Problems of neural network theory in their perspective

Galushkin A.I.
Scientific Center of Neurocomputers, Moscow
E-mail: scn@mail.cnt.ru

Abstract. The author’s viewpoint on the main perspective lines in the development of the neural network theory in the domain of the neural network algorithms for different mathematical tasks’ solutions is presented. The main objective is to analyse the neural network theory aspects that would improve neural network task solution algorithms resulting by the increase in the equivalent capacity-to-cost ratio of the computational systems. The perspective problems of the neural network theory are considered in the order, structure and denotations corresponding to the presently developed technique of the multi-layer neural network synthesis [1-3].

Introduction

The neural network theory presents the logical basis for the solution of the tasks of general and applied mathematics in the same way as earlier Boolean logic was the basis for the solution of the tasks by the computers with Neumann architecture.

The main reasons to write the present article are the following:
- the primitiveness of the neural network algorithm that are used on the initial stages of the tasks’ solution (as a rule, after some solution quality is achieved, there are no methods for its further improvement);
- the necessity for the elaboration of the neural network algorithms adequate to the task under consideration in the framework of some unified tasks solution technique in the neural network logical basis.

1. Typical input signals

The elaboration of the typical neural network input signal classes is necessary for the objective test of the adaptive neural network performance quality. At that, each input signal class is always specific for the specific solution task. The class of signals with Laplace transformation \{1, 1/s, 1/s^2, 1/s^3, \ldots\} is a typical example of the neural network input signal class. In this case, the test of the control system performance quality is carried out by the feed of the corresponding typical signals (6-function, unit step, linear signal, \ldots) to the system input with the following analysis of the transient process and of the error in the terminal steady state (the order of the control system astaticism).

The typical neural network input signal class must always possess some parameter characterizing the complexity of the solution task. This parameter is apparent in the aforementioned example. As far back as the beginning of 1960s, for the pattern recognition tasks oriented onto the performance with the random samples of complex unknown multi-modal distributions, the random samples of multi-modal distributions were suggested as typical neural network input signals. At that, the distribution modes represented a normal distribution and the mode centers were situated along the hyperbisector of the multidimensional space of indications alternating for each class [2, 3]. Two parameters were taken as the solution task complexity indexes, namely the number of the distribution modes and the variance of each separate mode. The indication of the modes belonging to the specific class in the case of self-learning neural network performance mode were absent.

The selection of the signal \(|\sin nx|\) (\(n\) is the index of complexity) to solve the task of forecasting the number of calls in a month at call-centers [4] is an example of the elaboration of typical multi-layer neural network input signals’ class.

The selection of the typical neural network input signals class is an important task for the researcher who desires to demonstrate, more or less objectively, the advantages of his neural network algorithm elaborated for the solution of the specific formulated task.

2. A priori probabilities of classes’ emergence

The necessity of taking into account the a priori probabilities of classes’ emergence appears in the different practical tasks. A typical example is the task of the letters recognition in the printed text of the scanned document in the case when the probability of any letter appearance is known. The possibility to use the a priori probabilities of the classes’ emergence during the adjustment of the multi-layer systems for the image recognition is investigated in the study [2]. This possibility was permanently used in [2] for the construction of different concrete systems. However, this technique requires an additional investigation in order to use it efficiently.

3. Matrices of costs for errors
As a rule, in the known studies, matrices of costs for errors that emerge in the course assignment of the image to this or that class are assumed to be diagonal. However, it is not often in agreement with reality. For example, in the case of the neurocomputer design for the mine recognition system with the use of the geolocator, in the matrix of costs for errors

\[ L = \begin{bmatrix} l_{11} & l_{12} \\ l_{21} & l_{22} \end{bmatrix} \]

the coefficients \( l_{21} \) (the cost for the error to take the mine as an irrelevant object) and \( l_{22} \) (the cost for the error to take the irrelevant object as a mine) cannot be equal in principle, and it must be taken into account during the adjustment of the multi-layer neural network in the similar way as it was carried out in the study [2].

4. The initial conditions for the adaptive adjustment of the weight coefficients

The choice of the initial weight coefficients of the adaptive neural network is an important condition to speed up the task solution procedure. Therefore, from our point of view, the widespread approach to choose the zero values for the weight coefficients or the random values with the uniform distribution in the given turn-down as the initial conditions is incorrect.

Even during the solution of the problem of recognition of two patterns' classes, it was clear that the initial weight coefficient values must be selected by generating the dividing surface configuration, implemented by the neural network, in the form of a multidimensional “chess-board” with the uniform distribution of “black” and “white” squares. Each color corresponds to the first and second classes of patterns in the physically implemented multidimensional space of indications [2, 3]. Such a multidimensional “chess-board” is formed by the hyperplanes corresponding to the neurons of the first layer.

The initial weight coefficients of the adaptive neural network can also be chosen as the weight coefficients of the neural network with variable structure after its learning stage termination.

The task of function approximation (extrapolation) is an example of the effective solution of the problem of the initial weight coefficients’ selection for the multi-layer neural network adjustment. In this case, assuming that the neural network is an effective nonlinear filter (extrapolator), it is expedient to choose the weight coefficients of the neural network implementing the equivalent nonlinear filter or extrapolator [5] as the initial conditions.

Hence, the choice of the initial conditions for the multi-layer neural network adjustment possesses the following properties:

- It is specific for each specific task that is solved by the neural network.
- It is aimed at the acceleration of the adjustment process (and therefore on the acceleration of the task solution) by putting the neural network into the domain of the global extremum of the secondary optimization functional.
- As a result, it allows one to increase the equivalent ratio between productivity and the cost during the specific task solution.

5. The control of parameters in the extremum searching procedure for the multi-extremum secondary optimization functional

The gradient local extremum searching procedure for the multi-extremum optimization functional is a very important element of the multi-layer neural network adjustment algorithm. In the simplest case, the weight coefficient \( K^* \) at the functional gradient is determined in the empirical way in the process of solution of each specific task and is left constant in the adjustment procedure. Since 1960s, researches have tried to make coefficient \( K^* \) be variable (decreasing) with time during the adjustment procedure. This was done in order to decrease the adjustment error in the steady state [10]. But this led to the significantly sharp increase in the adjustment duration (in the transient process of the multi-layer neural network adjustment). At present, a considerable part of this problem remains open. Some efforts are made to control the value of the coefficient \( K^* \) by the current error value and by the gradient functional [11].

6. The modifications of the global extremum searching algorithms for the multi-extremal secondary optimization functional

The secondary optimization functional in the multi-layer neural network is multi-extremal by definition. The reasons to this are the following:

- The input signal is rather complex (for example, the distribution of patterns aggregate in the multi-dimensional space of indications at patterns' recognition task solution is multi-modal).
- The variants of the task solution are multivariate.
- The open neural network structure is flexible.
The searching methods for the global extremum (or, in addition, for several local extremums) are at present only in the stage of formation.

One of such methods related to the set of ejections of the random initial conditions in the space of the neural network weight coefficients and to the search of the global extremum for the secondary optimization functional was presented and investigated in [2]. The convergence of the presented procedure by the number of initial conditions under the fixed number of local extremums was shown there. Some methods of this procedure resulting in the time decrease for the global extremum search are known. The method of “annealing” is an example.

The gradient algorithm is effective to find the local minimums in the case of neural network weights adjustment during the procedure of learning. In general, the mixture of the “annealing” and gradient methods is the most effective algorithm. First, the “annealing” algorithm is used to find the initial weights. The gradient descent algorithm is used thereafter to bring the system to the nearest local minimum. Then the “annealing” algorithm is used at this point again in order to leave this local minimum. These stages are repeated until one gets the possibility to leave the recurrent local minimum. In the latter case, it can be considered that the global minimum is obtained.

7. Filtering and extrapolation of the signal corresponding to the estimation of the secondary optimization functional gradient

As a rule, the decision to change the weight coefficients in the known neural network adaptation algorithms is taken on each operation cycle according to the results of one single pattern passing through the network. The experience of using the filter with the memory $m \neq 1$ [2] in the adjustment circuit showed the increase in the adjustment effectiveness for stationary and non-stationary patterns at the multi-layer neural network input. Some attempts to speed up the learning process through the application of the weight coefficients’ extrapolation procedure during the neural network adjustment are known from the literature [12]. The filter synthesis in the weight coefficients’ adjustment circuit is poorly investigated though it is a perspective algorithm in the general procedure of the multi-layer neural network synthesis.

8. The multi-layer neural network adaptation algorithms with the adjustment of the coefficients for the “slope” of the activation function

When using the activation function with a variable “slope”, the separate neuron in the neural network is described by the expression

$$y = \frac{2}{\pi} \arctg B \sum_{i=0}^{B} a_i x_i = \frac{2}{\pi} \arctg \left( B a_i \right) x_i$$

It is seen therefore that there is no sense to organize the adjustment circuit for the coefficients $B$ and $a_i$ at the constructing of adaptation algorithms. In the neural network consisting of the neuron set that contains a separate neuron subset or all the sets with the activation function of the same “slope”, the organization of the adjustment circuit for the coefficient $B$, as well as for the separate coefficients, is necessary in order to decrease the whole adjustment duration, i.e. to decrease the time required for the task solution.

9. Selection and analysis of the neural networks with the special structures adequate to the class of solution tasks

If the class of solution tasks allows one to define the class of effective neural network structures adequate to these tasks, then the elaboration of the special methods for the coefficient adjustment namely for this class of neural network structures will increase the adaptation effectiveness for the task solution of this specific class. We describe below some variants of neural network structures and the task classes adequate to them.

Cell-like neural network. Cell-like neural networks are the networks with the special topological structure that is adequate, in particular, to the task of patterns’ processing. In this case, the natural task parallelism results in the natural parallelism in the structure organization of the processing neural network.

Continual neural network. Continual neural networks [3] are used mainly in two cases:

- when the number of indications in the layer is large;
- when the signal or pattern processing is performed on a real time basis and without preliminary quantification of the input information.

Neural networks with lateral connections. This specific type of connections between layers in the multi-layer neural network is interesting from the viewpoint of implementation of the invariance to the transformation group and has been poorly investigated.

Neural networks with variable structure. Since 1960s, the multi-layer neural networks with variable structure have been an effective tool to solve the task of pattern recognition [1]. This is the variant of the
adjustment algorithms, which neural network structure (the number of neurons in the layers and the number of layers) grows in the process of adjustment up to a certain value of the solution quality index. The synthesis of the adjustment algorithms for the multi-layer neural network with variable structure is a promising technique to solve a wide range of practical tasks.

**Neural networks with random connections.** Already Rosenblat in his classical monograph [9], when suggesting to introduce the random connections between the retina and the first layer of the multi-layer neural network, noted that under the proper increase in the number of neurons, the systems’ reliability, related to the possible break-down of several neurons, increases. At present and in perspective of the development of microelectronic technology, the number of emulated neurons inside the super-large scale integration, on the board, and in the unit is quite large. And this number will continue to increase in the future. This makes the variant of the random connections more and more necessary for the implementation and research.

**Neural networks with feedback loops.** The conception of the neural network with feedback loops in its classical sense was introduced in [2]. In this case, the feedback channels are present in the structure of the adjusted multi-layer neural network. These channels are used for the transmission of the intermediate and output signals of the neural network to the input channels of the previous layers through the delay lines (for a given number of cycles). In the past, it was considered that such neural network structures could be used only for the design of the memory units of a different functional destination [13]. In the last 10-15 years, in foreign literature, the investigations of such neural networks, conventionally called recurrent ones [14], was sharply activated. At that, the range of their application increased including the tasks of function approximation and extrapolation and the systems’ dynamic control tasks.

**10. The activation function selection**

The activation function selection is an important element of the neural network synthesis procedure. More than ten types of the neuron activation function are described in the known literature. Usually their selection is performed arbitrarily. In the studies [1,2] the activation function $y = \frac{2}{\pi} \arctg Bg$ was used, where $g$ – is the analogous output signal of the neuron. At the end of $80^{th}$ - beginning of $90^{th}$, the sigmoid activation function became widely used. Wavelet networks are in fact the neural network with the activation function of a particular type.

As a rule, the introducing of the new peculiar type of activation function represents an attempt to make the neural network be more adequate to the solution task in order to decrease the number of neurons and adjusted coefficients. However, the desired goal is not always achieved because the task of the neural network adjustment simplification is also desired, in addition to the task of the simplification for the calculations of the output signal from the input signal. It is necessary to note that, the complication of the activation function results in the sufficiently sharp complication of the adjustment algorithm due to the fact that the computational units for the activation function derivatives’ calculations are used in the adjustment algorithm.

**11. The tasks related to supervisor qualification**

The elaboration of the neural network algorithms adequate to the real conditions of getting information for their adjustment requires an estimation of real supervisor qualification. Due to this requirement, along with the widespread learning modes of the multi-layer neural network in which it is assumed that the supervisor is aware about the patterns’ belonging to this or that class with unit probability, one must consider in more details three more learning modes introduced in [2]:  
- learning with supervisor having zero qualification (self-learning, clasterization);  
- learning with supervisor having finite qualification;  
- learning with supervisor having negative qualification (the “harm” mode in which the supervisor wittingly gives false information about the belonging of the pattern to this or that class).

**12. The invariance to the transformation group in the multi-layer neural networks**

The simplest and the most widespread subject of investigations today, but very far from completeness, is the neural network implementation of rotation, transition on the plane and the change of the pattern scale. When applied to the pattern, from the viewpoint of the task formulation, this is the simplest formalizable affinity transformation. The solution of the task to ensure the invariance to this transformation group is not yet completely found. However even at present, the neural network synthesis task to ensure the invariance to the more complex transformation groups, often non-formalizable, arises. The elements of mimicry in the task of face recognition by the portrait is an example. The task to ensure the
invariance can emerge also in the signal processing, for example to provide the sense of word recognition independently of the announcer, etc.

13. The adaptive control of the number of digits in weight coefficients during the process of neural network adjustment and functioning

The problem of choice of the number of digits in the weight coefficients during signal or pattern processing with information representation by different number of digits is very important and self-dependent. A low digit number in the weight coefficients results in the emergence of additional errors, and a large one leads to an increase in the system cost and in the process time consuming. The problem of the programmable variation of the weight coefficient digit numbers in the linear filters is maximally correctly solved at the hardware realization of Z-filters in super large-scale integration IMS A100 of Inmos firm as far back as 1986. In these super large-scale integrated circuits, the weight coefficient digit numbers of Z-filter are software programmable and can be changed by the user in the range 2-16 with a corresponding change in the processing speed. In the neural networks implementing nonlinear filters, the number of digits can be changed adaptively during the neural network functioning process according to some specified criterion, including the index of the task solution precision and the index of the realization complexity (the cost), or according to the duration of calculations. The problem of the development of adaptive algorithms for the digit number control and its hardware realization by super large-scale integrated circuits is the subject for future investigations.

14. The investigation of the neural network reliability

The property of gradual rather than catastrophic (as in the systems realized on the boolean logic elements) neural network degradation in the course of separate elements' going out of service is an important subject of investigations. The change of the solution task quality as a function of the degree and kind of breakdowns in the neural network structure is an example.

15. The neural network diagnostics

The elaboration of the special test system for the neural networks with different structures is an important element for the increase in the reliability of perspective neurocomputers. The development of such tests on the logical level is a branch of the perspective investigations in the neural network theory domain.

16. The parallelization of the neural network algorithms to different types of switching cores in super neurocomputers

The task of the designing of super neurocomputers will be evidently the main task in the domain of neurocomputers' technical implementation, and that is a promising domain of super computers. The super neurocomputer being a calculation system with massive parallelism will represent an aggregate of hosts, of some switching core and of neuroprocessors realizing hardware- or firmware-designed neural network fragments. The elaborated neural network task solution algorithm requires the parallelization on the number of neuroformations which operates with a specified switching core.

Additionally, some types of special neural network structures are also considered.

For each of neural network architectures, a specific parallelization technique must be designed corresponding to the given type of the switching core implemented in the super neurocomputer.

The parallelization algorithms and programs elaborated for the specific neural network architecture types must be the supplement to the operational system of the super neurocomputer.

It is necessary to note that for each type of the super neurocomputer switching core, one can synthesize the neural network architectures providing the minimal or predetermined loading for the switching core.

17. The analytical description of closed neural networks

We consider a neural network with a weight coefficient adjustment circuit as a closed neural network. It represents a complex nonlinear dynamic system with a rather complicated analytical description. The attempts to investigate analytically the dynamics of such systems were made repeatedly [17, 18, 19]. At present, this problem is poorly studied, and thus it is a challenging task for future researchers.

18. The problem of the small-size learning sample

The problem of the small learning sample in the multi-layer neural networks has a great practical sense and it was repeatedly investigated in numerous studies. For example, in [7, 8], the performance of neural network pattern recognition systems with the small-size learning sample is considered. The comparison with the non-parametric statistical classifiers under the small sample conditions is done in terms of recognition probability error. The presented experimental results make it possible to
substantiate the choice of this or that method under the constraints for the size of the learning sample and under the requirements for the recognition precision and reliability. It is shown in [8] that the effective tool of the errors’ elimination under the small-size learning sample conditions is the method based on the combination of different neural network architectures.

The further development of the approaches suggested in [1, 2] is realized in [3, 9] using the results of investigations of the multi-layer neural network adaptation. The problem raised in [3] relates to one of the domains of the general technique of multi-layer neural network adaptation algorithms that is called planning of experiments. The scientifically substantiated implementation of this approach makes it possible to select either the value or the range of parameter values for the multi-layer neural networks and for their adaptation algorithms that cannot be selected unambiguously on the previous stages of the synthesis [3, 9].

Conclusion

Unfortunately, perhaps due to the poor information distribution, a large number of “home-bred” neural network algorithms emerges. The achievement of the first positive results on the basis of these algorithms can provoke an illusion about the “completeness” of the neural network theory. However, this theory is only in the initial phase of its development. Evidently in the present work, not all the list of neural network theory problems is enumerated. The gradual progress in this domain must improve the solution of the vast number of existent tasks as well as pose new tasks.

References