NEURAL NETWORKS ANR THEIR LEARNING

Ivanko A.F., Ivanko M.A., Sorokina P.E.

Moscow Polytechnic University, Moscow, e-mail: alekfed@mail.ru

Despite the fact that artificial neural networks became popular quite recently almost every day on the Internet we hear something new about their achievements. They are already actively used in many areas of human life. Thus, the action of such systems can be seen, for example, in online stores advertisements based on your purchases, text recognition, cell phone cameras, weather forecasts, in control applications using voice commands, and, of course, video surveillance systems simply cannot do without them. Otherwise in things without which our modern life is no longer possible. Since machines are beyond things like fatigue, diseases, and feelings that can distract ordinary people from their work, such networks learn quite quickly, efficiently, and can always show results. Maybe in some cases it will be wrong, but in time, after some learning, they will be able to minimize the possibility of mistake. The article discusses artificial neural networks, their main types of learning, and tasks to which systems trained in its various types can be applied. In the final part of the article, a conclusion regarding the current stage of development of neural networks and their future prospects is made.

The human body is built in an incredibly complex way. Consisting of living cells, tissues, organs, and perfect systems, it never ceases to amaze with its ability to be constantly updated and to resist infections and diseases. Thus, every day we make about 10 thousand blinks, 29 thousand sighs, and the heart manages to make 100 thousand beats. One of the most important systems, thanks to which all the listed things and the life we are used to are possible is nervous. Consisting of two parts, central, which includes the spinal cord and brain, and peripheral, located outside of them, it not only coordinates and reconciles the work of all organs and the rest of the systems in the human body, but is the first one that responds to any changes in the external environment and is the basis of human mental activity [1].

The human brain is a complex structure that consists of a large number of neurons – the main functional cells of the nervous system, and their numerous processes. Connected with each other, they form a unified system. As a result, such habitual actions inherent only to a person, such as the ability to reflect, and even the transformation of thoughts into conscious speech, we take as something ordinary and proper. For centuries scientists have tried to disassemble all the processes of functioning and perception of the brain. Although at first glance it seems so simple, but in fact they are still trying to finally describe all the occurring in the brain processes.

Except the processes occurring in the brain, neurons – the main functional elements of the nervous tissue of the organism that permeates the entire human body – are responsible, including for movement in space and for the correct functioning of the senses, that is, for human interaction with internal and external environment. The main functions of nerve cells are: obtaining information, for which dendrites, short cell processes are responsible; its processing, for which the neuron itself is responsible; and the subsequent conduction and transmission of information to the subsequent neurons, for which axons, the longest cell branches are responsible. When a pulse is supplied to a neuron that is sufficient to overcome the excitation threshold of a nerve cell, determined by the connected dendrites of neighboring cells, the signal is transmitted to the next neuron and the next one from it until the pulse is small enough for the neuron to ignore it. Thanks to the aforementioned processes that connect cells of the nervous system and synapses, junctions of the axon of one neuron and the dendrite of another, the nervous system forms, along which the impulses move, and the prototyping of which in the middle of the last century began to gain popularity and for many years ahead predetermined the development of computer technology.

Purpose of the study

So what is a neural network? It is a kind of learning system, created, as the name implies, in the way, similar to human brain neurons. This is a kind of its machine interpretation, which, acting on the basis of a specific predetermined algorithm, by analyzing the results of previous calculations and originally provided information, corrects the answer, minimizing the possibility of error in subsequent operations [2]. It is a kind of computing system that allows you to give the most accurate answer for given conditions.

In the early forties of the twentieth century, American artificial neural network (INS) theorist McCulloch Pitts proposed an artificial

neuron model that simply explains the basis of all neural networks, the basic functional unit of which is previously described ordinary neuron. In the simplified artificial architecture all the important functions of the original cell are saved. In that case, the synapse is characterized by its own weight and to determine the impulse needed to continue or attenuate the signal, the weight characteristic of an individual neuron is multiplied by the value of its input signal. Further, the value of each signal, due to the fact that several processes of other cells fit one neuron, is added to the other values, which happens in the so-called adder, which in our case acts as a replacement for the body of the original cell. After completing the addition, the final value enters the activation function, which compares it with the level of excitation of the neuron and decides whether or not to transmit the signal to the processes of the next cell. Of course, this model does not take into consideration many of the parameters of a normal biological neuron, which we have not previously mentioned in the description of its action, and the absence of which, according to some experts, affects the work of the artificial architecture, causing errors. For example, in a natural neuron important factors are the number of synoptic connections between two cells, how well these compounds are developed, where these compounds are located on the neuron's body, etc. But even without these parameters, artificial neural networks strongly resemble the usual nervous system.

The most commonly used neural networks today have a multi-layered structure, which allows us to achieve good results. Unlike single-layer ones, where signals from the input layer are immediately applied to the output layer, in multilayer ones everything is much more complicated. Created like the aforementioned model of artificial neural networks, they consist of an input layer that receives information, a hidden layer in which the network tries to find the correct coefficients, and an output layer that gives us the result. The existence of hidden layers at the same time and complicates the process of learning networks, and opens up great opportunities. So, in the case of their absence, the network simply remembers the input data, and as soon as it is given to the input, but slightly modified, the output signal can be very far from the truth, because the number of layers and their constituent elements directly affects the complexity of the calculated functions responsible for calculating and issuing the correct result. Further, the article speaks about networks with a multilayer structure.Research methods.

According to scientists, there are about 100 billion neurons in a person's head ($1 * 10^{11}$). Of course, this number is not entirely accurate, but it is as close to reality as possible. Even if humanity manages to create such a model, and a network with such a number of neurons will be constructed in the laboratory, the next equally important step will be to create multiple connections between them and to establish the correct weights for each individual connection, which is another problem [3, 4].

As we have said, these coefficients are needed to calculate the final function, which determines the signal level and is responsible for its subsequent transmission or attenuation. The process of finding correct coefficients is called neural network training, in result of which network performance and the quality of its output data are improved. In the very best cases, the process of adjusting the synoptic scales, or as we called them before, the coefficients, occurs every time as it goes to the correct answer. It is a training, to the consideration of which we have come a long way, and which is the main topic of the article, can occur with a teacher, and accordingly without it.

In the first case, the model is given a markup data set, it remembers the information provided, and subsequently compares it with new incoming data, and, based on the comparison, gives the answer. So, for example, a tagged data set depicting a different species of bird will teach her to predict the answer and distinguish the dove from the hummingbird in the future. This type of training is used if we have a huge amount of data, and all of them are completely reliable, which, unfortunately, not often occurs in practice. In this case, we give the system input, and the result we want it to achieve. The model must understand what result to strive for, and in the process of learning, after a long analysis, anticipate the answer.

In the second case, as well as in the previous one, the neural network is given input data for analysis, it remembers them, and comparing the criteria from the previous input data sequences with new information, it gradually develops its algorithm for predicting the answer, decreasing each time probability of error. A huge amount of information is initially loaded into the network, and the model is trying to find any regularities on its own without outside help.

Due to the fact that learning takes place in different ways, in most cases the results obtained by the first and second models will be radically different. In view of this, different tasks are given to networks with different learning methods.

For networks that have been trained in the first way, most often are given for solving the regression and classification problems. In regression problems based on the given features of the object, we want to predict any other quantitatively previously unmeasured feature. In other words, in the case of, for example, linear regression, we are trying to establish a relationship between one or several input variables and one final variable, trying to find a function that will approximate a set of initial points. In classification problems on the same input data, we are trying to predict not a quantitative, but a qualitative feature. So, initially, information about some "classes" is loaded into the system, and the network, based on examples of input data, should put incoming information into one of them [2, 5, 6].

In fact, all elements of the artificial network, in the case of this training, when input is a couple of reference input actions and desired output signals, interconnected using bidirectional channels, which makes it possible to use the inverse and forward propagation of error to find the correct result. In the first case, input data is given, then weights are corrected and a simple prediction of the answer is given, while in the second case everything is a little more complicated: like in any other training, the networks are fed with the data input, with their subsequent propagation in the direction of the outputs, and only after that, the calculation and the backward propagation of the corresponding error and the adjustment of the weights occur. It looks like a decomposition into a Taylor series, only we are trying to optimize not all the members of a function, but only the very first one [7, 8].

We now turn to learning without a teacher. For it, the most popular systems use tasks of:

– clustering. Here, comparing the input data, the model organizes and divides them into special groups – clusters. Similar data by the selected parameters are recorded in one cluster, similar in other parameters – to another. There is something similar to the classification task mentioned above, but it should be understood that in this case the network itself determines the groups into which the input data is divided, whereas in the previous one they were incorporated into the system initially.

– detection of anomalies. As you can see, here the system is looking for objects, the signs of which seem to it to be very different from the signs specified at the beginning of the program, or from the signs that the neural network has identified during work. The characteristics of these objects are not registered in the training set, and their detection becomes the main goal of this method.

– searching for association rules – when the system examines the input information and, on its basis, tries to recommend data that is similar to it according to some criteria. This is one of the most popular tasks in many online stores, when on the basis of purchases made, the system offers products similar to them.

– reduction of the dimension, or, as they are otherwise called, autoencoders, which encode the input information, and then decode it, receiving data as close as possible to the original information, which allows to reduce the amount of information while maintaining its basic positions [3, 4, 6].

Except of these two methods, of course, there exist all sorts of their combinations. Here it is necessary to mention training with reinforcements. To some extent, it combines the previous two methods. So, reminding learning without a teacher, an unallocated data set is given to the system, and it also tries to predict the correct answer itself, but when it finds it, we intervene and in some way encourage our model. The system itself chooses which answer to come to, but it stops only after finding the right option and receiving a reward. The network eventually learns from its mistakes, and better finds the right path and the appropriate answer.

In addition to everything listed in any network is laid a so-called loss function, which calculates the difference between real and received answers. It gives an efficiency mark of the quality of work, and shows how one or another weights are suitable for solving this problem. Gradual minimization of errors, respectively, leads to improved and closest to real output data. It turns out that the smaller this function, the better result the selected model will give. The simplest and most known loss function is the standard deviation, which shows the average degree of variation in magnitudes relative to its expectation [7].

Another of the important functions, which was already mentioned in passing earlier, is the activation function, which decides which number to transfer further. After counting all the signals, it should convert the received number to another one, processed by some pledged function, and already issue it as the result of the program. As you can understand, different activation functions are used for different tasks. One of the simplest functions is the single-hop function, the result of which can be only one of two integers: one and zero, which respectively mean positive and negative results. Here the

function compares the input value with some fixed constant, and simply says more or less of it was the number supplied. There is also a huge number of sigmoidal functions in which the input value is compared with the number obtained from the calculation of a complex formula. Among them, the logistic function and the hyperbolic tangent function are among the most popular. Here the network is not limited to two integers, and can give, in the first case, a fractional number in the range from 0 to 1, and in the second, a fractional number from -1 to 1. Despite the fact that the network should automate all the factors needed to find the correct output values, some of them are manually configured. For example, one of these important parameters is the speed of its learning, because if you choose too small number, the system will go to the answer for a very long time, whereas in the opposite case, it may not notice the correct answer and skip it, or find it, but remote from the arithmetic mean by a number close to the value of its error, which, of course, is also the correct answer, but which, with the correct value of speed, can be found and more precisely. These so-called hyper-mamarameters, which, unfortunately, have to be guessed and corrected manually over and over again, also include the number of hidden layers and elements that make them up, which affects the complexity of the activation function mentioned in the previous paragraph and how its output will be close to real [8].

Scientists are going through all sorts of neural network architectures, trying to teach each of them. Based on this, they draw conclusions regarding a particular network or method of its training. Finding the right network architecture and its training are the main problems in creating a properly functioning neural network. As practice shows, small narrowly specialized networks, which are sharpened for one task, give a better result than those that have several specified goals. So in many sci-fi works, after creating a fully working neural network to facilitate a person's life, something is not going according to plan and leads to terrible consequences [9,10]. It would seem that the whole process should go through without various types of problems, but in fact, due to the fact that a wide range of possibilities may incorrectly influence the finding of weights and the creation of incorrect connections within the network, the neural network leads to incorrect results.

Conclusion

Of course, any cars have their drawbacks. For example, we cannot explain why a neural network will give one answer or another, and how it came to it. We cannot even guarantee, with an absolute probability, the unambiguity of obtaining a result when re-entering the same data. But, on the other hand, the same network can easily adapt to changes in the environment, solve problems at an unusually high speed, and with unknown regularities of the input data. They are also potentially fault-tolerant – if suddenly the work in any neuron does not go according to plan – the performance practically does not decrease. It is better, of course, to correct the mistake, but if there is no such possibility, then it is possible to continue work under such conditions [11, 12].

Progress does not stand still and inexorably moves us forward. There is no sphere of human life in which neither now nor in the future can one imagine the replacement of human labor by machine labor. It is already clear that soon artificial neural networks will serve us well. They will be able to automate many processes, reach places that cannot be reached by people, help scientists, and simplify our lives.

References

1. Human nervous system [Electronic resource] / Electron. Dan. Access mode: https://fit-baza.com/nervnaya-sistema-cheloveka/#i-2.

2. Training a neural network with a teacher, without a teacher, with reinforcement – what is the difference? Which algorithm is better? [Electronic resource] / Electron. Dan. Access mode: https://neurohive.io/ru/osnovy-data-science/obuchenie-s-uchitelem-bez-uchitelja-s-podkrepleniem/.

3. Chapter 3. Basics of ANN [Electronic resource] / Electron. Dan. – Access mode: https://neuralnet.info/chap ter/%D0%BE%D1%81%D0%BD%D0%BE%D0%B2%D1%8B-%D0%B8%D0%BD%D1%81 81 /.

4. How the neural network works: algorithms, training, activation functions and losses [Electronic resource] / Electron. Dan. Access mode: https://neurohive.io/ru/osnovy-data-science/osnovy-nejronnyh-setej-algoritmy-obuchenie-funkcii-aktivacii-i-poteri/.

5. Neural networks for beginners. Part 1. [Electronic resource] / Electron. Dan. Access mode: https://habr.com/ru/post/312450/.

6. Neural networks for beginners. Part 2 [Electronic resource] / Electron. Dan. Access mode: https://habr.com/ru/post/313216/.

7. Mean square deviation [Electronic resource] / Electron. Dan. Access mode: https://wiki.loginom.ru/articles/mean-square-deviation.html.

8. Building neural networks of direct distribution using adaptive elements [Electronic resource] / Electron. Dan. Access mode: http://jre.cplire.ru/jre/aug12/4/text.html.

9. The advantages of neural networks [Electronic resource] / Electron. Dan. Access mode: http://www.aiportal.ru/articles/neural-networks/advantages.html.

10. Advantages and disadvantages of neural networks as a means for processing knowledge [Electronic resource] / Electron. Dan. Access mode: https://studopedia.org/9-161389.html.

11. Ivanko A.F., Ivanko M.A., Kulikova E.V. Moscow Polytechnic University, Moscow, Security of information data. European Journal of Natural History. 2018. № 4. P. 118–120.

12. Ivanko A.F., Ivanko M.A., Vinokur A.I., Kulikova E.V. Computer-based artificial intelligence systems. European Journal of Natural History. 2019. № 1. P. 49–53.