Semantic Supervised Training for General Artificial Cognitive Agents

Roman V. DUSHKIN[a,1] and Vladimir Y. STEPANKOV[b]
[a] *Artificial Intelligence Agency, Volkonsky 1st Lane, 15, 127473 Moscow, Russia*
[b] *National Research Nuclear University MEPhI, Moscow 115409, Russia*

Abstract. The article describes the authors' approach to the construction of general-level artificial cognitive agents based on the so-called «semantic supervised learning». Within this approach in accordance with the hybrid paradigm of artificial intelligence, both machine learning methods and methods of the symbolic approach («good old-fashioned artificial intelligence») are used. A description of current problems with understanding of the general meaning and context of situations in which narrow AI agents are found is presented. The definition of semantic supervised learning is given and its relationship with other machine learning methods is described. In addition, a thought experiment is presented, which shows the essence and meaning of semantic training with a teacher, which makes it possible to «educate» a general-level AI agent. It opens the opportunity to apply not only general-level knowledge about the world around the agent, but also introduce the personal experience, which, according to the authors, will lead to a full understanding of the context and, ultimately, to the construction of general artificial cognitive agents. The article also provides a possible architecture for an artificial cognitive agent that can implement semantic supervised learning. The novelty of the work lies in the authors' approach to combining various methods of artificial intelligence within the hybrid paradigm. The relevance of the work is based on the ever-increasing interest in general-level artificial intelligence methods in science, technology, business and world politics. The article is theoretical. The article will be of interest to specialists in the field of artificial intelligence (especially in the direction of building artificial general intelligence), philosophy of consciousness, and in general to all those who are interested in up-to-date relevant information about the approaches and methods of implementing general artificial cognitive agents.

Keywords. Artificial intelligence, agent, agent-based approach, artificial general intelligence, supervised learning, semantics, context processing, personal experience, ontology, architecture

1. Introduction

The problem of creating general level artificial intelligence (Artificial General Intelligence, AGI) is becoming a technological trend of modern times again [1].

---

[1] Corresponding Author: Roman V. DUSHKIN, Artificial Intelligence Agency, Volkonsky 1st Lane, 15, 127473 Moscow, Russia; E-mail: drv@aiagency.ru.

Successes in the construction of narrow artificial cognitive agents which have already surpassed the human level for some specific tasks [2] give hope that with the proper level of technical elaboration it will be possible to move from narrowly focused artificial intelligence systems to artificial general intelligence. At the same time, there is a lot of skepticism about this trend. Since there are reasonable doubts that the extensive development of narrow artificial intelligence systems will make it possible to make a qualitative transition to AGI [3]. Intensification of research and the renewal of interdisciplinary dialogue are required to search for possible solutions in the field of building artificial general intelligence.

It should be noted that such an intensification of research can take place in several directions simultaneously. The authors see at least the following:

- Development of new types of architecture of artificial cognitive agents which allow them to solve problems inherent for AGI: independent goal-setting and the ability to independently learn new skills.

- Using the achievements of related sciences and organizing a comprehensive interdisciplinary dialogue for a deeper understanding of the nature of human intelligence.

- Application for the design and implementation of the general-level cognitive agent of the new knowledge gained from working with the natural intelligence of humans.

In this work the attempt is made to present the authors' vision and understanding of how one of the processes of «growing» an general-level artificial cognitive agent should be arranged by teaching such an agent general knowledge about its environment (not necessarily coinciding with human reality) for gaining the ability to solve arbitrary cognitive tasks later. An outline of the architecture of such an artificial cognitive agent is given.

2. Recognition and Understanding

All research in the field of artificial intelligence could be divided into several large classes then, in fact, everything can be put into two large paradigms – top-down and bottom-up [4]. This division was proposed back in the days of the first researchers of artificial intelligence at the Computer Science and Artificial Intelligence Laboratory of the Massachusetts Institute of Technology — John McCarthy and Marvin Minsky. The first advocated the methods of the top-down paradigm (the so-called «pure artificial intelligence»), and the second — for the methods of the bottom-up paradigm (respectively, «dirty artificial intelligence»).

The top-down approach in artificial intelligence is focused on methods of working with knowledge — these are knowledge bases and methods of acquiring, representing and processing knowledge, and various kinds of symbolic computation methods. The main feature of the approach is the ability to make decisions and explain the principles and reasons for making exactly such decisions. The disadvantages of this approach include the difficulty of learning for knowledge-based systems [5].

The bottom-up approach, based on training of artificial cognitive agents on data (machine learning, artificial neural networks), allows to effectively teach agents to solve

specific problems. But at the same time, in general terms, such agents remain «black boxes» with the complexity of explaining decision options [6].

Traditionally, bottom-up AI techniques have been used to solve pattern recognition tasks, to find hidden patterns in large data sets, and other similar tasks. In particular, the task of pattern recognition (in a general sense) has become the area where deep learning methods in some special cases have already surpassed the cognitive abilities of the human brain. For these particular cases, the accuracy of neural network solutions in pattern recognition is already higher than that of humans [7]. So there are no fundamental obstacles to increasing the accuracy even higher, as well as to expand the methods to solve new tasks.

At the same time, pattern recognition does not lead to understanding [8]. In general, the problem of understanding is acute today. It stood in the history of the development of artificial intelligence methods, since the understanding of the current cognitive agent's situation, especially taking into account the context and memory of the agent's «life experience», is the key to competent decision making. Without understanding the general meaning, it will only be possible to build narrow artificial intelligence systems for solving insular specific problems [8]. This thesis can be illustrated by the example of insects — their natural neural networks are quite capable of recognizing images, but no understanding arises in them. The bee visually recognizes flowers from which delicious nectar can be extracted, but does not understand exactly (at least in the human sense) what it does, why it does and for what purpose.

However, the methods of the top-down paradigm of artificial intelligence also do not lead to the emergence of understanding in artificial cognitive agents, despite the fact that these methods are initially aimed at working with knowledge and meaning. The processing of the context of the situation in which the artificial cognitive agent could be found is still an unsolved problem. As well as the use of «general common sense» has not yet been implemented. However, the situation is gradually changing — the achievements and successes of the artificial neural network GPT-3 amaze the imagination [9].

Nevertheless, the problem remains open. With all the advances in pattern recognition, the understanding of meaning by artificial cognitive agents leaves much to be desired. This is also because modern solution architectures are initially sharpened for solving a specific applied problem, that is, from their very inception, they are a narrow artificial intelligence. Therefore, new approaches and architectures are required.

3.   Semantic Learning

In order for a cognitive agent to understand the meaning of the situation in which it finds itself, it's training must be carried out in a way to ensure that it creates a large set of associative connections between units of knowledge [8]. Understanding of the meaning in this case arises after the activation of all relevant units of knowledge by associative connections when the agent focuses on the selected set of stimuli that form a description of the current state of the environment in which the cognitive agent functions.

Therefore, to teach an artificial cognitive agent, which must understand the meaning of the situation in which it finds itself, the new machine learning methods should be used. One of the composite methods that can be proposed for this is based on the simultaneous use of machine learning principles and methods of semantic situation analysis. This

method could be called «semantic learning», and it is based on the same learning principles that are used by a human in the process of its growth and education.

Considering, semantic learning is based on the following principles:

- At the basic level the agent need to use the three classic machine learning methods — supervised, unsupervised and reinforcement learning — all these machine learning methods are used to solve various problems [10], namely:

- o Unsupervised learning — identifying various hidden patterns in a variety of external stimuli, clustering objects and phenomena, simplifying input data.

- o Supervised learning — correlating objects and phenomena identified during unsupervised learning with the specified semantic classes.

- o Reinforcement learning — developing optimal (or at least suboptimal) strategies for working in an agent's environment to achieve particular given goals.

- After obtaining data from the basic level of machine learning, the agent needs to use the methods of top-down artificial intelligence to build semantic relationships and enrich the semantic network. In this process, the agent itself can act from an active position, interacting with the teacher in such a way as to receive more information about those aspects of it's environment that it still knows little about. Thus, semantic learning is carried out, which further leads to the agent's understanding of the meaning of situations in which it functions in accordance with [8].

Further in this work, at the basic level, only the supervised machine learning option will be considered, which, together with the semantic component, leads to the use of supervised semantic learning.

The general scheme of work of supervised semantic learning is as follows:

- Using the methods of supervised machine learning, the agent recognizes the states in which it finds itself in its environment. These recognized situations are placed in its «operative memory» for further use in the process of building semantic relations between a set of concepts, which includes concepts (named or yet unnamed from the point of view of the agent objects and phenomena) from the recognized situation.

- Then, by comparing the recognized information about the current state in which the agent is, with the knowledge available in its knowledge base, it tries to create new associative links between the concepts available in its associative memory and the new ones obtained from the previous stage.

- If gaps remain in the set of connections of new objects and situations at the previous stage, the agent performs some protocol of interaction with the teacher to build the missing associative connections by actively obtaining information in a dialogue with the teacher. An important feature of the described approach is the presence of procedures for identifying such gaps and as a consequence for launching an interaction protocol.

In order to clarify the presented ideas, it is possible to make the following thought experiment. It is based on introspection and observation of preschool children learning in terms of study of new concepts and related objects or phenomena of objective reality.

Imagine a preschooler who studies the world around him with interest. He walks around the garden being on a summer vacation visiting his grandmother and then he sees something on the sorrel garden bed that he has never seen before — some kind of slimy brown lump attached to the stem of the plant. The preschooler asks his grandmother what it is and she answers: «It's a slug». The preschooler finds a few more pieces of slugs and asks if they are slugs. Having received a positive answer, he is satisfied and does not ask anything else. However, for training purposes it would be enough to look at one slug from different angles or even in motion.

After some time, the learning process begins to cycle in his nervous system. The images of several slugs that he saw earlier begin to spin in his head along with the word «slug», which he heard and remembered. It can be assumed that this process is unconscious and is associated with the consolidation of long-term memory during sleep. But it is also possible that the echoes of this process break through to conscious perception in the form of reflexive learning loops at which time the child may even repeat the word «slug» out loud.

After the consolidation of memory which completes the supervising training in the framework of pattern recognition the next stage of training begins — the semantic one. The new concept should be built into the semantic network that the preschooler already has — and it does not matter that this semantic network is encoded in the neural networks of a living cognitive agent by means of sparse coding — the substrate in the current context does not matter. It is only important that there are numerous associative connections of different types between the set of cortical columns, in which the new concept of «slug» is encoded. It should be noted that semantic information should have at least two levels — linguistic and, in fact, semantic.

The linguistic level is usually built implicitly on the basis of those linguistic patterns that already exist in the linguistic space of the preschooler. After some time, he asks: «Slug — who is it?» And they answer him: «Not who, but what. It's a snail». New linguistic information was implicitly received — the word should be declined according to the type of an animate being. And if for the English language this example is a bit artificial, for many other languages the definition of gender is not so trivial. Using the analogy with already learned models and examples of words in tune with the new one does not give an answer. The phonetic consonance of words does not allow to unambiguously determine the gender of the object expressed by it. So, the preschooler continues assuming that the word «slug» is neuter since this can reasonably be assumed on the basis that it is a snail. «Can slug be touched? Will it not bite?» And again, he receives a corrective signal: «The slug can be touched». At this stage most of the linguistic information necessary for the correct use of the word in utterances is recorded in the semantic network.

At the same time, a part of the semantic information has already been formed — the IS-A relation has been received, the «slug» node has been built into the hierarchy of living beings in the «snail» group. Some of the semantic information is obtained from observations — various attributes of an object, such as color, softness, etc., as well as various associative connections, also obtained from observation, such as «creature» (if the slug moved), «lives on sorrel», etc. But then the preschooler can clarify various

additional information in order to extend more and more associative links from the new concept that has so interested him to the existing knowledge.

Thus, the use of a similar approach for training an artificial cognitive agent designed to solve arbitrary problems will create a conceptual field of knowledge for it. This field will consist of two components: universal and individual. The individual conceptual field includes the «personal life experience» of an artificial cognitive agent, while the universal conceptual field is general semantic information on human knowledge [11]. Unlike humans, the general knowledge for an AI agent can be obtained by integrating with such information systems as universal ontologies (for example, CYC [12] or the semantic web [13]).

4.   Possible Architecture for an Artificial Cognitive Agent

Having considered the basic principles of semantic supervised learning, it makes sense to start designing a possible architecture of an artificial cognitive agent in which this type of machine learning is possible. This section provides a draft of such an architecture for further discussion in the expert community and testing in practice.

The architecture of an artificial cognitive agent with the possibility of implementing semantic supervised learning is presented in Figure 1.
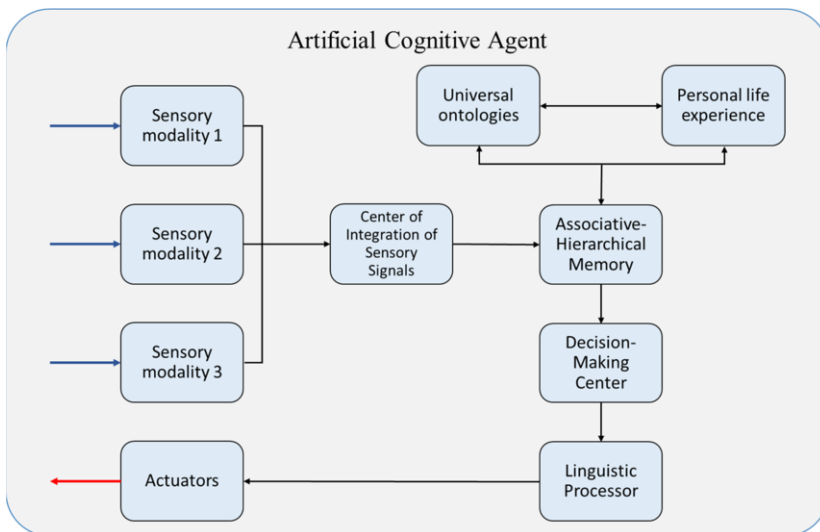

Figure: 1. General architecture of artificial cognitive agents for supervised semantic learning.

The presented diagram underlines the high-level structure of an artificial cognitive agent, within which semantic supervised learning could be implemented. The learning cycle in this case is organized as follows:

●   The artificial cognitive agent receives signals from the environment through sensors that can be combined with additional elements such as filters and preprocessors. This combination of sensors with additional input signal processing elements is considered together as a separate sensory modality. The

artificial cognitive agent can have more than one sensory modality that receives signals of a different nature from the environment and transforms them into a sequence of data suitable for further processing.

●   Further, all data received from different sensory modalities are sent to the Center of Integration of Sensory Signals, where the process of multisensory integration is carried out, which includes a low-level fusion of data and their mutual enrichment to build a holistic description of the current situation perceived in the environment. If an artificial cognitive agent has only one sensory modality, then the Center of Integration may be absent, although there are serious concerns that such an agent will be able to use supervised semantic learning. On the other hand, chatbots, as cognitive agents receiving input signals of only one modality through the messenger console, may well use the described approach. This issue requires additional detailed research and is beyond the scope of this work.

●   After the process of sensory integration, a holistic description of the perceived external environment from the Center of Integration is transferred into the Associative-Hierarchical Memory, which also actively interacts with two data banks — general information about the environment and the personal experience of the cognitive agent. Both of these data banks are filled by the process of semantic supervised learning, and this process is carried out constantly and continuously, including the current act of perception and interaction with the environment. The data related to the current perception are recorded to data banks — existing information is updated; new information is added if necessary.

●   In the associative-hierarchical memory all the necessary associations and hierarchical concepts associated with the perceived situation in the environment are «excited». This set of excited concepts is passed on to the Decision-Making Center, where a list of possible reactions of the AI-agent to the stimuli received from the environment is formed. Then the optimal response is selected and transmitted to the Linguistic Processor (and other low-level command blocks for actuators, if any exists). The optimal response is selected by the means of symbolic inference within the Decision-Making Center or with any sort of emotional simulation, but this question is still to be investigated.

●   The Linguistic Processor receives the formed decision from the Decision-Making Center and forms on its basis the reaction of the cognitive agent and sends it to the actuators (executive devices). The mentioned mechanisms are the output interfaces of the cognitive agent with the environment, transferring to it what they received from the Linguistic Processor in a form understandable to other agents.

By itself, the artificial cognitive agent plays an essential role in organizing supervised semantic learning. It should be the initiator of questions to the teacher, the answers to which will help it to actualize its semantic network. The mechanisms must be implemented in the agent that determine the incompleteness in the semantic network and form the corresponding reactions aimed at filling this incompleteness. These mechanisms must be implemented in the Decision-Making Center.

It should be noted that the presented scheme and its further description are deliberately devoid of indications of any specific technologies or methods for

implementing individual components. The presented architecture is positioned in a generalized way so that when implementing the cognitive AI-agent with the possibility of supervised semantic learning, this scheme could be implemented in specific ways of functional execution. For example, within the framework of the hybrid approach in artificial intelligence [14], professed by the authors, the Center of Integration of Sensory Signals can be implemented in the form of neural networks of suitable architecture, Associative-Hierarchical Memory — in the form of a special semantic maps [15], The Decision-Making Center is in the form of a symbolic inference machine [16], and the Linguistic Processor is again in the form of neural networks [17].

5. Conclusion

The article presents the authors' vision of the opportunities that open up to researchers in the field of building and «growing» general-level artificial cognitive agents. The method of supervised semantic learning is presented, which is aimed not only at the formation of hidden models of pattern recognition or the relationship between stimuli and reactions of an artificial agent, but at the construction of a semantic map of the general description of the surrounding reality and the agent's personal experience of functioning in it. The presented approach probably has the right to exist and be further developed in the framework of interdisciplinary research in the field of building artificial general intelligence.

References

[1] Williams A. E. (2020). A Model for Artificial General Intelligence // In book: Artificial General Intelligence. — July 2020. — DOI: 10.1007/978-3-030-52152-3_38.
[2] He K., Zhang X., Ren Sh., Sun J. (2015) Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification / IEEE International Conference on Computer Vision (ICCV 2015). 1502. — DOI: 10.1109/ICCV.2015.123.
[3] Harnad S. (1990) The Symbol Grounding Problem // Physica D. 42 (1-3). — p. 335-346.
[4] Dushkin R. V. (2018) Obzor podkhodov i metodov iskusstvennogo intellekta [The Review on approaches and methods of artificial intelligence] // Radioelektronnye technologii [Radioelectronic technologies], № 3, 2018. — p. 85-89. — URL: https://goo.gl/nYczgo.
[5] López M. (2015) Four unsolvable problems of symbolic AI // Revista de Filosofia. 40. — p. 81-104. — DOI: 10.5209/rev-resF. 2015.v40.n1.48441.
[6] [6] Nikolenko S., Archangelskaya Ye., Kadurin A. (2018) Glubokoye obucheniye. Pogruzheniye v mir neyronnykh setey [Deep Learning. Dive into the world of neural networks]. — SPb: Piter, 2018. — 480 p. — ISBN 978-5-496-02536-2.
[7] Khanam S., Tanweer S., Khalid S. (2020) Artificial Intelligence Surpassing Human Intelligence: Factual or Hoax // The Computer Journal, January 2020. — DOI: 10.1093/comjnl/bxz156.
[8] Dushkin R. V. (2020) Kritika «Kitayskoy komnaty» J. Searla s pozitsii gibridnoy modeli postroeniya iskusstvennykh kognitivnykh agentov [Criticism of J. Searle's «Chinese Room» from the perspective of a hybrid model for constructing artificial cognitive agents] // Sibirsky filosofsky zhurnal [Siberian philosophical magazine], № 2, 2020. — p. 30-47.
[9] Benzon W. (2020) GPT-3: Waterloo or Rubicon? Here be Dragons. — Preprint. — DOI: 10.13140/RG.2.2.18525.03048.
[10] Shumsky S. A. (2020) Mashinny intellekt. Ocherki po teorii mashinnogo obucheniya i iskusstvennogo intellekta [Machine intelligence. Essays on Machine Learning and Artificial Intelligence Theory]. — Moscow: RIOR, 2020. — 340 p. — ISBN: 978-5-369-01832-3.
[11] Heck R. G. (2009). Logic, semantics, ontology — PhD Thesis, January 2009. Massachusetts Institute of Technology, Dept. of Linguistics and Philosophy, 2009.

[12] Witbrock M. et al. (2005) Knowledge Begets Knowledge: Steps towards Assisted Knowledge Acquisition in Cyc // In: Papers from the 2005 AAAI Spring Symposium on Knowledge Collection from Volunteer Contributors (KCVC). — pp. 99-105. Stanford, California, March 2005.

[13] Yu L. (2014) A Developer's Guide to the Semantic Web. — 2nd ed. Springer. — ISBN 978-3-662-43796-4.

[14] Dushkin R. V., Andronov M. G. (2020) The Hybrid Design for Artificial Intelligence Systems // In: Arai K., Kapoor S., Bhatia R. (eds) Proceedings of the 2020 Intelligent Systems Conference (IntelliSys), Volume 1 (1250). — Springer, Cham, 2020. — P. 164-170. — ISBN 978-3-030-55179-7. — DOI: https://doi.org/10.1007/978-3-030-55180-3_13.

[15] Georgakopoulos T. (2019) Semantic Maps // In book: Oxford Bibliographies in LinguisticsPublisher: New York: Oxford University Press, January 2019. — DOI: 10.1093/obo/9780199772810-0229.

[16] Bibel W., Kurfess F., Aspetsberger K., Hintenaus P., Schumann J. (1986) Parallel Inference Machines // In book: Future Parallel Computers, An Advanced Course, Pisa, Italy, June 1986. — Chapter: 5. — Publisher: Springer Verlag, Berlin, Lecture Notes in Computer Science 272. — Editors: P. Treleaven, M. Vanneschi. — DOI: 10.1007/3-540-18203-9_5.

[17] Liu Zh., Lin Y., Sun M. (2020) Representation Learning and NLP // In book: Representation Learning for Natural Language Processing, July 2020. — DOI: 10.1007/978-981-15-5573-2_1.