UDC 681.327.12(045)
DOI:10.18372/1990-5548.58.13508

[1]M. Z. Zgurovsky,
[2]O. I. Chumachenko,
[3]V. S. Gorbatiuk

STRUCTURAL-PARAMETRIC SYNTHESIS OF THE FEEDFORWARD NEURAL NETWORKS WITH SIGMOID PIECEWISE-TYPE NEURONS

[1,2,3]National Technical University of Ukraine "Ihor Sikorsky Kyiv Polytechnic Institute", Kyiv, Ukraine
E-mails: [1] mzz@kpi.ua, [2]chumachenko@tk.kpi.ua, [3]vladislav.horbatiuk@gmail.com

Abstract—*The method of structural and parametric synthesis of feedforward neural networks is considered, which includes Sigmoid Piecewise neurons, used in the process of a predictive model construction. The article describes the principles of a new developed Sigmoid Piecewise neuron constructing. It is proposed the method of structural and parametric synthesis with a single layer of Sigmoid Piecewise neurons.The training algorithm of proposed neural networks is developed. The results of the effectiveness research of Sigmoid Piecewise neurons on real samples are presented.*

Index Terms—Feedforward neural networks; time series; sigmoid piecewise; training algorithm.

I. INTRODUCTION

The use of artificial intelligence technologies eliminates some of the classical approaches assumptions and drawbacks. At the moment, artificial neural networks are the most promising direction of the intellectual technologies – due to their versatility and impressive results obtained in various areas of their use, such as image and video analysis, text analysis, language analysis, and others.

Artificial neural networks (ANNs) represent a system of interconnected and interacting simple processors – artificial neurons.

Neural networks are not being programmed in the usual sense of the word, they are being trained. From a mathematical point of view, neural networks training is a multi-parameter problem of nonlinear optimization. Training ability is one of the main advantages of neural networks in comparison with traditional algorithms. Technically, training is the process the final goal of which is to find the relationship coefficients the between neurons. During the training process, the neural network is able to detect complex interdependencies between input and output data, as well as to perform generalization.

II. PROBLEM STATEMENT

The mathematical problem of approximating an unknown function having a set of noisy observations occurs when solving a large number of practical problems – time series forecasting, image classification, speech recognition – all these problems get reduced to the approximation problem of some form most of the times. At the moment, the best results in this area are obtained by using neural networks-based approaches [1]. Moreover, a large number of practical and theoretical studies indicate that the mathematical model of neurons that are used in the corresponding networks has a big influence on the overall quality of a final approximation model – for example, the use of ReLU neurons [2] has made it possible to significantly improve the best quality of network classification in the ImageNet competition. In this paper, we propose a new model of the Sigmoid Piecewise (SP) neuron together with a greedy algorithm for pretraining a single-layer feedforward network [3], which uses these neurons in its hidden layer.

III. PROBLEM SOLUTION

The continuous function of many variables $f(\vec{x}), \vec{x}\in R^n$, which satisfies certain additional properties, can be approximated with any given accuracy [4] by a piecewise linear function that cuts the space R^n in convex sets $S_1,...,S_N \subseteq R^n$, which do not intersect, and defines the corresponding linear function $f_i(\vec{x}) = \vec{w}_i^{\mathrm{T}}\vec{x},\ i=1,...,N, \vec{w}_i \in R^n$ on each set.

The issue with using piecewise linear functions as a basic neurons is that they are non-continuous and certain issues arise when trying to use gradient-based optimization methods [5] for their parameters tuning. Thus, a neuron with a following function model is proposed:

$$\text{sigm\_piecewise}(\vec{x};\vec{w}_+,\vec{w}_-,\vec{h}) = \frac{\vec{w}_+^{\mathrm{T}}\vec{x}}{1+e^{-k\vec{h}^{\mathrm{T}}\vec{x}}} + \frac{\vec{w}_-^{\mathrm{T}}\vec{x}}{1+e^{k\vec{h}^{\mathrm{T}}\vec{x}}},\ k>0,$$

which is a continuous approximation of the piecewise-linear model:

$$\text{piecewise\_linear}(\vec{x};\vec{w}_+,\vec{w}_-,\vec{h}) = \vec{w}_+^{\mathrm{T}}\vec{x} \cdot step'(\vec{h}^{\mathrm{T}}\vec{x}) + \vec{w}_-^{\mathrm{T}}\vec{x}\cdot step''(-\vec{h}^{\mathrm{T}}\vec{x}),$$

where

$$step'(x) = \begin{cases} 1, x \ge 0, \\ 0, x < 0, \end{cases} \quad step''(x) = \begin{cases} 1, x > 0, \\ 0, x \le 0. \end{cases}$$

Parameters' vectors $\vec{w}_+, \vec{w}_-, \vec{h}$ have the following meanings:

– vector $\vec{h}$ specifies a hyperplane that divides the space R^n into 2 half-spaces

– vector $\vec{w}_+$ specifies weights of the piecewise linear function in the half-space where $\vec{h}^{\mathrm{T}} x \ge 0$

– vector $\vec{w}_-$ specifies weights of the piecewise linear function in the half-space where $\vec{h}^{\mathrm{T}} x < 0$.

Structurally, a neuron with a sigm_piecewise model can be represented in the following way where AF is the activation function of the neuron sigm_piecewise (Fig. 1).

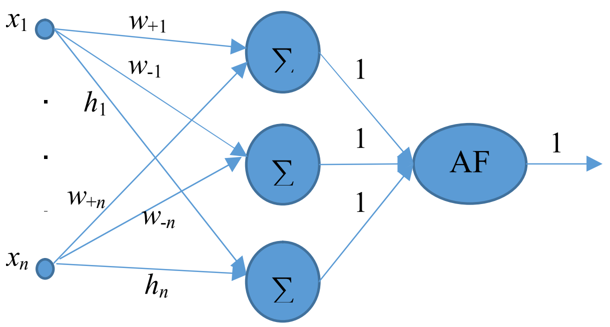

Fig. 1. Structural representation of a neuron with a sigm_piecewise model

The method for pretraining [6] parameters in single-layer networks of the following form is proposed, i.e. networks in the hidden layer of which the neurons with the same model $f(\vec{x};\vec{w})$ are used where $\vec{w}$ is the parameters vector of the neuron model, and the output layer of which consists of one neuron - a weighted adder (Fig. 2).

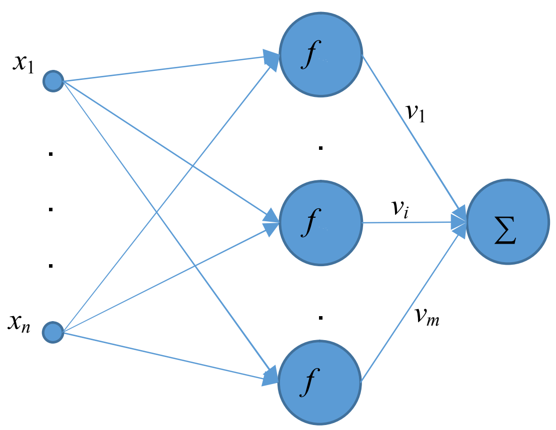

Fig. 2. Single-layer network

The idea of the method is to gradually increase the number of neurons in the hidden layer, and while adding a new neuron only its parameters and the parameter which is responsible for this neuron weight in the neuron-adder of the output layer are being trained- all other parameters are being fixed.

The following condition must be fulfilled in order to use this method – for any training set $< X^{(\mathrm{T})}, y^{(\mathrm{T})} : X^{(\mathrm{T})} \to \Re >$ and for any given error value $\varepsilon > 0$, which is to be achieved, a certain number of neurons in the hidden layer m, and such parameters vectors $\vec{w}_1, ..., \vec{w}_m, \vec{v}; \vec{v} \in R^m$ exist that the network error of the described type with such parameters will be less or equal to ε:

$$\forall < X^{(\mathrm{T})}, y^{(\mathrm{T})} : X^{(\mathrm{T})} \to R >, \varepsilon > 0 : \exists m \in N, \vec{w}_1, ...,$$
$$\vec{w}_m, \vec{v} : E(X^{(\mathrm{T})}, y^{(\mathrm{T})}, \mathrm{Net}(\vec{x}) = \sum_{i=1}^{m} v_i f(\vec{x}; \vec{w}_i)) \le \varepsilon$$

If this condition is not met, then the situation may arise when the addition and training of any new neurons number will not reduce the network error in the training set. For example, this condition is not satisfied for a network with linear neurons in a hidden layer, i.e. when $f(\vec{x};\vec{w}) = \vec{w}^{\mathrm{T}}\vec{x}$ – since the linear combination of any number of such neurons will always remain a linear function of $\vec{x}$, so if the examples in the training set are not described ideally by some linear function, then there is some limit value of the error $\varepsilon_{threshold}$, and the value less that it can not be obtained with any network configuration having linear neurons in the hidden layer.

Having a training and validation set of the form $< X^{(\mathrm{T})}, y^{(\mathrm{T})} : X^{(\mathrm{T})} \to \Re >$ and $< X^{(V)}, y^{(V)} : X^{(V)} \to \Re >$ respectively, the pretraining of single-layer network of the described type according to the proposed method consists of the following steps:

1) On the first iteration, the network of the described type with one neuron in the hidden layer is trained.

2) After i iterations we have a network with i neurons in the hidden layer.

3) On iteration $i+1$ another neuron is added to the hidden network layer, after which the network training is performed one more time, but all the network parameters, except for this neuron parameters and its weight in the neuron-adder of the output layer, are fixed (Fig. 3).

4) At every iteration i the following parameters are calculated:

– the current value of the network error on the validation set – it is defined as $E^{(V)}(i)$;

– the minimum value among all the errors in previous iterations – it is defined as $E\_\min^{(V)}(i) = \min_{j=1,\ldots,i-1} E^{(V)}(j)$.

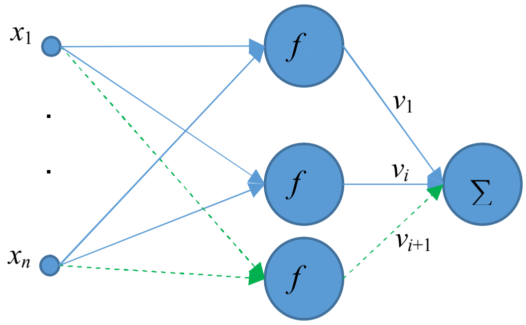

Fig. 3. Another neuron adding to the network a hidden layer

5) New neurons adding is performed while the following condition is fulfilled-during the I last iterations the minimum value of the network error on the validation set was sufficiently improved, that is:

$$E\_\min^{(V)}(i) \le a \cdot E\_\min^{(V)}(i-I), \quad a \in (0,1],\ I \in N,\ \forall i > I,$$

where a is the constant that specifies the desired level of minimum error value improvement for I iterations – usually it is chosen from a set of numbers $\{0.9, 0.99, 0.999, \ldots\}$ – that is with $a = 0.9$ by I iterations, the minimum network error on the validation set should decrease by at least 10%, with $a = 0.99$ – not less than 1%, and so on.

6) After the stop, the network returns to the iteration, where the minimum error in the validation set was reached.

The use of this method for the pretraining of the described networks type is appropriate when in order to achieve a satisfactory value of the network error on the training set it is enough to have a network with a relatively small number of neurons in the hidden layer – from 10 to 10,000 (depending on the computing capacity of the computer being used for training, and the time allocated for training), since on each iteration the parameters of only one neuron are being trained. This situation usually arises when the unknown approximation function is quite simple, or when the model of the neurons used in the hidden layer is a rather complicated function.

IV. Study of the Sigmoid Piecewise-Type Neurons Efficiency on Real Sets

Let us perform a comparative test of the new neuron efficiency in the approximation problem on real set. We will compare it with the ReLU neuron. For a given set $< X,\ y : X \to R >$ the benchmark consists of the following steps:

1) A certain level of error is defined $\text{MSE}_{target} > 0$ which is to reach.

2) At the first iteration of the test, two fully connected feedforward networks [7] with one hidden layer with 1 neuron, and one output linear neuron are trained: the network with ReLU neuron in the hidden layer, and the network with SP neuron in the hidden layer are being trained.

3) If the error is both networks $< \text{MSE}_{target}$ is the test is over.

4) After i iterations we have a network with i neurons in the hidden layer.

5) On iteration $i+1$ one more neuron of the corresponding type is added to the hidden layer of each network, after which the training of both networks is repeated, but all network parameters, except for the parameters of new neurons and their weights in the output linear neuron, are fixed.

6) For each network, the current error MSE_{i+1} is calculated and remembered - that is the error of the corresponding network, achieved while using $i+1$ neurons.

7) If on iteration $i+1$ the error of both networks has become $< \text{MSE}_{target}$ is the test is stopped.

8) After stopping the test, we have data of the form “number of neurons in the hidden layer” → “reached value of the approximation error”.

The comparative test was performed on the time series of daily rates of USD to EUR (open access):

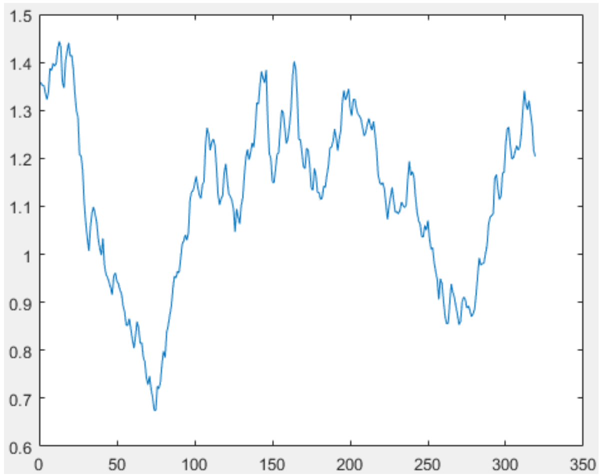

Fig. 4. Time series of daily rates of USD to EUR

The training set $< X_t, y : X_t \to \Re >$ was constructed from the output time series by the method of time series embedding with the embedding dimension $m = 5$ and the forecast horizon $k = 4$ – that is for the value g_{i+4} forecast the values

$g_{i-4}, g_{i-3}, g_{i-2}, g_{i-1}, g_i$ have been used. Mean squared error of a naive model $\hat{g}_{i+4} = g_i$ equals $\mathrm{MSE}_{naive} = 7.16 \cdot 10^{-2}$. Let us set the level of the mean square error that needs to be achieved $\mathrm{MSE}_{target} = 5 \cdot 10^{-2}$, and we will "greedily" train both networks until the specified error level is reached, and then we build a plot of the neurons number vs the mean square error of the network, consisting of this number of greedily trained neurons, for both networks:

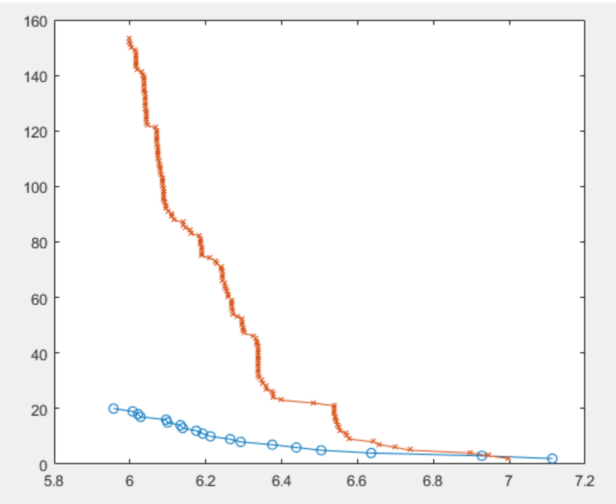

Fig. 5. Plot of the number of neurons vs the mean square error of the network

As we can see, starting approximately from the error value $\mathrm{MSE}_{treshold} \approx 6.5 \cdot 10^{-2}$, the "weighted" ReLU -type neurons were required 3 times more than sigm_piecewise -type neurons to achieve this error value, after which the gap in the required number of neurons only increased - and hence, taking into account that the sigm_piecewise -type neuron has approximately 3 times more parameters – to achieve an error less than when $\mathrm{MSE}_{treshold}$ using – sigm_piecewise type neurons fewer parameters were required than when using "weighted" ReLU-type neurons – which indicates a greater efficiency of the SP-type neuron in terms of the "parameters number necessary to achieve the given error level" criterion if the error level is small enough.

V. CONCLUSION

A new model of artificial Sigmoid Piecewise neuron has been developed, which is based on the use of two types of weight coefficients and a special activation function that allows to adjust simultaneously and independently the parameters of the hyperplane that breaks the input space into 2 half-spaces, and the parameters controlling the initial value of the neuron on different half-spaces. This allows approximating rather complex functions using a small number of neurons – thus reducing the number of parameters that need to be adjusted. In addition, due to the separation of the parameters being responsible for the separation of the hyperplane and the parameters that correspond to the original value of the neuron in different half-spaces, neural network based on the Sigmoid Piecewise neurons can "automatically" clusterize the space of input examples and approximate the given function separately on each cluster – thus improving the accuracy of the final forecasting network while training on heterogeneous samples.

A special greedy algorithm for pretraining of feedforward networks with one hidden layer composed of Sigmoid Piecewise neurons have been developed as well. The application of this algorithm allows to automatically select the optimal (in a certain sense) structure of the neural network (similar to [8]) thus reducing the need for expert intervention in the process of building a prediction model.

REFERENCES

[1] Jürgen Schmidhuber, "Deep learning in neural networks: An overview," *Neural networks*, vol. 61, pp. 85–117, January 2015. doi.org/10.1016/j.neunet.2014.09.003

[2] Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *The Journal of Machine Learning Research,* vol. 15, no. 1, pp. 1929–1958, 2014.

[3] Olivier Delalleau, and Yoshua Bengio, "Shallow vs. deep sum-product networks," *Advances in Neural Information Processing Systems.* 2011, pp. 666–674.

[4] Stephen Boyd, and Lieven Vandenberghe, *Convex optimization*. Cambridge university press. 2004.

[5] Martin Riedmiller, and Heinrich Braun, "A direct adaptive method for faster backpropagation learning: The RPROP algorithm," *IEEE International Conference on Neural Networks*. 1993, pp. 586–591.

[6] Hugo Larochelle, Yoshua Bengio, Jérôme Louradour, and Pascal Lamblin, "Exploring strategies for training deep neural networks," *Journal of machine learning research*, 10, no., pp.1–40, Jan. 2009.

[7] Donald F. Specht, "A general regression neural network," *IEEE transactions on neural networks* 2, no. 6, 1991, pp. 568–576.

[8] Sung-Kwun Oh, and Witold Pedrycz, "The design of self-organizing polynomial neural networks," *Information Sciences* 141, no. 3–4, pp. 237–258, 2002.

Received October 19, 2018

Zgurovsky Michael. Doctor of Engineering Science. Professor. Academician of the National Academy of Sciences of Ukraine.
Rector of National Technical University of Ukraine "Ihor Sikorsky Kyiv Polytechnic Institute," Kyiv, Ukraine.
Education: Kyiv Polytechnic Institute (now National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute"), Kyiv, Ukraine, (1975).
Research area: cybernetics, system analysis, decision-making theory.
Publications: 50 inventions, author and co-author of over than 400 scientific works, 38 monographs and textbooks.
E-mail: mzz@kpi.ua

Chumachenko Olena. Candidate of Science (Engineering). Assosiate Professor.
Technical Cybernetic Department, National Technical University of Ukraine "Ihor Sikorsky Kyiv Polytechnic Institute," Kyiv, Ukraine.
Education: Georgian Politechnic Institute, Tbilisi, Georgia, (1980).
Research area: system analysis, artificial neuron networks.
Publications: more than 80 papers.
E-mail: chumachenko@tk.kpi.ua

Gorbatiuk Vladyslav. Post-graduate student.
Technical Cybernetic Department, National Technical University of Ukraine "Ihor Sikorsky Kyiv Polytechnic Institute," Kyiv, Ukraine.
Education: National Technical University of Ukraine "Ihor Sikorsky Kyiv Polytechnic Institute" (2017).
Research area: deep learning, forecasting methods.
Publications: 10.
E-mail: vladislav.horbatiuk@gmail.com

М. З. Згуровський, О. І. Чумаченко, В. С. Горбатюк. Структурно-параметричний синтез нейронних мереж прямого поширення, з нейронами типу Sigmoid Piecewise

Розглянуто метод структурного та параметричного синтезу нейронних мереж прямого поширення, до складу яких входять нейрони типу Sigmoid Piecewise, котрі використовувалися при побудові прогнозуючої моделі. У статті описано принцип побудови нового розробленого нейрону типу Sigmoid Piecewise. Наведено результати дослідження ефективності нейронів типу Sigmoid Piecewise на реальних вибірках.

Ключові слова: штучні нейронні мережі; часові ряди; сигмовидна кускова функція; пряме поширення.

Згуровський Михайло Захарович. Доктор технічних наук. Професор. Академік НАН України.
Ректор Національного технічного університету України «Київський політехнічний інститут ім. Ігоря Сікорського», Київ, Україна.
Освіта: Київський політехнічний інститут (тепер Національний технічний університет України «Київський політехнічний інститут ім. Ігоря Сікорського»), Київ, Україна, (1975).
Напрям наукової діяльності: кібернетика, системний аналіз, теорія прийняття рішень.
Кількість публікацій: 50 винаходів, автор та співавтор понад 400 наукових праць, 38 монографій та підручників.
E-mail: mzz@kpi.ua

Чумаченко Олена Іллівна. Кандидат технічних наук. Доцент.
Кафедра технічної кібернетики, Національний технічний університет України «Київський політехнічний інститут ім. Ігоря Сікорського», Київ, Україна.
Освіта: Грузинський політехнічний інститут, Тбілісі, Грузія, (1980).
Напрям наукової діяльності: системний аналіз, штучні нейронні мережі.
Кількість публікацій: більше 80 наукових робіт.
E-mail: chumachenko@tk.kpi.ua

Горбатюк Владислав Сергійович. Аспірант.
Кафедра технічної кібернетики, Національний технічний університет України «Київський політехнічний інститут ім. Ігоря Сікорського», Київ, Україна.
Освіта: Національний технічний університет України «Київський політехнічний інститут ім. Ігоря Сікорського», Київ, Україна, (2017).
Напрям наукової діяльності: глибоке навчання, методи прогнозування.
Кількість публікацій: 10.
E-mail: vladislav.horbatiuk@gmail.com

М. З. Згуровский, Е. И. Чумаченко, В. С. Горбатюк. Структурно-параметрический синтез нейронных сетей прямого распространения, с нейронами типа Sigmoid Piecewise
Рассмотрен метод структурного и параметрического синтеза нейронных сетей прямого распространения, в состав которых входят нейроны типа Sigmoid Piecewise, которые использовались при построении прогнозирующей модели. В статье описан принцип построения нового разработанного нейрона типа Sigmoid Piecewise. Приведены результаты исследования эффективности нейронов типа Sigmoid Piecewise на реальных выборках.
Ключевые слова: искусственные нейронные сети; временные ряды; сигмовидная кусочная функция; прямое распространение.

Згуровский Михаил Захарович. Доктор технических наук. Профессор. Академик НАН Украины.
Ректор Национального технического университету Украины «Киевский политехнический институт им. Игоря Сикорского», Киев, Украина.
Образование: Киевский политехнический институт (ныне Национальный технический университет Украины «Киевский политехнический институт им. Игоря Сикорского»), Киев, Украина, (1975).
Направление научной деятельности: кибернетика, системный анализ, теория принятия решений.
Количество публикаций: 50 изобретений, автор и соавтор более 400 научных работ, 38 монографий и учебников.
E-mail: mzz@kpi.ua

Чумаченко Елена Ильинична. Кандидат технических наук. Доцент.
Кафедра технической кибернетики, Национальный технический университет Украины «Киевский политехнический институт им. Игоря Сикорского», Киев, Украина.
Образование: Грузинский политехнический институт, Тбилиси, Грузия, (1980).
Направление научной деятельности: системный анализ, искусственные нейронные сети.
Количество публикаций: более 80 научных работ.
E-mail: chumachenko@tk.kpi.ua

Горбатюк Владислав Сергеевич. Аспирант.
Кафедра технической кибернетики, Национальный технический университет Украины «Киевский политехнический институт им. Игоря Сикорского», Киев, Украина.
Образование: Национальный технический университет Украины «Киевский политехнический институт им. Игоря Сикорского», Киев, Украина, (2017).
Направление научной деятельности: глубокое обучение, методы прогнозирования.
Количество публикаций: 10.
E-mail: vladislav.horbatiuk@gmail.com