

Determination of Structural Parameters of Multilayer Perceptron Designed to Estimate Parameters of Technical Systems

Zhengbing Hu

School of Educational Information Technology, Central China Normal University, Wuhan, China
E-mail: hzb@mail.ccnu.edu.cn

Igor A. Tereykovskiy

Faculty of Applied Mathematics, National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Kyiv, Ukraine
E-mail: terejkowski@ukr.net

Lyudmila O. Tereykovska

Kyiv National University of Construction and Architecture, Kyiv, Ukraine
E-mail: terejkowski@ukr.net

Volodymyr V. Pogorelov

Faculty of Applied Mathematics, National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Kyiv, Ukraine
E-mail: volodymyr.pogorelov@gmail.com

Received: 07 February 2017; Accepted: 08 March 2017; Published: 08 October 2017

Abstract—The paper is dedicated to the problem of efficiency increasing in case of applying multilayer perceptron in context of parameters estimation for technical systems. It is shown that the increase of efficiency is possible by adaptation of structure of the multilayer perceptron to the problem specification set. It is revealed that the structure adaptation lies in the determination the following parameters:

1. The number of hidden neuron layers;
2. The number of neurons within each layer.

In terms of the paper, we introduce mathematical apparatus that allows conducting the structure adaptation for minimization of the relative error of the neuro-network model generalization. A numerical experiment to demonstrate efficiency of the mathematical apparatus was developed and described in terms of the article. Further research in this sphere lies in the development of a method for calculation of optimum relationship between the number of the hidden neuron layers and the number of hidden neurons within each layer.

Index Terms—Neuro-network model generalization, Structure of multilayer perceptron, Hidden neuron layer, Hidden neuron, Structure adaptation.

I. INTRODUCTION

One of the most significant spheres of the development of today's technical systems for controlling and diagnosing of the objects of various designations is the improvement of the mathematical apparatus that is used for the estimation of the controlled parameters [2, 3, 8]. It is believed that the large perspectives of such the improvement are related to the use of the neuro-network models (NNM) [1, 6]. Prerequisite of the use is established NNM's ability to solve effectively weakly formalized problems. This, in addition to the rest of the facts, allows significant simplification and shortening of the process of the development of mentioned controlling and diagnosing systems. Although, rather a large number of various NNM types are currently in use, multilayer Perceptron (MLP) is used commonly as basic model. For instance, MLP lies in the basis of deep neuron networks used by Google and Facebook companies for the recognition of both voice signals and pictures. In addition to this, the mentioned NNM type unambiguously demonstrates its prospects for estimation of controlled parameters for:

1. Cyber-attacks recognition;
2. Voice signals;
3. Technical conditions of building constructions [1--4].

At the same time, practical experience of using of the neuro-network recognition systems as well as results of the researches [2, 3, 7] shows that it is necessary to increase efficiency of MLP application. This goal can be reached by adaptation of MLP structure to the problem specifications set. This statement is based on the fact that the MLP structure influences significantly its generalization error. In its turn, the application efficiency principally depends generalization error.

II. ANALISYS OF EXISTING PUBLICATIONS

Generally, MLP (shown on Fig. 1) corresponds to neuro-network that consists of a number of layers. Each layer, in its turn, consists of artificial neurons that are

connected in series [4, 5].

We should note that in case of MLP shown on Fig. 1 numbers of input and output neurons equals to N_x and N_y , number of hidden neurons equals to H , and numbers of those in the first, h -th and the last hidden layers are J_1, J_h and J_H , respectively. We are going to consider classic MLP that has no inverse links. In this model, each neuron within the hidden layer receives all the output signals from the preceding layer, and its output signal can be transmitted to any neuron belonging to the next layer. Linear activation function is used for the neurons within the input and output layers. Within the hidden neurons logistic sigmoidal function is used.

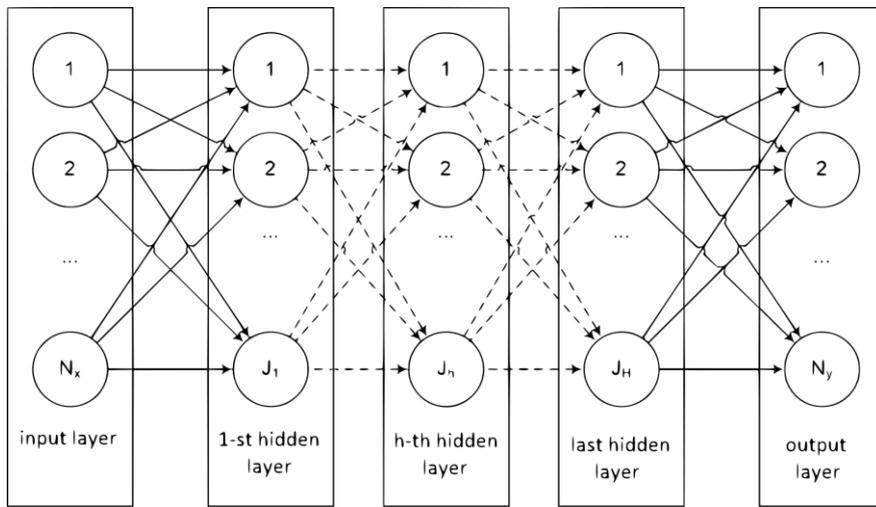


Fig.1. Structure of MLP.

Principal parameters to define MLP structure are number of the input neurons, number of hidden layers of neurons, number of neurons within each hidden layer and number of output neurons.

Based on the results of [1, 4, 7] we can suppose within the first approximation that the number of input and output neurons corresponds to the number of the parameters being registered as well as the number of the objects recognized. These values are specified a priori. They are based upon the problem specifications set and are not subject to any change. Therefore, MLP structure can be optimized but at the expense of the number of the hidden layers of neurons as well as the number of the hidden neurons within each of the layers. The following calculation expressions are submitted in [1, 3]:

$$\frac{N_y \times P}{1 + \log_2 P} \leq L_w \leq N_x \times \left(\frac{P}{N_x} + 1 \right) \times (N_x + N_y + 1) + N_x, \quad (1)$$

$$\frac{P}{10} - N_x - N_y \leq L_w \leq \frac{P}{2} - N_x - N_y, \quad (2)$$

$$L_w < P \times \epsilon_{\max}, \quad (3)$$

where N_x, N_y are the numbers of input and output neurons;

L_w is the number of synaptic bonds;

P is the number of the training examples;

ϵ_{\max} is the maximum allowable recognition error.

But practical experience and results [5] shows that the available solutions of (1-3) require significant improvements.

It is reasonable to highlight papers [1, 4] in which some expressions were derived for the determination of optimum number of hidden neurons for the two-layered perceptron:

$$N_1^{opt}(\min) = Round \left(\begin{matrix} (0,4N_x + 0,2) \times \\ \times \frac{\sqrt{PN_x}}{N_x + N_y} \end{matrix} \right), \quad (4)$$

$$N_1^{opt}(\max) = Round\left(\frac{2\sqrt{PN_x}}{N_y}\right), \quad (5)$$

where $N_1^{opt}(\max)$ and $N_1^{opt}(\min)$ are maximum and minimum boundary values of the optimum number of hidden neurons;

$Round(X)$ is the operation for the determination of the nearest integer number derived of argument P .

Use of (4, 5) restrict the search area of the optimum number of the hidden neurons, but the area of use of these two expressions is limited by two-layered Perceptron. Besides to this, additional limitations are imposed to it with the following prerequisite:

$$P \geq 10N_x. \quad (6)$$

The prerequisite has too general nature and does not take into account the amount of the classes for recognition, that should be represented in the training sample.

We shall note as well that many papers denoted to the MLP application for the estimation of the parameters of the technical systems do not contain substantiated method of calculation of its structural parameters. Even approaches for the determination of the best parameters of the MLP structures contain some contradictions. Thus, they use principle of minimization of synaptic bonds in [1] which are necessary for the network training on the specified set of examples in [1]. But the specified principles contradict [7] where they suggest that the reduction of the network size does not lead to the increase of its generalization possibilities.

At the same time, general specific feature of the papers that have been analyzed is the fact that the main conditions of problem domain that influences MLP's structure are:

1. The number of input and output parameters;
2. The amount of training sample;
3. The maximum allowable recognition error.

So, the purpose of this research is to develop a procedure for the determination of the MLP structure parameters and determination of the number of these parameters.

III. DEVELOPMENT OF THE MATHEMATICAL APPARATUS

According to [3, 6] development of the MLP structural model is based on the determination of the total number of synaptic links. We are going to take minimization of the minimum relative MLP error criterion as base of this determination according to [6, 7]:

According to [3, 6] development of the structural MLP model is based on the determination of the total number of synaptic links.

$$G_{MLP} \rightarrow \min, \quad (7)$$

where G_{MLP} is the relative MLP error.

In terms of the paper, the term, relative MLP error means division of generalization error (ε) means by the number of synaptic links (L_w):

$$G_{MLP} = \frac{\varepsilon}{L_w}. \quad (8)$$

We should note that the generalization error of the neuro-network corresponds to the error of the recognition of the parameters which were not used for the MLP training. In other words, this is the error of the recognition of the parameters which are not constituent part of the training, test, or validation samples. Rewriting (7) taking into account (8) we obtain the following:

$$\frac{\varepsilon}{L_w} \rightarrow \min. \quad (9)$$

We should note that for the full-linked MLP the following is true:

$$L_w = N_x \times J_1 + J_1 \times J_2 + \dots + J_{h-1} \times J_h + \dots + J_{H-1} \times J_H + J_H \times N_y \quad (10)$$

In case, when number of neurons in each of the hidden layers is the same, it is possible to transform (10) as follows:

$$L_w = (H-1) \times J^2 + (N_x + N_y) \times J, \quad (11)$$

where J is the number of neurons in a hidden layer;

H is the number of hidden layers of neurons.

Taking into account the results of [1, 4] we should suppose that when solving (9), it is possible to state that the generalization error and the number of synaptic links are continuous quantities. The assumption allows determining of optimum number of synaptic links (L_w^{opt}) basing upon the solution of the following equation:

$$\frac{\partial L_w}{\partial \varepsilon} = 0. \quad (12)$$

Let us detail constituent equation (12). Generalization error is calculated as sum of the approximation error (ε_a) and model description error (ε_o)

$$\varepsilon = \varepsilon_a + \varepsilon_o. \quad (13)$$

We should mention that the approximation error relates to the memorizing of the training data by MLP, but description one relates to the generalization of these data.

According to [1, 7] we can assume that the approximation error is proportional to the ratio of the sum of input and output neurons and number of synaptic links:

$$\varepsilon_a \sim \frac{N_x + N_y}{L_w}. \quad (14)$$

Error of the model description is calculated as follows:

$$\varepsilon_o \sim \frac{L_w}{P}, \quad (15)$$

where P is the number of training examples.

Upon substitution of (14, 15) in (13) we derive the following:

$$\varepsilon \sim \left(\frac{N_x + N_y}{L_w} + \frac{L_w}{P} \right). \quad (16)$$

Substituting (16) in (12) we can derive the following upon conducting trivial conversions:

$$L_w^{opt} \sim \sqrt{P \times (N_x + N_y)}, \quad (17)$$

where L_w^{opt} is optimum number of synaptic bonds of MLP.

Taking into account (11) we can put down proportion (17) in the following form:

$$\left((H^{opt} - 1) \times J_w^{opt^2} + (N_x + N_y) \times J_w^{opt} \right) \sim \sqrt{P \times (N_x + N_y)}, \quad (18)$$

where J_w^{opt} is optimum number of neurons within a single hidden layer;

H^{opt} is optimum number of the hidden layers of neurons.

Introduction of the proportionality factor allows writing (18) in the form of the following quadratic equation:

$$\left((H^{opt} - 1) \times J_w^{opt^2} + (N_x + N_y) \times J_w^{opt} \right) - k \sqrt{P \times (N_x + N_y)} = 0, \quad (19)$$

where k is some proportionally factor.

We should note that the solution of equation (19) foresees determination of the two unknown values of J_w^{opt} and H^{opt} . In addition to this, we shall estimate the magnitude of k factor preliminarily.

Let us make use of expression [4] which suggests that

under equal calculation possibilities number of synaptic links of MLP is somewhat lower than the synaptic bonds of the two-layered perceptron. It allows assuming of the following in the first approximation:

$$L_d = L_w, \quad (20)$$

where L_d is the number of synaptic links of the two-layered perceptron.

The number of the synaptic bonds of the two-layered perceptron is determined as follows:

$$L_d = (N_x + N_y) \times N_1, \quad (21)$$

where N_1 is the number of the hidden neurons within the two-layered perceptron.

Taking (20, 21) into account we can rewrite expression (19) as follows:

$$(N_x + N_y) \times N_1^{opt} - k \sqrt{P \times (N_x + N_y)} = 0, \quad (22)$$

where N_1^{opt} is the optimum number of the hidden neurons within the two-layered perceptron.

Determination of optimum number of the hidden neurons is reduced to the following calculation:

$$N_1^{opt} = k \times \frac{\sqrt{P \times (N_x + N_y)}}{(N_x + N_y)}. \quad (23)$$

We are going to use Hecht-Nilsen theorem for the determination of factor for which it was proved that minimum number of neurons equals to:

$$N_1^{\min} \geq 2N_X + 1. \quad (24)$$

We, too, shall take into account that maximum number of hidden neurons should not exceed the number of the training examples:

$$N_1^{\max} \leq P. \quad (25)$$

Equating (23) with (24) and (23) with (25) we derive the following:

$$\begin{cases} k \times \frac{\sqrt{P \times (N_x + N_y)}}{(N_x + N_y)} \geq 2 \times N_X + 1, \\ k \times \frac{\sqrt{P \times (N_x + N_y)}}{(N_x + N_y)} \leq P. \end{cases} \quad (26)$$

As the consequence, we are going to take into consideration specifics of the actual problems related to the evaluation of the parameters of the technical systems for more exact determination of the factor. As a rule, in these problems:

1. The number of the output parameters does not exceed the number of the input ones;
2. The output neuron corresponds to each recognition class;
3. The number of the training examples for each recognition class should exceed the number of the input parameters at least 10 times;
4. Total amount of all the training examples should exceed the number of the recognized classes at least 10 times as well.

Therefore, we can assume the following without infringing inequality (26):

$$P = 10 \times N_y = 10 \times 10 \times N_x = 10^2 \times N_x. \quad (27)$$

We are going to derive the following upon substitution of (27) in (26):

$$\left\{ \begin{array}{l} k \geq \frac{0,1 \times (2 \times N_x + 1) \times (N_x + N_y)}{\sqrt{N_x \times (N_x + N_y)}} \\ k \leq 10 \times \sqrt{N_x \times (N_x + N_y)}. \end{array} \right. \quad (28)$$

We can assume the following in case of rather great number of input parameters:

$$2 \times N_x + 1 = 2 \times N_x. \quad (29)$$

This allows reduction of (28) as follows:

$$\left\{ \begin{array}{l} k \geq 0,2 \times \sqrt{N_x \times (N_x + N_y)}, \\ k \leq 10 \times \sqrt{N_x \times (N_x + N_y)}. \end{array} \right. \quad (30)$$

Substituting (30) in (23) we can rewrite expressions for the evaluation of the range of optimum number of the hidden neurons of the two-layered perceptron in the following form:

$$N_1^{opt}(\min) = 0,2 \times \sqrt{N_x \times (N_x + N_y)} \times \frac{\sqrt{P \times (N_x + N_y)}}{(N_x + N_y)} \quad (31)$$

$$N_1^{opt}(\max) = 10 \times \sqrt{N_x \times (N_x + N_y)} \times \frac{\sqrt{P \times (N_x + N_y)}}{(N_x + N_y)}, \quad (32)$$

where $N_1^{opt}(\max)$, $N_1^{opt}(\min)$ are maximum and minimum boundaries of the range of optimum number of

hidden neurons in the hidden perceptron.

Upon appropriate reductions and taking into account that the number of the hidden neurons should be integer number we can finally derive the following equation for the two-layered perceptron:

$$N_1^{opt}(\min) = Round(0,2 \times \sqrt{N_x \times P}). \quad (33)$$

$$N_1^{opt}(\max) = Round(10 \times \sqrt{N_x \times P}). \quad (34)$$

Use of expressions (11, 20, 21, 33, 34) allows rewriting equations for the determination of optimum MLP structure in the following form:

$$(H^{\min} - 1) \times J_w^{\min^2} + (N_x + N_y) \times J_w^{\min} = \quad (35)$$

$$= (N_x + N_y) \times Round(0,2 \times \sqrt{N_x \times P})$$

$$(H^{\max} - 1) \times J_w^{\max^2} + (N_x + N_y) \times J_w^{\max} = \quad (36)$$

$$= (N_x + N_y) \times Round(20 \times \sqrt{N_x \times P})$$

where H^{\min} , H^{\max} are minimum and maximum number of the hidden layers of neurons within MLP;

J_w^{\min} , J_w^{\max} are minimum and maximum number of neurons within single hidden layer.

Further refinement of the number of hidden neurons within the limit from H^{\min} to H^{\max} is recommended to do by conducting of numerical experiments related to minimization of the approximation error. We should note that the determination of optimum ratio of the number of hidden layers and the number of neurons within the single hidden layer requires further studies.

Let us consider change of the search range of optimum number of hidden neurons of the two-layered perceptron when using the model developed in order to estimate information value of the results obtained. We shall evaluate change of the range with the help of the following expression:

$$\delta = \frac{N_1^{opt}(\max) - N_1^{opt}(\min)}{N_1^{\max} - N_1^{\min}}. \quad (37)$$

Let us consider the problem that is related to the evaluation of the parameters of computer systems for the purpose of revelation of cyber-attacks. We are going to take into account that the number of the input parameters is ca. 100 ($N_x \approx 100$) for the majority of similar problems, but the number of the training examples required for the effective training of MLP equals to 20,000 ($P \approx 20,000$). We assume that we apply a two-layered perceptron with a single output ($N_y = 1$). If we use the mentioned parameters in expressions (33, 34) we

derive the following: $N_1^{\min} = 282$, $N_1^{\max} = 14140$. Using these results in (37) we derive that $\delta \approx 0.7$. Therefore, range of search of optimum number of hidden neurons narrowed nearly 1.5 times.

Perspectives of the further studies are related to the improvement of the architectural parameters of MLP. First of all, it is reasonable to the development the detailed method for the adaptation of the number of hidden layers to the applied problem specifications set.

IV. CONCLUSIONS

It was shown that increasing efficiency of the multilayer Perceptron application for the estimation of the parameters of the technical systems is possible at the expense of the adaptation of its structural parameters to the problem specifications set. It was proposed to conduct adaptation from the positions of minimization of the relative error of the neuro-network model. Mathematical apparatus was developed which allows calculation of the most acceptable range of the number of neurons within the hidden layers.

REFERENCES

- [1] Ezhov A. A. *Neurokompyuting i ego primeneniya v ekonomike i biznese* / A. A. Ezhov, S. A. Shumskiy. – M.: MIFI, 1998. – 224 s.
- [2] Korchenko A. *Neyrosetevyie modeli, metody i sredstva otsenki parametrov bezopasnosti Internet-orientirovannykh informatsionnykh sistem: monografiya* / A. Korchenko, I. Tereikovskiy, N. Karpinskiy, S. Tyinyimbaev. – K. : TOV «Nash Format» – 2016. – 275 s.
- [3] Tarasenko V. P. *Metod zastosuvannya produktsiinykh pravyl dlia podannia ekspertnykh znan v neiomerezhevykh zasobakh rozpoznavannya merezhevykh atak na kompiuterni systemy* / V. P. Tarasenko, O. H. Korchenko, I. A. Tereikovskiy // *Bezpeka informatsii*. – 2013. – T. 19, № 3. – S. 168-174.
- [4] Rudenko O.H. *Shtuchni neironni merezhi. Navch. posib.* / O.H. Rudenko, Ye.V. Bodianskiy. – Kharkiv: TOV "Kompaniia SMIT", 2006. – 404 s.
- [5] Tereikovskaya L. *Prospects of neural networks in business models* [Text] / L. Tereikovskaya, O. Petrov, M. Aleksander // *TransComp* 2015. 30 November – 3 December, 2015, Zakopanem, Poland. – P. 1539–1545.
- [6] Tereikovskiy I. *Neironni merezhi v zasobakh zakhystu kompiuternoi informatsii* / I. Tereikovskiy. – K. : PolihrafKonsaltnh. – 2007. – 209 s.
- [7] Tereikovskiy I. A. *Optymizatsiia struktury dvokhsharovoho perseptronu, pryznachenoho dlia rozpoznavannya anomalnykh velychyn ekspluatatsiinykh parametrov kompiuternoi merezhi* / I. A. Tereikovskiy // *Naukovo-tekhnichnyi zbirnyk «Upravlinnia rozvytkom skladnykh system» Kyiv. nats. un-tu budivnytstva i arkhitektury*. – 2011. – Vyp. 5. – S. 128–131.
- [8] Zhengbing H. *The Analysis and Investigation of Multiplicative Inverse Searching Methods in the Ring of Integers Modulo M* / H. Zhengbing, I. A. Dychka, M. Onai, A. Bartkoviak // *Intelligent Systems and Applications*, 2016, 11, P. 9-18.

Authors' Profiles



Zhengbing Hu: Ph.D., Associate Professor of School of Educational Information Technology, Central China Normal University, M.Sc. (2002), Ph.D. (2006) from the National Technical University of Ukraine —Kiev Polytechnic Institute. Postdoc (2008), Huazhong University of Science and Technology, China. Honorary Associate Researcher (2012), Hong Kong University, Hong Kong. Major interests: Computer Science and Technology Applications, Artificial Intelligence, Network Security, Communications, Data Processing, Cloud Computing, Education Technology.



Igor A. Tereikovskiy graduated from National Aviation University, Kiev, Ukraine. Currently, he is a Doctor of Science, professor at Faculty of Applied Mathematics, National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Ukraine. He has currently published more than 100 publications.

His research interests are information security, neural network systems for cyber-attacks recognition, recognition of voice signals.



Lyudmila O. Tereikovska graduated from State Academy of Light Industry of Ukraine, Kiev. Currently, she is a PhD, associate professor at Kyiv National University of Construction and Architecture, Ukraine. She has currently published more than 20 publications.

Her research interests are data mining, development of neural network systems, recognition of voice signals, the construction of distance education systems, recognition of cyber-attacks.



Volodymyr V. Pogorelov graduated from Faculty of Applied Mathematics, National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute" in 2014. He is currently a PhD student in Computer Science at National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute". He has currently published 9 publications.

His research interests are cyber-attack recognition, neural networks, artificial intelligence, recognition of voice signals, functional programming, theory of formal languages and compilation.

How to cite this paper: Zhengbing Hu, Igor A. Tereikovskiy, Lyudmila O. Tereikovska, Volodymyr V. Pogorelov, "Determination of Structural Parameters of Multilayer Perceptron Designed to Estimate Parameters of Technical Systems", *International Journal of Intelligent Systems and Applications(IJISA)*, Vol.9, No.10, pp.57-62, 2017. DOI: 10.5815/ijisa.2017.10.07