Annals of DAAAM for 2011 & Proceedings of the 22nd International DAAAM Symposium, Volume 22, No. 1, ISSN 1726-9679
ISBN 978-3-901509-83-4, Editor B. Katalinic, Published by DAAAM International, Vienna, Austria, EU, 2011

Make Harmony between Technology and Nature, and Your Mind will Fly Free as a Bird

Annals & Proceedings of DAAAM International 2011

ANALYSIS OF ITERATIVE LEARNING ALGORITHMS FOR THE MULTILAYER PERCEPTRON NEURAL NETWORK

BACEK, T[omislav]; MAJETIC, D[ubravko]; BREZAK, D[anko] & KASAC, J[osip]

Abstract: *In this paper, a comparison of different algorithms, used in the training of a multilayer feedforward neural network (MLP), is presented. Tested algorithms, which are of the first or the second order, include both local and global adaptation techniques. Prediction of nonlinear dynamic Glass-Mackey system is used as a benchmark problem. To improve training speed and efficiency, bipolar sigmoidal activation function with adaptive gain parameter is used. Furthermore, modification of random weight initialization is proposed.*

Key words: *static neural network, adaptive activation function, prediction, nonlinear chaotic system*

1. INTRODUCTION

Neural networks (NN) are used in a wide variety of applications due to their capacity to learn and to generalize. They are of especially great interest in areas where problems cannot be solved by conventional methods, such as dynamic system prediction. For this reason, these types of problems can be used as a benchmark tests. Several experiments have so far shown that neural networks can successfully deal with nonlinear systems (Novakovic et al., 1998). They can be thought of as a mapping function, which is basically a solution of prediction problems.

Most widely used algorithm in the training MLP networks is the Error Back Propagation (EBP). In order to enhance the training capability of the basic EBP algorithm, several modifications are included – momentum, activation function (AF) with adaptive parameter and modified weight initialization method. Yet, none of these modifications managed to prevail the main limitation of the EBP method – dependence on the size of the partial derivative. Therefore, in this paper, the EBP method was compared with several major training algorithms.

Beside aforementioned modified EBP algorithm, another three frequently used algorithms, namely Conjugate Gradient (CG), Resilient Backpropagation (RPROP) and Levenberg-Marquardt (LM), are also tested and compared. Since there are several known versions of the CG and the RPROP algorithms, two versions of each are tested in this paper. During all tests, bipolar sigmoidal activation function, 6-13-1 network structure and initial weights have not been changed.

Main goal of our research is to find the best MLP network learning algorithm in regression and classification problems. A part of this comprehensive research is presented in this paper.

2. FEEDFORWARD NEURAL NETWORK

Neural network used in this paper is a three-layered feedforward NN. Input of the *i*-th neuron of the *k*-th layer (with the exception of an input layer) is a sum of weighted outputs of neurons of the (*k-1*)-th layer, (1). Bias is the only neuron that has no input, because its output is always one. It controls shape, orientation and steepness of sigmoidal activation function (AF), and therefore needs to be included. Neural network task in this study was to predict value of only one point ahead by using values from the 4 past and a present point. Therefore, input layer has 5 neurons (plus bias), while output layer has one neuron. Hidden layer can have arbitrary number of neurons, so 12 neurons (plus bias) are chosen in this paper. Sigmoidal AF of hidden layer neurons is given in (2), whereas AF of the output layer neuron is a simple linear function with unit gain.

$$net_i^{(k)} = \sum_{i=1}^{I-1} w_{ij} \cdot y_j^{(k-1)} , \qquad (1)$$

$$y_i^{(k)}(net_i) = \frac{2}{1+e^{-c \cdot net}} - 1 \, , \qquad (2)$$

where y and *net* denote neuron output and input, respectively, and c is AF’s adaptive gain parameter.

In order to improve learning, a modification of random weight initialization is proposed (Nguyen & Widrow, 1990),

$$\mathbf{W} = 0.7\, H^{\frac{1}{L}}(-1 + 2 \cdot rand) \, , \qquad (3)$$

where H and L denote the number of neurons in layers connected with the weight vector $\boldsymbol{W}$, former referring to succeedding and latter to preceding layer.

3. LEARNING ALGORITHMS

As mentioned before, four learning algorithms are tested and compared. The basic EBP algorithm (Novakovic et al., 1998) has slow convergence in case of small learning parameter η, and can lead to oscillations in case of big η. Hence, the basic EBP is modified with both 1st (α) and 2nd (β) order momentum, latter being set to *(α-1)/3*. The RPROP algorithm (Igel & Husken, 2000) has so far been presented in four versions. Since modified versions proved to outperform basic versions, they are also used in this paper. Modified RPROP versions are reffered to as iRPROP+ and iRPROP-. The CG algorithm (Kasac et al., 2009) tested in this paper uses Fletcher-Reeves (FR) or Dai-Yuan (DY) method for finding parameter β, because the FR method is the most widely used, and the DY method is proved to achieve the same level of accuracy as the FR method with the substantial reduction of the computational time (Kasac et al., 2009). In conclusion, the fourth analysed algorithm was the LM algorithm (Hagan & Menhaj, 1994). In order to have more influence on the network behavior, coefficient β is actually given as β_{dec} and β_{inc}. Former is used to multiply parameter μ when error decreases, while latter is used when error increases.

Performance index used in this paper is the sum of squared errors,

$$E = \sum_{i=1}^{N} (d_i - O_i)^T \, (d_i - O_i) \, , \qquad (4)$$

where N is the training set size, while d_i and O_i denote desired and actual network response, respectively. All error measures are reported using non-dimensional Normalized Root Mean Square error index – NRMS (Lapedes & Farber, 1987).

4. NONLINEAR CHAOTIC SYSTEM

Chaos is a common property of all nonlinear dynamic systems, with a wide variety of nonlinear behaviors, which makes it a great benchmark for testing different signal processing techniques. Since its definition is simple, but its behavior hard to predict, Glass-Mackey chaotic system is proposed as a NN benchmark (Lapedes & Farber, 1987). Discrete Glass-Mackey dynamic system is defined as (Novakovic et al., 1998)

$$x(n-1) = \frac{1}{1+b}\left[x(n-1) + \frac{ax(n-\tau)}{1+\, x^{10}(n-\tau)}\right], \qquad (5)$$

where a and b are constants, and τ is time delay. Sampling time is T_0=1s. In this paper, a=0.2, b=0.1 and τ=30.

In order to predict the behavior of nonlinear chaotic system, i.e. signal value in P-th point ahead, a mapping function $f(\bullet)$ needs to be determined from

$$x(n+P) = \, f\big(x(n), x(n-\Delta), \ldots, x(n-m\Delta)\big), \qquad (6)$$

where P denotes number of points ahead, Δ denotes signal delay, and m is an integer constant. In this paper, P=Δ=6, m=4.

The Glass-Mackey discrete-time series benchmark, used in this paper, was generated using Eq. (5) and consisted of 1000 points. First 500 points were used for learning, whereas the remaining 500 points were used for the testing of algorithms.

5. EXPERIMENTAL RESULTS

Every network learning process was carried out using 35000 learning steps. During this process, network was tested after every 1000 steps for there is no guarantee that the test error will have strictly decreasing manner as learning proceeds. If test error decreased compared to a previous one, weights were saved. Otherwise, they were not considered. Table 1 shows learning and test errors for all algorithms, as well as step in which the smallest registered test error encountered. Comparison of NRMS test error curves for the EBP, iRPROP-, CG DY and LM algorithms is presented in Fig. 1.

	$NRMS_{learning}$	$NRMS_{test}$	$NRMS_{test}$ step
EBP	0.0662	0.0936	34000
iRPROP-	0.0644	0.0834	16000
iRPROP+	0.0635	0.0834	19000
CG FL	0.0621	0.0831	12000
CG DY	0.0532	0.0823	12000
LM	0.0379	0.0745	6000

Tab. 1. Experimental results of a feedforward NN six-step-ahead prediction

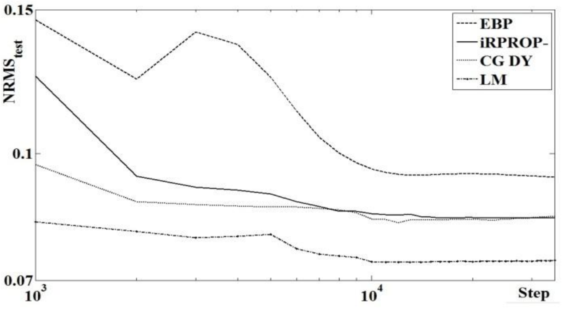

Fig. 1. Comparison of NRMS test error curves

From the presented results it can be seen that the best results were accomplished with the LM algorithm. The problem with LM algorithm is time consumption, rising up from computa-tional requirements of each step. Nevertheless, this drawback is surpassed by the increased efficiency (after only 2000 steps LM algorithm already outperformed the best results of all other tested algorithms).

Fig. 2 depicts the best NN test result. It can be seen than NN learned its prediction task on previously unseen data with high accuracy, which confirms NN generalization capabilities.

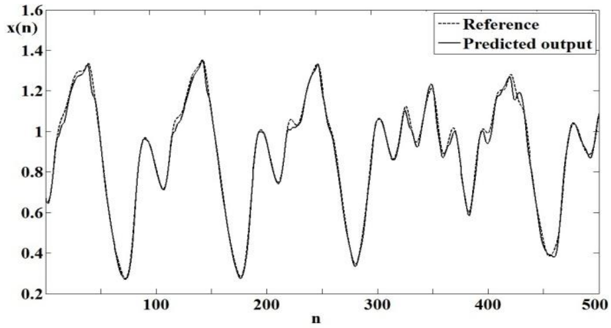

Fig. 2. Prediction of the Glass-Mackey time series using the 6-13-1 feedforward NN trained with LM algorithm

6. CONCLUSION

Comparison of different learning algorithms, used in the training of a static NN, is presented. Modified weight initialization method and an adaptation of AF gain parameter are included to improve learning capabilities. Also, modified versions of simple EBP, RPROP and LM algorithms are tested. For this purpose, prediction of nonlinear chaotic system is used.

Criteria used for the evaluation of learning algorithms which influenced the neural network performance were efficiency and accuracy of the neural network, with the emphasis on the accuracy, due to its direct relation to the generalization capability. In our experiments, Levenberg-Marquardt algorithm proved to be the best algorithm regarding both criteria. Both versions of the RPROP and CG algorithm achieved comparable results, whereas EBP turns out to be the algorithm with the poorest learning and especially generalization capabilities.

Future work will be directed towards analysis of presented algorithms and their modifications on different regression and classification benchmark problems and will be published in our upcoming publications.

7. REFERENCES

Hagan, T.M. & Menhaj, M.B. (1994). Training Feedforward Networks with the Marquardt Algorithm, *IEEE Transactions on Neural Networks*, Vol. 5, No. 6, pp. 989-993, ISSN 1045-9227, November 1994

Igel, C. & Husken, M. (2000). Improving the Rprop Learning Algorithm, *Proceedings of the Second International Symposium on Neural Computation, NC' 2000*, Bothe, H. & Rojas, R.(Ed.), pp. 115-121, ICSC Academic Press, 2000

Kasac, J.; Deur, J.; Novakovic, B. & Kolmanovsky, I.V. (2009). A Conjugate Gradient-based BPTT-like Optimal Contol Algorithm, *3rd IEEE Multi-conference on Systems and Control*, ISSN 1085-1992, Saint Petersburg, 2009

Lapedes, A. & Farber, R. (1987). Nonlinear Signal Processing Using Neural Networks:Prediction and System Modeling, *Techical Report*, Los Alamos National Laboratory, Los Alamos, New Mexico, 1987

Nguyen, D. & Widrow, B. (1990). Improving the Learning Speed of 2-Layer Neural Networks by Choosing Initial Values of the Adaptive Weights, *Proceedings of the International Joint Conference on Neural Networks (IJCNN)*, Vol. 3, pp. 21-26, San Diego, CA, USA, 1990

Novakovic, B.; Majetic, D. & Siroki, M. (1998). *Artificial Neural Networks*, Faculty of Mechanical Engineering and Naval Architecture, ISBN 953-6313-17-0, Zagreb, Croatia