UDK 681.51:622.7

A. I. KUPIN, Y. O. KUMCHENKO

State institution of higher education «Kryvyi Rih National University», Kryvyi Rih, Ukraine

USAGE OF TRAINING METHODS TO PARAMETERIZATION OF MULTILAYER NEURAL COMPUTING STRUCTURES FOR TECHNOLOGICAL PROCESSES

The analysis of the existing training methods of multilayer neural network computing structures is carried out. By the use of computer simulation the most effective training methods are investigated. Recommendations of selected methods usage by examples of multilayer approximation tasks for technology of benefication are given. As software environments three independent application program packages (neuroemulators) of type were applied to computer simulation: Neuro Solution, Statistica Neural Networks and MATLAB Neural Networks Tools (NNT). Based on the results received in the course of research the comparative analysis was carried out them.

Key words: *Multilayer neural network, computer simulation, training, approximation, identification, classification, technological processes, Levenberg–Marquardt, Gauss-Newton, Conjugate gradient, back propagation.*

Introduction

Nowadays more often to solve applied problems of information and automation under conditions of complex production different intelligent control technologies are used [1]. Thus one of the basic approaches for mathematical models making in the process of approximation, identification, classification is the use of multilayer neural networks (NN) with different architectures.

At present, there are no clear answers to specific questions of unique choice of architecture and the most effective training method (parameterization) in the theory of artificial neural networks. Therefore, most researchers act empirically, choosing from the set of potential alternatives the best variant by certain criteria and under specific technology condition.

1. Analysis of recent research, publications, and presentation of task

For training (parameterization) multilayer neural network structures intended for further identification and control of complex technological processes (TP) in real time, it is necessary to apply methods that meet certain requirements. According to [2] these requirements include: rate of convergence, computational robustness, demands to the computer main memory (RAM) and so on. At present, among the existing methods the so-called methods of the 2nd order meet the requirements the best. They are

- Levenberg–Marquardt;
- Gauss-Newton;
- Conjugate gradient.

Therefore, further analysis, research and selection of potentially effective methods of training neural network structures of technological purposes proposed in [1], will be limited to the set of these methods. From the point of view of automation of further calculations and modelling it is very important that these methods are implemented in the most powerful software packages of emulating neural network structures (MATLAB Neural Tools, Neuro Solutions, Statistical Neural Network, etc.) [5, 6].

2. Material description and results

All these methods are based on functional expansion into the Taylor series up to the 2-nd order. This expansion near the Θ^* point (theoretical parameters optimum of NN) will be as follows [4]:

$$\begin{aligned} V_M\{\Theta,S,\Xi\} &= V_M\left\{\Theta^*,S,\Xi\right\} + \\ &+(\Theta-\Theta^*)^T V'_M\left\{\Theta^*,S,\Xi\right\} + \\ &+\frac{1}{2}(\Theta-\Theta^*)^T V''_M\left\{\Theta^*,S,\Xi\right\}(\Theta-\Theta^*) = \\ &= V_M\left\{\Theta^*,S,\Xi\right\} + (\Theta-\Theta^*)^T G(\Theta^*) + \\ &+\frac{1}{2}(\Theta-\Theta^*)^T H(\Theta^*)(\Theta-\Theta^*), \end{aligned} \quad (1)$$

where $V_M\{\cdot\}$ is the objective function criteria; Θ^* - vector of parameters which are subject to adjustment (NN architecture, weighting factor, regression depth); S - types of regression models which are used; Ξ - statistical data access for training; $G(\Theta^*)$, $H(\Theta^*)$ are the gradient and hessian at the optimum point.

The gradient is defined as

$$G(\Theta^*) = V'_M\left\{\Theta^*, S, \Xi\right\} = \left.\frac{dV_M\left\{\Theta^*, S, \Xi\right\}}{d\Theta}\right|_{\Theta=\Theta^*} \quad (2)$$

and the matrix of other derivatives - hessian or Hessian matrix

$$H(\Theta^*) = V''_M\left\{\Theta^*, S, \Xi\right\} = \left.\frac{d^2V_M\left\{\Theta^*, S, \Xi\right\}}{d\Theta^2}\right|_{\Theta=\Theta^*} \quad (3)$$

Sufficient conditions of minimum of function are zero gradient and positive hessian definition. They are

$$\begin{cases} G(\Theta^*) = 0, \\ H(\Theta^*) > 0. \end{cases} \quad (4)$$

In most cases, finding the minimum may be reduced to the iterative procedure like:

$$\Theta^{(i+1)} = \Theta^{(i)} + \mu^{(i)} f^{(i)}, \quad (5)$$

where Θ^* is the current iteration parameters (i); f(i) - search direction; $\mu^{(i)}$ - step of the current iteration algorithm.

At the same time linear approximation of a prediction error according to the output signal at the output of neural network $d\hat{y}(t\,|\,\Theta)$ is applied as follows:

$$\begin{aligned} \tilde{\varepsilon}(t,\Theta) &= \varepsilon(t,\Theta^{(i)}) + (\varepsilon'(t,\Theta^{(i)}))^T(\Theta - \Theta^{(i)}) = \\ &= \varepsilon(t,\Theta^{(i)}) - (\psi(t,\Theta^{(i)}))^T(\Theta - \Theta^{(i)})^T, \end{aligned} \quad (6)$$

where $\psi(t,\Theta) = \dfrac{d\hat{y}(t\,|\,\Theta)}{d\Theta}$, t is discrete time.

Modified criterion (1) for the i iteration is:

$$V_M\left\{\Theta, S, \Xi\right\} \approx L^{(i)}(\Theta) = \frac{1}{2M}\sum_{i=1}^{M}[\tilde{\varepsilon}(t,\Theta)]^2, \quad (7)$$

where $L^{(i)}(\Theta)$ is the approximate value of the modified criterion, M – the number of training sample templates.

The search direction in the Newton-Gauss method is based on criterion approximation definition $L^{(i)}(\Theta)$ near the current iteration [2-5]. In turn, the conjugate gradient method is based on the search direction change (RESTART) to the gradient direction (antigradient) in a sharp slowdown of convergence. Thus there are different approaches and algorithms of implementation of these procedures for both methods (many versions [7]).

However, no algorithm takes into account that the global minimum $L^{(i)}(\Theta)$ can be located outside the current iteration, as a result the search will be incorrect. Therefore, it will be more rational to assess the reasonability of minimum search $L^{(i)}(\Theta)$ in the area of current iteration. For that according to the algorithm method of Levenberg–Marquardt (known in literature as Levenberg–Marquardt methods, Levenberg scheme, and the method of Levenberg–Marquardt) radius sphere $\delta^{(i)}$ is chosen. Then the optimization problem can be formulated as following system

$$\begin{cases} \hat{\Theta} = \arg\min L^{(i)}, \\ \left|\Theta - \Theta^{(i)}\right| \le \delta^{(i)}. \end{cases} \quad (8)$$

An interactive minimum search procedure in the presence of limitation in the system includes the following stages

$$\begin{cases} \Theta^{(i+1)} = \Theta^{(i)} + f^{(i)}, \\ \left[R(\Theta^{(i)}) + \lambda^{(i)} I\right] f^{(i)} = -G(\Theta^{(i)}), \end{cases} \quad (9)$$

where $\lambda^{(i)}$ is a parameter that defines the area $\delta^{(i)}$.

Hypersphere of radius $\delta^{(i)}$ is defined as an area within which $L^{(i)}(\Theta)$ can be considered as an adequate criterion approximation $V_M\left\{\Theta, S, \Xi\right\}$.

The feature of the method is the procedure of determining of the interaction between $\delta^{(i)}$ and $\lambda^{(i)}$ parameter. As there is no unique dependence between them in practice several heuristic procedures are used [2]. For example, the gradual increase of $\lambda^{(i)}$ until the criterion $L^{(i)}(\Theta)$ will reduce, and then iteration is completed. Values of $\lambda^{(i+1)}$ parameter for the next operation are reduced.

Also an alternative approach, based on a comparison of actual reduction criterion and predicted reduction based on approximation $L^{(i)}(\Theta)$ is used. As a measure of the approximation accuracy the factor r(i) is considered

$$r^{(i)} = \frac{V_M\left\{\Theta^{(i)}, S, \Xi\right\} - V_M\left\{\Theta^{(i)} + f^{(i)}, S, \Xi\right\}}{V_M\left\{\Theta^{(i)}, S, \Xi\right\} - L^{(i)}(\Theta^{(i)} + f^{(i)})}. \quad (10)$$

In the case of approaching factor r(i) to 1, $L^{(i)}(\Theta)$ is an adequate approximation of $V_M\left\{\Theta, S, \Xi\right\}$ and the value of λ decreases, that corresponds to the increase of $\delta^{(i)}$. On the other hand, small or negative factor leads to the need of increasing λ. Based on this the general scheme of the algorithm is as follows:

1. Choose the initial parameter vector value that must be adjustment Θ (0), and the factor of λ (0).
2. Determine the search direction from the equations set (5).
3. If $r^{(i)} > 0,75 \;\Rightarrow\; \lambda^{(i)} = \lambda^{(i)} / 2$.
4. If $r^{(i)} < 0,25 \;\Rightarrow\; \lambda^{(i)} = 2\lambda^{(i)}$.

5. If $V_M\left\{\Theta^{(i)}+f^{(i)},Z^P,\Xi\right\}<V_M\left\{\Theta^{(i)},Z^P,\Xi\right\}$ take as a new iteration $\Theta^{(i+1)}=\Theta^{(i)}+f^{(i)}$ and define $\lambda^{(i+1)}=\lambda^{(i)}$.

6. If the stopping criterion is not achieved, go to the step 2.

The criterion value that minimizes can be presented in the following form

$$L^{(i)}(\Theta^{(i)}+f)=V_M\left\{\Theta^{(i)},S,\Xi\right\}+ +f^T G(\Theta^{(i)})+\frac{1}{2}f^T R(\Theta^{(i)})f. \quad (11)$$

Substituting to (2) the expression for determining the search direction, which was obtained from the ratio

$$R(\Theta^{(i)})f^{(i)}=-G(\Theta^{(i)})-\lambda f^{(i)}, \quad (12)$$

get

$$V_M\left\{\Theta^{(i)},S,\Xi\right\}-L^{(i)}(\Theta^{(i)}+f^{(i)})= =\frac{1}{2}\left(-(f^{(i)})^T G(\Theta^{(i)})+\lambda^{(i)}\mid f^{(i)}\mid^2\right). \quad (13)$$

Ratio (8) allows at the algorithm stages 3 and 4 to determine the factor $r^{(i)}$ using the expression (10).

Based on the general technique of intellectual neural multidimensional identification [8] using the methods of computer simulation the investigation of model structures based on neural network autoregressive predictors in terms of TP magnetite quartzite concentration was conducted. The investigation included the following steps:

– choice of teaching methods, evaluation of the model regression depth (number of delayed signals at the input and output);

– application of teaching methods (the rate of convergence, accuracy);

– direct and inverse prediction;

– testing of derived systems at nonlinearity.

Analysis and choice of the base set of teaching methods for identification models were carried out based on the methodology described in [2]. The main stages of the investigation are:

1. For the simulation experiments the simplest model type NNARX (Neural Network based AutoRegressive eXogenous signal) was chosen. In order to simplify the analysis the same regression depth ($l_1=l_2=2$) was adopted on the basis of previous results [1, 8].

2. Templates of NN of modelling structures in bases of NN of direct distribution (НПР), radial-basic functions (RBF) that full the coherent (FCNN, recurrent) are prepared. For all models the NN with one latent layer by the formula: 16-8-8 (corresponding quantity neurons on a structure input, in the latent layer and on an output) was applied.

3. Tenfold training and testing of all specified NNS of structures with application of four methods of training has been carried out: back propagation (back propagation or BP – a method, as the actual standard of NN training [2-6]), Gauss-Newton (GN - method), Levenberg–Marquardt.

4. LM) and Conjugate gradient. (CG). Statistical sample of indicators has been applied to training Northern Mining Complex ("SevGOK", Kryviy Rih, Ukraine) by the formula: 350-280-70 (total of templates, quantity of templates for training, quantity of templates for verification). Base indicators of first and last stage TP were thus analyzed.

5. Average indicators of convergence (the quantity of epoch or iterations for training), robust (a root-mean-square error – MSE, the generalised root-mean-square error -- NMSE [6]) and the applied computing resources (main memory) has been brought to tabl. 1.

6. On the basis of the results received in the course of research there was comparative analysis carried out.

As program environments for computer modelling there were applied three independent packages of applied programs (neural simulator) type: Neuro Solution, Statistica Neural Networks and MATLAB Neural Networks Tools (NNT). Corresponding results of modelling in these different packages approximately coincide. Also all received results coincide well enough with resulted in [1, 2].

In the course of computer modelling a system hardware-software platform has been applied:

– personal computer with working parameters CPU Pentium IV 2.66 Hz/RAM 2 Gb;

– operating system Windows 7.

On fig. 1 curves which show change of criterion of root-mean-square error MSE in the course of training of model of type NNARX for different bases of neural network structures are resulted. Similar results have been received by the author for others extended autoregressive predictors models NNARXMAX (NNARX + Moving Average, exogenous signal), NNOE (Neural Network Output Error).

The analysis of results of computer modelling allows making certain generalisations in the form of the following conclusions.

Results of training intellectual neural models of type NNARX qualitatively almost identical if they are accordingly grouped (calusterized) by identical methods of training (GN, CG, LM).

From the point of view of speed of convergence and robust the most perspective the method of Levenberg–Marquardt. (LM), but its resources consumption is the greatest. The standard method of training of the NN, based on back propagation (BP), has

Table 1

Comparative estimation of accuracy, resources consumption and speed of convergence of potential algorithms of investigated neural structures training

Algorithm of training	Convergence, Epoch (itera-tions)	MSE	NMSE	COM-PUTER resources, Mb
1. Basis NN (multilayered perceptron)				
1.1. BP	568	1,198596	1,76165223	30
1.2. GN	303	1,161828	1,96306745	24
1.3. LM	177	0,778172	1,45139743	35
1.4. CG	425	0,888760	1,45448391	21
2. Basis RBF (radial-basic functions)				
2.1. BP	196	1,85732511	2,111487478	30
2.2. GN	65	1,19651332	2,131730124	25
2.3. LM	31	0,79076953	1,906790835	35
2.4. CG	87	0,89815021	1,912728683	21
3. Basis FCNN (full coherent neural networks)				
3.1. BP	837	1,0915434	1,60226771	33
3.2. GN	451	1,0807423	1,77265223	27
3.3. LM	265	0,7223413	1,21234453	37
3.4. CG	637	0,8684867	1,26644234	22

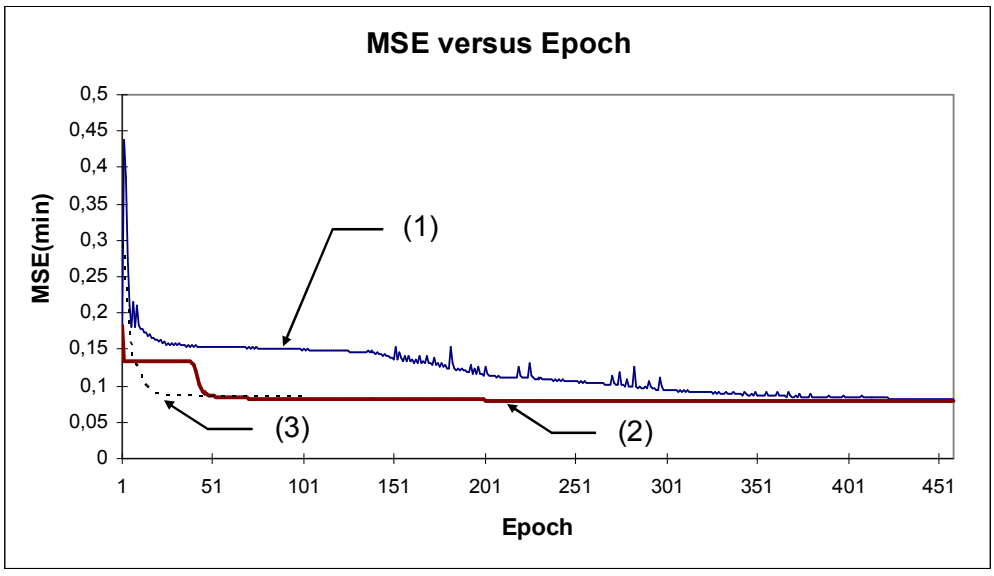

Fig. 1. Change of criterion MSE from quantity of iterations (epoch) at training neural identification model NNARX: 1 – two-layer perceptron which was trained for CG - method; 2 – a network of radial-basic functions (RBF) for GN - method; 3 – full coherent and partially recurrent a network for LM - method

shown good enough robust, but its speed of coincidence slow enough, and requirements concerning resources are too big. Approximately identical and balanced enough results methods of Gauss-Newton (GN) and Conjugate gradient (CG) have shown.

In view of the above-stated tests it is possible to recommend to apply for approximation complex TP and using recurrent dynamic neural structure under condition of possibility of their hardware realisation (for example, neuro-graphic processors) or application of the parallel and distributed computing [9]. The latest is immediate prospects for continuation of the further researches in this direction.

References

1. Kupin, A. Neural identification of technological process of iron ore beneficiation [Text] / A. Kupin // Proceedings of 4th IEEE Workshop on Intelligent Data Acquisition and Advanced Computing Systems Technology and Applications (IDAACS'2007). – Dortmund, Germany, 2007. – P. 225–227.

2. Omatu, S. Neuro-Control and its Applications. Springer-verlag [Text] / S. Omatu, M. Khalid, R. Yusof. – London, 1996. – 272 p.

3. Dorf, R. Modern control systems [Text]/ R. Dorf, R. Bishop. – Prentice Hall, 2001. – 832 p.

4. He, X. A new method for identifying orders in input-output models for nonlinear dynamical systems [Text] / X. He, H. Asada // Proceedings of the American Control Conference. – San Francisco, California, 1993. – P. 2520-2523.

5. Billings, S. A. Nonlinear System Identification: NARMAX Methods in the Time, Frequency, and Spatio-Temporal Domains [Text] / S.A. Billings. – London : Wiley, 2013. – 400 p.

6. Schwenker, F. Three learning phases for radial-basis-function networks [Text] / F. Schwenker, H. Kestler, G. Palm // Neural Networks. – 2001. – № 14. – P. 439–458.

7. Kilian, C. Modern Control Technology [Text] / C. Kilian // Thompson Delmar Learning. – London, 2005. – 608 p.

8. Kupin, A. Identification of technological process of iron ore concentrating with using neural nets [Text] / A. Kupin // Proceedings Of the 3rd International Conference ACSN-2007.–Ukraine. – Lviv : Publishing House of Polytechnic National University, 2007. – P. 83–84.

9. Sundararajan, N. Parallel Architectures for Artificial Neural Networks [Text] / N. Sundararajan, P. Sundararajan // Computer Society Press. – London, 1998. – 412 p.

Поступила в редакцию 20.03.2014, рассмотрена на редколлегии 24.03.2014

Рецензент: д-р техн. наук, проф. Ю. П. Кодратенко, Национальный университет кораблестроения им. адм. Макарова, Николаев, Украина.

ИСПОЛЬЗОВАНИЕ МЕТОДОВ ОБУЧЕНИЯ ДЛЯ ПАРАМЕТРИЗАЦИИ МНОГОМЕРНЫХ НЕЙРОСЕТЕВЫХ СТРУКТУР ТЕХНОЛОГИЧЕСКОГО НАЗНАЧЕНИЯ

А. И. Купин, Ю. А. Кумченко

Сделан анализ существующих методов обучения многомерных нейросетевых структур. Путем компьютерного моделирования исследованы наиболее эффективные методы обучения. Даны рекомендации применения выбранных методов на примере задач многомерной аппроксимации для обогатительной технологии. В качестве программных сред для компьютерного моделирования были применены три независимых пакеты прикладных программ (нейроэмуляторов) типа: Neuro Solution, Statistica Neural Networks и MATLAB Neural Networks Tools (NNT). На основании полученных в процессе исследования результатов был проведен их сравнительный анализ.

Ключевые слова: Многоуровневая нейронная сеть, компьютерное моделирование, обучение, аппроксимация, идентификация, классификация, технологические процессы, Левенберга-Марквардта, Гаусса-Ньютона, Сопряженного градиента, обратное распространение.

ВИКОРИСТАННЯ МЕТОДІВ НАВЧАННЯ ДЛЯ ПАРАМЕТРИЗАЦІЇ БАГАТОВИМІРНИХ НЕЙРОМЕРЕЖЕВИХ СТРУКТУР ТЕХНОЛОГІЧНОГО ПРИЗНАЧЕННЯ

А. І. Купін, Ю. О. Кумченко

Зроблений аналіз існуючих методів навчання багатовимірних нейромережевих структур. Шляхом комп'ютерного моделювання досліджено найбільш ефективні методи навчання. Надані рекомендації застосування обраних методів на прикладі завдань багатовимірної апроксимації для збагачувальної технології. В якості програмних середовищ для комп'ютерного моделювання були застосовані три незалежних пакети прикладних програм (нейроемуляторів) типу: Neuro Solution, Statistica Neural Networks та MATLAB Neural Networks Tools (NNT). На підставі отриманих у процесі дослідження результатів був проведений їх порівняльний аналіз.

Ключові слова: Багаторівнева нейронна мережа, комп'ютерне моделювання, навчання, апроксимація, ідентифікація, класифікація, технологічні процеси, Левенберга-Марквардта, Гауса-Ньютона, сполучених градієнтів, зворотне поширення.

Купін Андрій Іванович – д-р техн. наук, професор, завідувач кафедри комп'ютерних систем та мереж ДВНЗ «Криворізький національний університет», м. Кривий Ріг, Україна, e-mail: kupin@mail.ru.

Кумченко Юрій Олександрович – аспірант, ДВНЗ «Криворізький національний університет», м. Кривий Ріг, Україна, e-mail: ksm06@ukr.net.