Neural Networks in Semantic Analysis

1st Alexey Averkin
Dorodnicyn Computing Centre
FRC CSC RAS
Moscow, Russia
averkin2003@inbox.ru

2nd Sergey Yarushev
Department of Informatics
Plekhanov Russian University of Economics
Moscow, Russia
Yarushev.SA@rea.ru

***Abstract*—This paper presents research of the possibilities of application deep neural networks in semantic analysis. This paper presents the current situation in this area and the prospects for application an artificial intelligence in se-mantic analysis and trend and tendencies of this science area. For better un-derstanding future tendencies of researches in semantical area we present detailed review of the studies in semantic analysis with using artificial intelligence, studies about a human brain.**

***Keywords*—Semantic Analysis, Deep Neural Networks, Forecasting, Processing of Nat-ural Language.**

I. Introduction

Machine learning is a central technology of artificial intelligence. In recent years, in this area all the attention was focused on the deep learning technology as a method for automatically extracting characteristic values. Huge amounts of time series data collected from the devices, especially in the era of Internet of Things. Applying deep learning that data and classifying them with a high degree of accuracy, it is possible to carry out further analysis with a view to creating new products and solutions and open up new areas of business. Deep learning [1] technology “Fig. 1”, which is seen as a breakthrough in the development of artificial intelligence, delivers the highest accuracy of image recognition and speech, but it is still applicable only to limited types of data. In particular, it has so far been difficult to classify accurately automatically variable time series data coming from the devices connected to the Internet of things. The theory and practice of machine learning are experiencing this "deep revolution" caused by the successful application of methods Deep Learning (deep learning), representing the third generation of neural networks. In contrast to the classical (second generation) neural networks 80-90-ies of the last century, a new learning paradigm allowed to get rid of some of the problems that hindered the dissemination and successful application of traditional neural networks. Networks trained by deep learning algorithms, not just beat the best in terms of accuracy, alternative approaches, but also in a number of tasks showed the beginnings of understanding the meaning of information supplied (for example, image recognition, analysis of textual information, and so on). Most successful modern industrial methods of computer vision and speech recognition built on the use of deep networks, and IT-industry giants such as Apple, Google, Facebook, buying groups of researchers involved in deep neural networks. Deep Learning - is part of a broader family of machine learning methods - learning concepts, which feature vectors are located directly on a variety of levels. These features are automatically determined and associated with each other, forming output data. At each level, abstract presented features based on attributes of the previous level. Thus, the deeper we move forward, the higher level of abstraction. The neural network is a plurality of layers of a plurality of levels with feature vectors that generate output data.

The reported study was funded by RFBR, project number 20-010-00828.

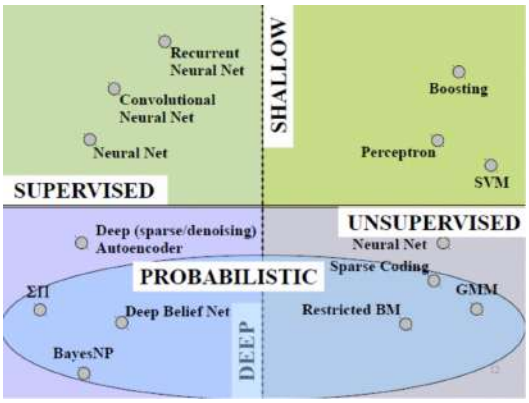

Figure 1. Classification of deep architectures.

II. Histrory of Deep Learning

The team of graduate students from the University of Toronto (Canada), led by Professor Jeffrey Hinton [2] (Geoffrey E. Hinton), won the competition held by the pharmaceutical company Merck. With access to a limited set of data describing the chemical structure of the molecules 15, a group of Hinton was able to create and use a special software system, which is to determine which of these molecules will be more effective than other work as drugs. The peculiarity of this operation lies in the fact that system designers have used artificial neural network based on the so-called "deep learning"

(Deep Learning). As a result, their system was able to conduct the necessary studies and calculations on the basis of very small set of input data: usually for training neural networks before using the required load in a vast array of information. Achieving Hinton team looks particularly impressive when you consider that the application form has been submitted at the last moment. Moreover, the system of "deep learning" was created in the absence of concrete data on the interaction of molecules with the proposed target. The successful application of the technique of "deep learning" was another achievement in the development of artificial intelligence, which is very rich in the second half of 2012. So, Jeff Dean this summer (Jeff Dean) and Andrew Ng (Andrew Y. Ng) from Google revealed a new system for image recognition to determine the accuracy of the cat in the picture 15.8%, where the cluster system for training of 16 thousand units to use the network ImageNet, contains 14 million images of 20,000 different objects. Last year, the Swiss scientists have demonstrated a system that is better recognized a person signs in the photos (the accuracy was 99.46% on a set of 50,000 images, with people maximum accuracy was 99.22%, and the average accuracy for a group of 32 people was "only" 98.84%). Finally, in October of this year, Richard Rashid (Richard F. Rashid) [3], coordinator of the Microsoft academic programs shown at a conference in Tianjin (China) technology live translation from English into Mandarin Chinese, maintaining the original voice. All these technologies, demonstrating a breakthrough in the field of artificial intelligence, in one way or another based on the technique of "deep learning". The main contribution to the theory of in-depth training is now making just Professor Hinton, who, incidentally, is the great-grandson of George Boole, an English scientist, the inventor of Boolean algebra underlying modern computers. The theory of deep learning complements conventional machine learning technology with special algorithms for the analysis of the input data at several levels of presentation. The peculiarity of the new approach is that the "deep learning" is studying the subject, until it finds enough informative presentation levels to take into account all the factors that could affect the characteristics of the studied subject. Technology Artificial Intelligence (AI) in general, and in-depth training in particular, are now widely used in different systems, including voice assistant Apple Siri-based Nuance Communications technologies and recognition of addresses on Google Street View. However, scientists carefully evaluate progress in this area since the history of AI is replete with big promises and no less loud crunches. Thus, in the 1960s, scientists have believed that before the establishment of full-fledged AI there are only about 10 years. Then, in the 1980s, there was a whole wave of young companies that offer a "ready-AI", then in this area has come true "Ice Age" - right up to the present time, when the enormous computing power available in the cloud services, opened a new level implementation of powerful neural networks using new theoretical and algorithmic basis. New learning paradigm realizes the idea of training in two phases. In the first stage of a large array of un-partitioned data using avtoassotsiatorov (stratified by their learning without a teacher) extracted information about the internal structure of the input data. Then, using this information in a multi-layer neural network, trained with her teacher (tagged data) by known methods. The number of untagged data desirable to be as large as possible. Markup data may be much less. In our case, this is not very important. Deep learning the most active and the most powerful tool in artificial intelligence area now. Based on the well-reviewed seen on many popular technologies of deep learning. Major corporations sent huge amounts of money in research in this area. The most popular field of application of deep learning today are the systems of computer vision, facial recognition, text, search engines and ontological system. To date, the use is gaining popularity depth training in forecasting time series. Based on this, in our further work we will carry out experiments in this direction. Today, deep learning technology based on the principles of service, which makes their use more convenient and faster in most areas. Processing information in natural language analysis of the relationship between the collection of documents and terms presented in this document, understanding and determination of the direction and theme of the text - all tasks of semantic analysis. Latent semantic analysis are the search giant to find the text of one subject. A lot of work is carried out in the field of construction of semantic models for processing quality of the text, understanding the logical relationships, optimization of knowledge bases, as well as the widest range of applications. How do children learn language? In particular, as they associate it with a sentence structure meaning? This issue is certainly associated with issue that is more global - how the brain connects the sequence of characters for building symbolic and sub-symbolic representations? In this paper, we consider a number in the field of semantic analysis using artificial neural networks applications.

III. Review of text classification approaches

Chinese scientists [4] in their study presented a new model of text classification using a neural network and its learning by back propagation, as well as with the modified method. Use an effective method to reduce the characteristics of selecting the sample dimension, that of a bear increasing system performance. Standard learning algorithm backpropagation trained quite slow, so the researchers modified the algorithm to increase the speed of learning. Traditional word-matching based on the classification of text using vector space model. However, in this approach takes no account semantic relationships between terms, which can lead to deterioration of

the classification accuracy. Latent semantic analysis can overcome the problems caused by the use of statistically derived conceptual indices and not just individual words. It creates kontsetualnye vector spaces, in which each term or a document is submitted as a vector in space. This not only significantly reduces the dimension, but also can detect important associative relations between terms. The researchers tested their model on a set of 20 news data, experimental results showed that the model with the modified method of backpropagation proposed in this paper, have surpassed the traditional teaching method. As well as the use of latent semantic analysis for this system allows you to dramatically reduce the dimension that can achieve good results classification. Indian scientists compared the performance of common neuronal back propagation network with a combination of the neural network with the method of latent semantic indexing of text classification problem. In traditional neural networks with backpropagation error process of adjusting the scales locked in a local minimum, as well as the speed of learning network of this type is rather low, which entails a decrease in performance. In view of these shortcomings, scientists have decided to make a combination of latent semantic indexing and this neural network. Latent semantic representation of the structure of data in the low-dimensional space in which the documents, the terms and words sequences were also compared. The one-dimensional decomposition technique used in latent semantic analysis, multi-dimensional matrix in which the terms are broken down into a set of K orthogonal factors, which the original text data is changed to a smaller semantic space. The new vector documents can be found in the K-dimensional space. Also, the new coordinates are queries. Performance tested a combination of these methods on the basis of the classification methodology 20 newsgroups from different categories, such as sports, medicine, business, politics, and others. As a result, these methods can significantly reduce the dimension and get the best results of the classification of the text. One of the key figures in the study of human brain is Mauntkasl [5]. In this paper, he summarized his many years of research, he argues that, despite the diversity of their functions, all sections of the cerebral cortex are arranged, in principle, the same. This means that learning and pattern recognition takes place uniformly in the cortex, and the variety of its functions is a consequence of the diversity of signals processed by different parts of the cortex. According Mauntkaslu bark has a two-dimensional honeycomb structure. The basic functional unit of the cortex is a mini-column of about 30 microns in diameter, consisting of about 100 neurons. Such mini-columns interconnected by positive and negative lateral connections. Moreover, the latter shall be sharp, but with a certain delay relative to the first. As a result, at the same time excited by a pool adjacent mini-column, involuntarily forcing the recall T. Kohonen self-organizing map [6]. As a result, throughout the cortex, we are seeing signs of self-organizing maps: detectors similar signals are located next to each other. Experiments show that the elementary detectors in the area of these maps the order of 0.1 mm 2, ie, they contain a mini-columns 102 or 104 neurons. Such functional units Mauntkasl calls the macro columns. It is they who determine the "resolution" of the cortex, and limit the number of signs that can remember people (only a few million). But the reliability of the memory is guaranteed by a large number of neurons that make up the macro column. So we keep his memory throughout life, even with the death of a substantial part of the neurons. Thus, Kohonen maps are evidently the most appropriate tool for simulating the operation of the cortex. You just need to teach them how to work with dynamic patterns, with which only the brain works, because its main task - foresight. As people acquire language, as well as two or more different languages with a nervous system remains an open question. To solve this problem, the French scientists, led by Peter blast furnace [7] proposed to build a model that will be able to learn any language from the very beginning. In this work, they offer the neural network approach, which handles proposals for like, word for word, without prior knowledge of the semantics of words. The proposed model does not "pre-bound" structure, but only random and trained compound model is based on Reservoir Computing technology. Scientists for robotic platforms through which users can teach the robot the basics of the English language, to later give him a different job, developed earlier model. In this paper, was added ability to handle rare words in order to be able to keep the dictionary size is quite small in the processing of natural language. Moreover, this approach was extended to the French language and shows that the neural network can learn two languages at the same time. Even with a small body language model is capable to learn and generalize knowledge in a monolingual or bilingual. This approach may be a more practical alternative for small shells of different languages than other training methods based on large data sets. Many studies performed in the language processing by neural networks [8], and in recent years with the use of so-called Echo State Networks [9]. How the human brain processes the proposals that a person reads or hears? The task of understanding how the brain does it, occupies a central place in the research of scientists from the field. proposals processed in real time. Previous words in a sentence may affect the processing time in hundreds of milliseconds. Recent neurophysiological studies suggest that it is the frontal part of the brain plays a crucial role in this process. Hinaut [10] conducted a study that gives some insight into how certain aspects of the treatment of the proposals in real time occur, based on the dynamics of periodic

cortical networks and plasticity in cortico-striatal system. They model the prefrontal area BA47 using recurrent neural network to obtain on-line input word categories in the processing of proposals. The system is trained on pairs of sentences in which the meaning is encoded as a function of the activation of the corresponding role of verbs and nouns in sentences. The model examines the expanded set of grammatical structures and demonstrates the ability to generate new designs. This shows how much early in the sentence a parallel set of predicates makes sense. The model shows how the online responses to the speech are influenced by the preceding words in a sentence and the previous proposals in the discourse, providing a new perspective on the neurophysiology of the brain cortex to recognize grammatical structure. The study found that recurrent neural network can decode the grammatical structure of the sentences in real-time in order to get an idea of the meaning of sentences. Neural processing of natural language. The focus in cognitive science today is focused on the study of how the neural networks in the brain are used to read and understand the text. This issue explores a huge number of scientists in neuroscience along with recent studies that are designed to examine the brain processes involved in learning non-linguistic sequences or artificial grammar learning. Peter Ford Dominey [11] in their study attempted to combine data from several neurophysiological proposals processing models, through the specification of neural network model, the architecture of which is based on the well-known cortico-striatal-thalamo-cortical (KSTK) neuroanatomy system of human language. The challenge is to develop simulation models that take into account the limitations and neuroanatomical connections, and functional image data. In the proposed model, the structural cues encoded in a recurrent neural network in Kortikova BA47, activate circuit (KSTK) for modulating the flow lekskicheskoy semantic information in an integrated view on the meaning of the sentence level. The simulation results are shown in Caplan [12]. Modeling Language Authority spent S.A. Shumsky. In the work [13], the author puts forward three hypotheses: The first hypothesis is that the processing of the time series in the crust is done like modules, recognizing the typical temporal patterns, each with its own input. For example, the bark of the site responsible for the morphological analysis of words, recognize about 105 words and their constituent morphemes and syllables. Another cortical portion defining sentence structure, operates in the same way, but with a different primary alphabet, each character is no longer encodes a letter and a word unit. This area stores the characteristic patterns of the combination of words in a grammatically correct sentence. According to the second hypothesis, the input for the next cortical module, responsible for the analysis of the temporal structures of higher order thalamus serves as a compressed output signal from the previous module. According to the third hypothesis, in the "body language" there are two interchannel "deep learning": the grammatical and semantic. Similarly, the dorsal (scene analysis) and ventral (object recognition) channels the analysis of visual information.

REFERENCES

[1] Ranzato, M. A., & Szummer, M.: Semi-supervised learning of compact document representations with deep networks. In: Proceedings of the 25th international conference on Machine learning. pp. 792-799, ACM, 2008.

[2] Hinton, Geoffrey E., Simon Osindero, and Yee-Whye: "A fast learning algorithm for deep belief nets." Neural computation 18.7: pp. 1527-1554, 2012.

[3] Deep Learning: Microsoft's Richard Rashid demos deep learning for speech recognition in China, http://deeplearning.net/2012/12/13/microsofts-richard-rashid-demos-deep-learning-for-speech-recognition-in-china/, last accessed 2020/01/17

[4] Yu B., Xu Z., Li C.: Latent semantic analysis for text categorization using neural network. //Knowledge-Based Systems, . 21, №. 8, pp. 900-904, 2008.

[5] Mountcastle V.: The columnar organization of neocortex // Brain, vol. 120, P. 701–722, 1997.

[6] Kohonen T.: Self-Organizing Maps. Springer Verlag, 2001.

[7] Hinaut X. et al.: A Recurrent Neural Network for Multiple Language Acquisition: Starting with English and French //CoCo@ NIPS, 2015.

[8] Miikkulainen, R.: Subsymbolic case-role analysis of sentences with embedded clauses. Cognitive Sci 20: pp. 47–73, 1996.

[9] Hinaut X., Dominey P. F.: Real-time parallel processing of grammatical structure in the fronto-striatal system: a recurrent network simulation study using reservoir computing //PloS one, .8, №. 2, p. 52946, 2013.

[10] Frank, S. L.: (2006). Strong systematicity in sentence processing by an Echo State Network. In Proc. of ICANN, pp. 505–514, 2006.

[11] Peter Ford Dominey, Toshio Inui, Michel Hoen: Neural network processing of natural language: II. Towards a unified model of corticostriatal function in learning sentence comprehension and non-linguistic sequencing // Brain and Language. doi:10.1016/j.bandl.2008.08.002, 2008.

[12] Caplan, D., Baker, C., Dehaut, F.: Syntactic determinants of sentence comprehension in aphasia. Cognition, 21, 117–175, 1985.

[13] Shumski, S.: Brain and language: Hypotheses about structure of a natural language, 2017.

Нейронные сети в семантическом анализе

Аверкин А.Н. Ярушев С.А.

В статье описываются глубинные архитектуры искусственных нейронных сетей и возможности их применения в семантическом анализе. Рассматривается история разработки глубинных нейронных сетей. Исследуются современные тенденции в задачах семантического анализа, а также представляется кратких обзор исследований в данной области.

Received 21.12.19