. наук, советник директора по инновациям, Институт математики им. С. Л. Соболева, пр. ак. Коптюга, 4, 630090, г. Новосибирск, Россия. / Sviridenko Dmitrii, doctor of science, advisor of the director in innovations of the Sobolev institute of mathematics, 4 Acad. Koptyug avenue 630090 Novosibirsk Russia.