

Реализация в задачном подходе к Искусственному Интеллекту идеи П.К. Анохина об обобщении Теории Функциональных Систем на другие области знаний

Витяев Е.Е.

Институт математики им. С.Л. Соболева СО РАН, Новосибирск, Россия

vityaev@math.nsc.ru

Ключевые слова: теория функциональных систем; задача; искусственный интеллект

Мотивация и цель: В настоящее время широкое развитие получил **Общий искусственный интеллект (Artificial General Intelligence – AGI)**, основанный на том, что моделирование когнитивных процессов человека не является обязательным условием решения интеллектуальных задач в силу результатов глубокого обучения и больших лингвистических моделей. Соответственно, возникает понятие «Общего Искусственного Интеллекта», которым потенциально может обладать, как человек или живой организм с высокоразвитой центральной нервной системой, так и абстрактная робототехническая система. В [1] дается следующее определение AGI: «Это способность решать когнитивные задачи в целом, действуя целенаправленно, адаптируясь к условиям среды через обучение, минимизируя риски и оптимизируя потери на достижение своих целей».

Цель доклада показать, что предлагаемый нами Задачный подход к Искусственному Интеллекту не только решает сформулированную перед AGI задачу, но, что принципиально более важно, моделирует когнитивные процессы человека и поэтому является вариантом «Естественного интеллекта». Как показано в докладе он достаточно точно моделирует целенаправленную деятельность человека в соответствии с Теорией Функциональных Систем (ТФС) работы мозга [2]. Тем самым задачный подход в достаточно полной мере реализует мечту П.К. Анохина об обобщении ТФС на другие области знаний.

Методы и алгоритмы: **Задачный подход. Понятие «задача» в основаниях математики** [3]. Анализ понятия задачи начинается со следующих рассуждений д.ф.н. К.Ф. Самохвалова: «Я хочу пить» – что это значит? Нет, конечно, никакой ошибки полагать, что слова «я хочу пить» означают определенное состояние сознания, именуемое жаждой. Но тогда возникает новый вопрос: как ощущение жажды (хотения) связано с фактическим питьем (удовлетворением хотения)? Откуда я знаю, что удовлетворить жажду можно питьем? Содержится 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. знать и иметь критерий удовлетворения желания.

Таким образом, задача определена (осмыслена) тогда и только тогда, когда есть критерий решенности задачи – критерий проверки действительно ли предъявленное решение 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 формулируется иначе: не нужно для всей математики доказывать её непротиворечивость – это невозможно и ненужно. Надо решать задачи в рамках слабых формальных систем.

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 Цель уже достигнута. Поэтому формулировка Цели всегда должна включать некоторый критерий достижения цели. Достижение Цели дает Результат. Единственной физиологической теорией, в которой достижение Цели и получение Результата рассматривается как *решение мозгом Задачи* по удовлетворению некоторой потребности, является Теория Функциональных Систем [2]. 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. Согласно ТФС центральные механизмы функциональных систем, обеспечивающих целенаправленное поведение, имеют следующую однотипную архитектуру.

Афферентный синтез. Начальную стадию поведенческого акта любой степени сложности составляет афферентный синтез, включающий 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.

Обстановочная и пусковая афферентации фиксируют свойства обстановки и время начала поведенческого акта.

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 них. «Вытягивая» из памяти весь накопленный опыт, мотивационное возбуждение преобразуется в *конкретную цель*, определяющую способ своего достижения. Конкретная цель называется в ТФС *«высшей мотивацией»*.

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

Подкрепление. Санкционирующая стадия. Если в результате выполнения конкретного плана действий цель будет достигнута (потребность удовлетворена) и все подрезультаты акцептора результатов действия также достигнуты, то

возникает последняя санкционирующая стадия, в которой осуществляется удовлетворение потребности и занесение конкретного плана действий в память.

Результаты: Понятие «задача» в семантическом моделировании.

Распространим, как хотел П.К. Анохин, когнитивный процесс достижения цели в ТФС на процесс решения задач. Для этого уточним понятие «задача». Будем считать, что задача определена тогда и только тогда, когда в ней присутствуют:

- предметная область, зафиксированная в виде модели в некоторой онтологии, а также знания о предметной области, исходные данные, факты и гипотезы;
- запрос (вопрос), сформулированный в задаче, относящийся к предметной области, на который 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.

Решение задачи в ТФС может быть сформулировано так:

- предметная область – внешний мир, модель внешнего мира в виде «Образа мира», в котором осуществляется целенаправленное поведение;
- запрос к предметной области – это потребность-цель, которую надо удовлетворить, представленная некоторым целевым предикатом P_0 ;
- под решением задачи понимается объект, ситуация, событие внешнего мира, воспринимаемые «совокупностью афферентных раздражений» a , которое делает предикат P_0 истинным (удовлетворяет потребность). Истинность $P_0(a)$ и есть критерий решения задачи.
- принятие решений. конкретных планов действий по достижению цели P_0 может быть несколько. Во время принятия решений выбирается один план действий с учетом контекста, зафиксированного в модели.

Задачный подход предполагает, что истинным назначением ИИ является автоматизация решения задач и запросы должны формулироваться в терминах исполнимых спецификаций. В качестве концепции базовой модели вычислений предлагается взять концепцию Σ -определимости вычислений [5], дополнив ее процедурой проверки истинности Σ -формул на конструктивной модели M , рассматриваемой совместно с ее списочной надстройкой $NW(M)$ [6, 7].

Определим теперь в рамках данной концепции понятие «задача». Под решением 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. Так понимаемый подход к формулировке и решению задач носит название *семантического моделирования*.

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 гарантировать полиномиальную вычислимость специфицируемых в этом языке предикатов и функций [8, 9].

Финансирование: Работа выполнена по Государственному Заданию: «Логические исчисления и семантики, теории моделей и вычислимости» FWNF-2022-0011.

The implementation in the Task-driven approach to Artificial Intelligence the P.K. Anokhin's idea of generalizing the Theory of Functional Systems to other fields of knowledge

Vityaev E.

S.L. Sobolev Institute of Mathematics, SB RAS, Novosibirsk, Russia

vityaev@math.nsc.ru

Key words: theory of functional systems; task; artificial intelligence

Motivation and Aim: Currently, **Artificial General Intelligence** (AGI) has been widely developed, based on the fact that modeling human cognitive processes is not a prerequisite for solving intellectual problems due to the results of deep learning and large linguistic models. Accordingly, the concept of “General Artificial Intelligence” arises, which can potentially be possessed by both a person or a living organism with a highly developed central nervous system, and an abstract robotic system. In [1], the following definition of AGI is given: “It is the ability to solve cognitive tasks in general, acting purposefully, adapting to environmental conditions through learning, minimizing risks and optimizing losses to achieve one's goals”.

The purpose of the report is to show that our proposed Task Approach to Artificial Intelligence not only solves the problem formulated for AGI, but, more importantly, models human cognitive processes and therefore is a variant of “Natural Intelligence”. As shown in the report, he accurately models purposeful human activity in accordance with the Theory of Functional Systems (TFS) of the brain [2]. Thus, the task approach fully realizes P.K. Anokhin’s dream of generalizing TFS to other fields of knowledge.

Methods and Algorithms: **A task-driven approach.** The concept of “task” in the foundations of mathematics [3]. The analysis of the concept of a task begins with the following arguments of Dr. F.N. K.F. Samokhvalov: “I want to drink” – what does this mean? There is, of course, no mistake in assuming that the words “I am thirsty” mean a certain state of consciousness called thirst. But then a new question arises: how is the feeling of thirst (desire) related to the actual drinking (satisfaction of desire)? How do I know that drinking can satisfy thirst? Does the experience of thirst itself contain a consciousness of how this thirst can be satisfied? ... To know desire does not mean to know what is desired, but means the ability to know what is desired”, i. e. to know and have a criterion for satisfying desire.

Thus, the problem is defined (understood) if and only if there is a criterion for solving the problem – a criterion for checking whether the presented solution is really a solution to the problem. In [3] it was proved that only in “weak” formal systems (for which Godel’s theorem does not pass) we can determine by means of the formal system itself whether a certain text is a proof of the solution of the problem or not. As a result, Hilbert's program for the justification of mathematics is formulated differently: it is not necessary for all mathematics to prove its consistency – this is impossible and unnecessary. It is necessary to solve problems within the framework of weak formal systems.

The concept of “task” in cognitive sciences [4]. A generalization of the concept of a task in cognitive sciences is the concept of a **Goal**. The goal cannot be achieved without having a criterion for achieving it, otherwise it can always be assumed that the Goal has already been achieved. Therefore, the formulation of the Goal should always include

some criterion for achieving the goal. Achieving a Goal gives a Result. The only physiological theory in which achieving a Goal and obtaining a Result is considered as the brain's solution to the Problem of satisfying some need is the Theory of Functional Systems [2]. P.K. Anokhin wrote: "Perhaps one of the most dramatic moments in the history of studying the brain as an integrative education is the fixation of attention on the action itself, and not on its results... we can assume that the result of the "grasping reflex" will not be grasping itself as an action, but that set of afferent stimuli that corresponds to the signs of the "grasped" object". The "totality of afferent stimuli" is the criterion for achieving the goal in TFS. The goal in TFS is to satisfy some need. According to the TFS, the central mechanisms of functional systems that ensure purposeful behavior have the following architecture of the same type.

Afferent synthesis. The initial stage of a behavioral act of any degree of complexity is afferent synthesis, including the synthesis of motivational excitement, memory, situational and trigger afferentation.

Motivational excitement. The goal setting is carried out by the *need* that has arisen, which is transformed into motivational excitement.

Memory. Motivational excitement "extracts from memory" all possible ways to achieve a goal, as well as the entire sequence and hierarchy of results that must be obtained to achieve the goal.

Situational and trigger afferentation fix the properties of the situation and the time of the beginning of the behavioral act.

Decision-making. At the stage of afferent synthesis, motivational excitement can be extracted from memory (in a given setting) several ways to achieve a goal. At the decision-making stage, only one of them is selected. By "pulling" all the accumulated experience from memory, motivational excitement is transformed into a *specific goal* that determines the way to achieve it. A specific goal is called a "higher motivation".

The acceptor of the actions result. Motivational excitement also "extracts from memory" the entire sequence and hierarchy of sub-goals that must be obtained in the process of implementing the action plan and achieving the goal. This sequence is called in the TFS an acceptor of the results of an action, which represents the dominant need (Goal) of the body, transformed in the form of anticipatory excitation of the brain, as if into a kind of complex receptor for future reinforcement, which is a criterion for achieving a specific goal.

Reinforcements. The sanctioning stage. If, as a result of the implementation of a specific action plan, the goal is achieved (the need is satisfied) and all the sub-results of the acceptor of the action results are also achieved, then the last authorized stage arises, in which the need is satisfied and the specific action plan is stored in memory.

Results: The concept of "task" in semantic modeling. Let's extend, as P.K. Anokhin wanted, the cognitive process of the goals achievement in TFS to the process of solving problems. To do this, we will clarify the concept of "task". We will assume that a task is defined if and only if it contains:

- the subject area fixed as a model in some ontology, as well as knowledge about the subject area, initial data, facts and hypotheses;
- a query (question) formulated in the task, related to the subject area, to which we must receive an answer;
- the criterion for satisfying the request is in which case it can be considered that the answer to the request (question) has been received.

- in what context should we look for an answer to the query (question) – what goal are we pursuing when solving the problem, i. e. what do we expect from the result.
- The solution of the problem in the TFS can be formulated as follows:
- the subject area is the external world and the model of the external world in the form of an “Image of the world” in which purposeful behavior is carried out;
 - a request to the subject area is a need—a goal that needs to be satisfied, represented by some target predicate P_0 ;
 - the solution of the problem is understood as an object, situation, event of the external world, perceived by a “set of afferent stimuli” a , which makes the predicate P_0 true (satisfies the need). The truth of $P_0(a)$ is the criterion for solving the problem.
 - decision-making. There may be several specific action plans to achieve the P_0 goal. During decision-making, one action plan is selected, taking into account the context fixed in the model.

The task approach assumes that the true purpose of AI is to automate problem solving and queries should be formulated in terms of executable specifications. As a concept of the basic model of calculations, it is proposed to take the concept of Σ -definability of calculations [5], supplementing it with a procedure for verifying the truth of Σ -formulas on the constructive model M together with its list superstructure $HW(M)$ [6, 7].

Let us now define the concept of “task” within the framework of this concept. By solving the problem, we will understand a set of constants that makes the Σ -formula of the query when denoting its variables by constants true on the domain model. It is important to emphasize that the truth of the query formula obtained by substituting constants instead of variables is the criterion for solving the problem. The so-understood approach to problem formulation and solution is called semantic modeling.

It should be noted that from a practical point of view, for the specification of tasks, it is quite sufficient to use not the entire Σ -language, but its Δ_0 -fragment, which includes Σ -formulas containing only limited quantifiers of universality and existence. This Δ_0 fragment has the advantage that, under certain conditions, it is possible to guarantee the polynomial computability of predicates and functions specified in this language [8, 9].

Funding: The work was performed according to the State Assignment: “Logical calculus and semantics, model theory and computability” FWNF-2022-0011.

Список литературы/References

1. Ведяхин А. и др. Сильный искусственный интеллект. На подступах к сверхразуму. М.: Интеллектуальная Литература, 2021
2. Анохин П.К. Принципиальные вопросы теории функциональных систем. В: Философские аспекты теории функциональных систем. М.: Наука, 1978;49-106
3. Ершов Ю.Л., Самохвалов К.Ф. Современная философия математики. Новосибирск, 2007
4. Vityaev E.E. Purposefulness as a Principle of Brain Activity. In: Nadin M. (Ed.). Anticipation: Learning from the Past. Cognitive Systems Monographs. V. 25. Springer, 2015;231-254
5. Ершов Ю.Л. Определимость и вычислимость. Новосибирск, 2000
6. Гончаров С.С., Свириденко Д.И. Семантическое моделирование и искусственный интеллект. *Сибирский философский журнал*. 2018;16(4):5-25
7. Goncharov S.S. Conditional terms in semantic programming. *Sib Math J*. 2017;58(5):794-800
8. Гончаров С.С., Свириденко Д.И. Логический язык описания полиномиальной вычислимости. *Доклады РАН*. 2018;485(1):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;17:380-394