Task Approach to Artificial Intelligence

Sergei .S. Goncharov
Sobolev institute
of mathematics
4 Acad. Koptyug avenue,
630090 Novosibirsk Russia
S.S.Goncharov@math.nsc.ru

Dmitrii I. Sviridenko
Sobolev institute
of mathematics
4 Acad. Koptyug avenue,
630090 Novosibirsk Russia
dsviridenko47@gmail.com

Evgenii E. Vityaev
Sobolev institute
of mathematics
4 Acad. Koptyug avenue,
630090 Novosibirsk Russia
vityaev@math.nsc.ru

Abstract

The paper present a Task approach to Artificial Intelligence. We argue 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.

1 Introduction. Four main approaches to Artificial Intelligence

In modern Artificial Intelligence (AI) we can distinguish four main directions: Deep Learning, the Agent-based approach, Explainable AI (XAI) and the General AI (AGI). Let us briefly consider these directions.

1.0.1 Deep learning

We pay special attention to this direction, because very often artificial intelligence is mistakenly identified exclusively with deep learning. As well known, the basis of deep learning is the technology of neural networks, allowing to achieve phenomenal success in solving difficult cognitive tasks – enough to remember the achieved advantage of artificial intelligence systems over human in such highly intellectual games such as chess, GO and poker. However, for all the well-known advantages of neural networks, they have serious shortcomings.

Firstly, neural networks are a “black box” and do not provide an opportunity to explain the reasons for making certain decisions when using them. This significantly limits, and in some cases even makes it impossible to use neural networks in areas such as medicine, finance, military applications, logistics, where the price of the error is either too high or the explanation of the decision is necessary for legal reasons. For example, the neural network’s refusal to issue a financial credit or its recommendation to perform a dangerous surgery must be legally justified.

Secondly, the use of neural networks often faces another problem – the problem of “retraining”, which is a consequence of the fact that the neural network “remembers” the answers instead of inference of them based on the regularities and input data.

Third, neural networks have weak capacity for generalization. For example, a neural network trained to recognize elephants and whales, if a whale thrown ashore is presented, will see an elephant in it, and the elephant bathed in the surf will be recognized as a whale.

Fourthly, neural networks memorize individual, often random, details of the samples they receive during training and make decisions based on those details rather than on a fully generalized subject. For example,

In: Sergei S. Goncharov, Yuri G. Evtushenko (eds.): Proceedings of the Workshop on Applied Mathematics and Fundamental Computer Science 2020 , Omsk, Russia, 30-04-2020, published at http://ceur-ws.org

adding a non significant noise to an image or replacing the image with noise may result in the recognition of a nonexistent subject, while replacing only one pixel in the image may result in the recognition of a different subject than the one presented.

Fifth, neural networks are poorly adapted to solving computational problems and cannot work with symbolic and recursive structures.

Sixth, neural networks are characterized by "catastrophic forgetfulness" – they cannot be consistently trained to solve several tasks. As a consequence, neural networks are not able to solve complex tasks by decomposition and steps. Neither can they work in dynamically changing environments.

1.0.2 Agent approach

This approach, detailed in the monograph [Russell2006], acts in a certain sense as an alternative approach to neural network deep learning, since its clear advantage is its distinct integration, multitask nature. This circumstance is a consequence of a successfully chosen ontological base, where the concepts of "rational agent" and "external environment" are proposed as the basic concepts, and various problems of artificial intelligence are considered as tasks of interaction of "rational agent" with "external environment". In work [Russell2006] the total classification of rational agents and environments, and also tasks solved by agents is conducted. Let us draw the reader's attention to the important remark that the agent approach actually follows the task approach, but only without an explicit definition of the concept of "task" and all the components that bring this concept.

1.0.3 Explained Artificial Intelligence (XAI)

This approach to Artificial Intelligence has been formed relatively recently and requires methods used to solve problems of artificial intelligence, the mandatory presence of components that help understand how and why certain decisions are made (for example, in the same deep neural network training). One of the main methods of XAI are the following:

- a method that shows the contribution of each attribute to the resulting forecast, with the accompanying forecast explanation showing the predicted class together, for example, with the pixels that most strongly affected the forecast result;
- a method that highlights features specific to the forecast; note that in order to highlight features that affect the forecast of objects of some class, it is necessary to combine features by class objects.

It should be noted that for the right purposes and all the advantages of this approach, the above methods used in it, however, do not remove from the agenda many problems peculiar to neural networks.

1.0.4 Artificial General Intelligence (AGI)

This direction is a significantly rethought version of the so-called "Strong Artificial Intelligence", which, unlike the alternative to it "Weak Artificial Intelligence", which is solving individual problems, aimed at the ability to solve any intellectual problem solved by human. Taking into account the lack of accurate and generally accepted understanding of what the phrase "any (!) problem that can be solved by a human" means, as well as the presence of a number of known logical and mathematical results, limiting the cognitive capabilities of human, we can safely argue that the so understood goal of strong artificial intelligence is not correct and, therefore, is not achievable. In our opinion, the hopes of supporters and adherents of the Strong Artificial Intelligence on the omnipotence of deep learning will not help here either. Moreover, one should also take into account the disadvantages of deep learning, which were discussed earlier, as well as the fact that the practice of deep learning has shown that the modeling of cognitive processes of a human is not necessarily a condition for the successful solution of cognitive (intellectual) problems of the subject domain to which the task refers. As a matter of fact, this circumstance was the main reason for the critical rethinking of the goals of "Strong Artificial Intelligence" in the 90s and the emergence of the concept of "Artificial General Intelligence" (AGI), which to some extent can potentially possess a human or a living organism with a highly developed central nervous system and also abstract robotic system. Leading developers of AGI (Ben Grzel, Shane Legge, Pei Wang) define it more cautiously and quite comprehensibly as "*... the ability to solve cognitive problems in general, acting with purpose, adapting to the environment through training, minimizing risks and optimizing losses to achieve their goals*".

At present, according to the above definition, AGI, which is a very promising direction, is still mainly of a purely methodological and theoretical nature. Nevertheless, it should be noted that the so-called AGI is of an obvious integration nature and, as can be easily seen from its definition, is much in line with the agency approach. But, as well as the agent based approach, AGI is also characterized by the absence of a precise definition of the task, which, in the authors' opinion, is one of the reasons for its weak practical achievements.

The purpose of the present work is to show that our proposed task approach to artificial intelligence is capable not only of accommodating all of the above directions, satisfying their definitions, but also allows us to model purposeful human activity accurately enough in accordance with the physiological Theory of Functional Systems (TFS) of the brain [Anokhin78].

2 The task approach

Let's start with the presentation of the task approach by describing the results of its use in the analysis of fundamental problems of mathematical foundations. For this purpose, it is necessary, first of all, to give an accurate definition of the "task" concept arising in the mathematical foundations.

2.1 The concept of "task" in the mathematical foundations

The analysis of the task notion begins with the following simple reasoning formulated by Doctor of Philosophy K.F. Samokhvalov: "I want to drink – what does it mean? Of course, there is no mistake to suppose that the words "I am thirsty" simply mean this, where it is a certain state of consciousness that I am experiencing now and which I call thirsty. But then a new question arises: how does the feeling of thirst (of wanting) relate to actual drinking (satisfaction of wanting)? How do I know that it is possible to satisfy thirst by drinking? Is there a consciousness in thirst itself of what can be satisfied with this thirst? ... Knowing desire does not mean knowing what is desired, but the ability to know what is desired", i.e. to have the criterion of satisfying desire.

Thus, the task is defined (understood) if and only if we have the criterion of the task solution – the criterion of checking whether the presented solution is really the solution of the task. In mathematical theories, this criterion is usually considered to be the proof of the task solution. But this criterion is applied only when, within the framework of the formal system itself, we have both the proof of the task solution and the possibility to be convinced by the means of the system itself that this proof is actually the task solution.

In [Ershov–Samokhvalov2007] it was proved that only in "weak" formal systems (for which the Gödel theorem does not hold) we can determine, by means of the formal system itself, whether some text is the proof of the task solution or not.

As a result, the Hilbert program of justification of mathematics can be formulated differently: it is impossible and unnecessary for all mathematics to prove its consistency as it was declared in the Hilbert program. We propose to formulate and solve problems within the framework of weak formal systems corresponding to them.

2.2 The concept of "task" in cognitive sciences. Functional Systems Theory

The analogue of the "task" notion in cognitive sciences is the *Goal* concept. It should be noted that the Goal cannot be achieved without having a criterion of its achievement, otherwise it can always be considered that it has already been achieved. Hence the formulation of the Goal should always be accompanied by the definition of the criterion of its achievement. Achieving the Goal gives a certain Result.

The only physiological theory in which achievement of the Goal and obtaining of the Result is considered as a solution of the TASK by the brain to satisfy a certain need is P.K. Anokhin's Theory of Functional Systems (TFS) [Anokhin78]. This theory also reveals physiological mechanisms for achieving the Goal and solving this task by the brain. P.K. Anokhin wrote: "...perhaps, one of the most dramatic moments in the history of studying the brain as an integrative formation is the fixation of attention on the action itself, not on its results ... we can consider that the result of the "grasping reflex" will be not the grasping itself as an action, but the set of afferent irritations that corresponds to the features of the "grasping" subject. "The totality of afferent irritations" is the criterion of achieving the goal in TFS. Therefore, a necessary condition for a purposeful behavior of an intellectual agent is Goal setting, which includes a criterion of the goal achievement.

Definition of a Goal is paradoxical, as the criterion of achievement of a goal essentially does not contain any knowledge of how to achieve it, it is possible to define a goal without defining the way to achieve it. Let us call this paradox of the goal as a "goal paradox". As will be seen from the theory of functional systems, the brain in case of purposeful behavior permanently resolves the goal paradox, determining what, when and how it can be achieved.

The goal in TFS is satisfaction of some *Need*: "Every need, even at insignificant deviation of vital function from optimal level for metabolism (in what the need is manifested), is immediately perceived by specific receptor apparatuses" (criterion of goal achievement). The leading excitation ... determining the purposeful activity is the motivational excitement formed on the basis of the dominant need.

Interaction of results and goals in TFS is performed in several ways: on the "dominant principle", "results hierarchy" and "results models".

In relation to the dominant functional system, the other functional systems are structured in a hierarchy based on the principle of "results hierarchy": "For example, hungry rabbit is dominated by a functional system whose activities are aimed at finding food. At this time, other functional systems that determine, for example, cross pressure, respiration, separation, are aimed at better providing the dominant functional system.

According to P.K. Anokhin, the central mechanisms of functional systems that provide purposeful behavioral acts have the same architecture, which will be considered further.

2.2.1 Afferent synthesis

The initial stage of a behavioral act of any degree of complexity is afferent synthesis, which includes the synthesis of motivational excitation, memory, situational and starting afferentation.

Motivational excitation. Goal setting is carried out by the arising need, which is transformed into motivational excitation.

Memory. Motivational excitation "extracts from memory" all possible ways to achieve the goal, as well as the entire sequence and hierarchy of results that should be obtained to achieve the goal in some specific way.

Situational Afferentation. While recording a memory trace, the situation in which the result is attained is also being recorded. This situation is registered, along with the motivation, as a necessary precondition for attaining the goal. Thus, the motivational stimulus in this situation extracts only those ways of attaining the goal that are possible in the given situation.

Starting afferentiation. Starting afferentation is also a situational afferentation, only related to the time and place of the goal achievement. The starting afferentiation answers the question where and when the result can be achieved.

Thus, at the afferent synthesis stage, the goal paradox is largely resolved and it is determined what, when and how to achieve the goal.

2.2.2 Decision making

At the stage of afferent synthesis by the motivational excitation several ways of the goal achievement can be extracted from the memory. At the stage of decision making, only one of these ways is chosen – a concrete *action plan*. By "stretching" all the accumulated experience out of memory, the motivational excitation is transformed into a *concrete goal*, which defines the way of its achievement. This goal is called a "the highest motivation".

2.2.3 Acceptor of action results

Motivational excitation also "extracts from the memory" the entire sequence and hierarchy of results, which must be obtained to perform the action plan. This sequence is called in TFS the *acceptor of action results.* it is a dominant need (Goal), transformed in the form of advanced excitation of the brain, as a kind of *complex receptor* of future reinforcement, which is a *criterion for achieving a concrete goal.*

2.2.4 Reinforcement. The sanctioning stage

If the goal is achieved (the need is met) as a result of the execution of the concrete action plan and all the results of the acceptor of the action results are obtained, then the last sanctioning stage takes place, in which the need is met and the executed concrete action plan is recorded into the memory: "The purposeful behavior ... ends in the last sanctioning stage. At this stage, during the action of an irritant satisfying a leading need (reinforcement), the parameters of the achieved result cause flows of inverse afferentiation, which by all its properties corresponds to the previously programmed properties of the reinforcing irritant in the acceptor of action results". If one of the results of the acceptor of action results is not received and the corresponding goal is not reached, then there is an approximate–research activity arise: "The evaluation of the achieved result of the action takes place with the help of active orientation and research activity, which occurs in all cases when the result of the performed action does not correspond to the properties of the action result acceptor, i.e. in case of "mismatch" in the behavioral activity. Thanks to the inclusion of such a reaction, the afferent synthesis is immediately reconstructed, a new decision is made and a new program of actions is built".

2.3 The concept of "task" in semantic modeling

Thus, we have shown that task approach allows for a fairly accurate and complete reflection of cognitive processes in the brain, aimed at satisfaction some needs with specially organized purposeful behavior. Let us now expand the task approach, with the aim of providing an automated solution to the widest possible class of problems. To do this, of course, we need to clarify the "task" concept.

Further, we assume that a certain *task* is defined if and only if its formulation contains:

- an indication of the subject domain and knowledge about the subject domain recorded in the form of its model, including a description of the signature and structure of the language for describing the subject domain, a set of terms and concepts, their relationships, as well as initial data, facts and hypotheses;
- to what request (question), related to the subject domain and formulated in the task, we should get an answer;
- criterion of satisfaction of the request – in which case we can assume that the answer to the request (question) is received;
- in what context should we search for an answer to a query (question) what goal do we pursue while solving the problem, that is, what do we expect from the result and what are its consequences and what to do if the answer turns out to be negative.

The task approach we propose with respect to artificial intelligence assumes that the true purpose of artificial intelligence is to automate the process of problems solving, understanding the term "automation" in the broadest sense and believing that to solve a task, the components corresponding to it, such as subject domain description and queries, should be formulated in terms of executable specifications. It is proposed to take the concept of Σ–definability of computations [Ershov2000] as a concept of the basic computational model, supplementing it with an effective procedure for verifying the truth of Σ–formulas on the constructive model M, considered together with its list add-in HW(M) [Goncharov–Sviridenko1986, Goncharov–Sviridenko1989, Goncharov–Sviridenko2018a, Goncharov–Sviridenko2018b, Goncharov–Sviridenko2018c]. It should be noted that the idea of using the logical–mathematical theory of multisorted constructive models together with their list add–ins for specification and automatic solution of a wide range of tasks was formulated back in the 80s of the last century (see, for example, [Goncharov–Sviridenko1986, Goncharov–Sviridenko1989]) and is actively developing currently. It should be noted that the proposed concept allows for conservative thermal enrichment with conditional and recursive terms [Goncharov2017, Goncharov–Sviridenko2018b], which makes it possible within the framework of this theory to build developed logical and functional means of specification and problem solving. We now indicate how, within the framework of this concept, the concept of task is postulated.

So, it is assumed that we have at our disposal a multisorted constructive model M together with its list add-in HW (M), which acts as a kind of basic computer. Then the domain model of the task can be described in the language of predicate calculus of the signature of this basic constructive model M, together with its list add-in HW (M), as a set of Σ–definitions, i.e. Σ–formulas and Σ–terms of this language. Moreover, recursive schemes of Σ–definitions are allowed with some restrictions on the occurrence of definable predicates and terms in them. The query for the domain model will also be defined as an Σ–formula, in the record of which both signature constructions of the basic structural model and Σ–defined predicates and terms of the domain can be used.

By the solution of the task formulated in this way, we mean a set of constants making the Σ–query formula with its variables substituted by constants true in the domain model. It is important to emphasize that the truth of the Σ–query formula obtained by substituting constants instead of variables is the criterion for the problem solution. Note that there may be several sets of constants that make the Σ–query formula true, and therefore there is the possibility of choosing in some sense the best solution, taking into account the context of the problem solution. In this case, the criterion for the problem solution should also include the criterion for choosing the best response to the query. The so–understood approach to the formulation and automatic solution of tasks is called *semantic modeling*.

It should be noted that, from a practical point of view, 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 in the specification of tasks. This Δ_0 fragment has the advantage that under certain conditions it is possible to guarantee the polynomial computability of the predicates and functions specified in this language [Goncharov-et-al-2020, Goncharov–Sviridenko2018c].

It is easy to verify that, within the framework of semantic modeling, one can easily formulate the cognitive approach discussed above to formulate the Goals and find by the brain ways to achieve them [Vityaev2015]. From the point of view of semantic modeling, the tasks solved by the brain in satisfyng a certain needs can be formulated as follows:

– the domain model is the external world in which purposeful behavior is carried out;

– a request (Goal) to a subject domain is a need that must be satisfied; we denote it by the predicate P_0;

– by the solution of the so–understood task–goal, we understand the achieving an object, situation, event of the external world, perceived by the "totality of afferent stimuli" a, which makes the predicate P_0 true (satisfies the need) on a (that is P_0(a) is true). This is the criterion of the task–goal solution;

– note that there can be several specific action plans to achieve the Goal that make the predicate P_0 true. The choice of the desired plan is carried out during the decision making.

References

[Russell2006] S. Russell, P. Norwig. Artificial Intelligence: A Modern Approach. Kiev:Williams, 2006.

[Anokhin78] P.K. Anokhin Fundamental questions of the theory of functional systems. Philosophical aspects of the theory of functional systems. M.: Nauka, 1978. (in Russian)

[Ershov–Samokhvalov2007] Yu.L. Ershov, K.F. Samokhvalov. Modern Philosophy of Mathematics. Novosibirsk, 2007.

[Ershov2000] Yu.L. Ershov Definability and computability. Novosibirsk, 2000. (in Russian)

[Goncharov2017] S.S. Goncharov onditional terms in semantic programming. *Siberian Mathematical Journal*, 58(5):794–800, 2017.

[Goncharov–Sviridenko1986] S.S. Goncharov, D.I. Sviridenko Theoretical aspects of -programming. *Lecture Notes in Computer Science*, 215:169–179, 1986.

[Goncharov–Sviridenko1989] S.S. Goncharov, D.I. Sviridenko -programming. *Transl., II.Ser., Am. Math. Soc.*, 142:101–121, 1989.

[Goncharov–Sviridenko2018a] S.S. Goncharov, D.I. Sviridenko Semantic modeling and artificial intelligence. *Siberian Journal of Philosophy*, 16(4):5–25, 2018. (in Russian)

[Goncharov–Sviridenko2018b] S.S. Goncharov, D.I. Sviridenko Recursive terms in semantic programming. *SMJ*, 59(6):1279–1290, 2018. (in Russian)

[Goncharov–Sviridenko2018c] S.S. Goncharov, D.I. Sviridenko The logical language for the description of polynomial computability. *Doklady RAS*, 485(1):11–14, 2018. (in Russian)

[Goncharov-et-al-2020] S. Goncharov, S. Ospichev, D. Ponomaryov, D. Sviridenko The expressiveness of looping terms in the Semantic Programming. *Siberian Electronic Mathematical Reports*:380-394, 2020.

[Vityaev2015] E.E. Vityaev Purposefulness as a Principle of Brain Activity. *Anticipation: Learning from the Past, (ed.) M. Nadin. Cognitive Systems Monographs*, 25(13):231–254, 2015.