Australian Journal of Basic and Applied Sciences, 5(10): 571-576, 2011
ISSN 1991-8178

Presentation of a Combined-Clustered Network Pattern Based on Neural Networks for Improving the Performance of Pattern Recognition

Fatimah Masoudifar

Abstract: This paper presents a multi-layer combined network for improving the performance of common networks. We have chosen the clustered algorithm from structural architecture based on Radial Base Function (RBF) and applied it in the combined multi-layer network. Discovering the multi-layer architecture introduced here is mainly based on the algorithm of Modified Regression Predicted Error (MRPE). The ability of this architecture is that it can use a set of evaluated data and compare them with other network architectures. The network is evaluated using IRIS data. In addition to evaluating the credibility of network, its stability is also checked based on a structural procedure.

Key words: Neural networks, radial base function, algorithm of modified regression predicted error, combined multi-layer network.

INTRODUCTION

Nowadays in the medical and financial fields, analyzers are usually exposed to classified items based on the measured or historical data. In electric engineering, engineers have decided to categorize the distribution of transformers with sectorial discharge conditions. Geologists also categorize the shape of grains into angular or squared while pathologists divide the results of a cancer test into three groups of normal, semi-serious and serious. The main problem against analysts is how to categorize data with complete complexity and many dependant variables. A great deal of time and cost is being spent in order to evaluate the classified problems accurately (Ozgrid Business Applications).

Neural networks are efficient in solving some complicated problems such as classifying a huge number of data. Their initial applications go back to 19^{th} century when they were used to explain human brain operations as a special work. Today, famous algorithms of learning and architecture are applied based on neural networks in different practical programs. Multi-layer networks and Radial Base Functions (RBF) are among the most used applications.

MLP network is one of the multi-layer networks which can be learnt based on levels of autonomous decisions in the input space. Learning in these networks is usually a very time consuming procedure. RBF network is a special class of neural networks with a hidden simple layer and its linear weighting structure is one of its advantages over MLP networks. Parameters in the units of hidden layer RBF networks can be constant from the beginning, so the RBF network can be learnt without quick and nonlinear optimization. Regarding the high speed of the RBF networks and their other advantages these networks are not used usually; because in recognition of patterns, network accuracy is more important than its speed (Chathura *et al.,* 2008).

Performance of a neural network is dependent on the network structure and its learning algorithm. Thus, direct linear input connections from the input nodes to the output nodes are proposed to improve the performance and extension of it into nonlinear neural networks such as MLP and RBF. So, one RBF network with input linear connections can perform better with more efficiency than a standard RBF. Noteworthy is that these concepts also have some applications in MLP networks but the input linear connections cannot be explained by the regression learning algorithm. Modifications in the network architecture cannot increase performance of the MLP networks. For solving the problem, the MRPE algorithm is used in MLP networks. That's why we may call MLP networks based on MRPE as combined multi-layer networks, while MRPE algorithm and the input linear connections is remarked as a combined pattern in the network which is able to improve MLP networks (Lung, 2007).

This paper continues as following: Section 2 briefly studies the ideas which were the base for establishing combined multi-layer networks. Section 3 deals with explaining the clustered-combined multi-layer networks and testing the performance, it also speaks about the implementation of clustered-combined pattern and the rate of incompatibility. Section 4 consists of a case study. Finally in sections 5 and 6, the paper pays to discussion, conclusion and further studies.

Combined Multi-Layer Networks:

MLP multi-layer networks are artificial models of the neural network which converts an input set to an output approximate one. These networks show that they contain more than one neuron layer or nerve with activation function in them. Schematic architecture of MLP network can be seen in fig.1.

Corresponding Author: Fatimah Masoudifar
E-mail: edg2010edg@gmail.com

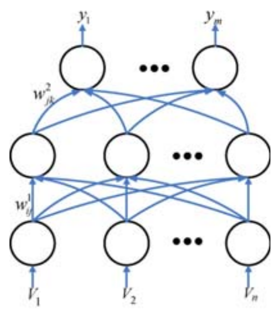

Fig. 1: Schematic architecture of MLP network.

Cybenko proved that only one MLP network with a hidden layer would be adequate for any continual function with specific accuracy. The MLP network with m outputs and n hidden nodes would be as follows (Sadik and Mustafa, 2007):

$$\hat{y}_k(t)=\sum_{j=1}^{n_h} w_{jk}^2 F\left[\sum_{i=1}^{n_i} w_{ij}^1 v_i^0(t)+b_j^1\right]; \quad \text{for } 1\le k\le m \tag{1}$$

Where, W^1_{ij} and W^2_{ij} are the related weights between input and hidden layer and between hidden layers and outputs, respectively. b^1_j and v^0_i is the threshold to input and nodes. N_i and n_h are the numbers of nodes for input and hidden states respectively. Since MLP networks are completely nonlinear and since assuming nonlinear equations as linear ones is done approximately, these two will provide results with high error levels which can be solved by using combined multi-layer networks (fig.2).

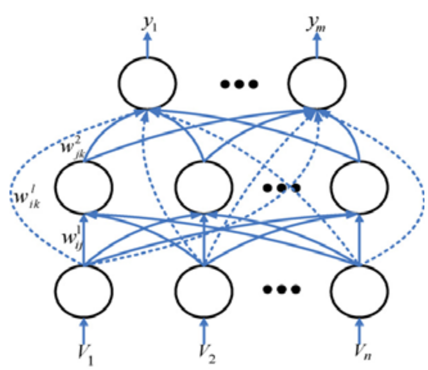

Fig. 2: Combined multi-layer networks.

Combined multi-layer networks can be used as optimal networks to modeling the linear and nonlinear systems. They let the input network to be connected directly to the output nodes with the same weight connections, so that the added linear system would be parallel to the original nonlinear system and to the original MLP model. The K^{th} output in an output layer or Y_k can be obtained from the following equation:

$$\hat{y}_k(t)=\sum_{i=1}^{n_h} w_{jk}^2 v_j^1+\sum_{i=0}^{n_i} w_{ik}^l v_1^0(t); \quad \text{for} \quad 1\le k\le m \tag{2}$$

So we will have:

$$v_j^1(t)=F\left[\sum_{i=1}^{n_i} w_{ij}^1 v_1^0(t)+b_j^1\right] \tag{3}$$

In this section our goal is to reach the minimum possible error which is calculated as below:

$$\varepsilon_k(t) = y_k(t) - y_k(t) \tag{4}$$

In the above formulae, y and y are the actual and predicted values respectively. As the prediction error should reach its minimum value, it is necessary to use proper algorithms. The present study has utilized learning algorithms of MRPEs which will be discussed more in the following.

The above mentioned learning algorithm was modified by Chen who proposed the following formulae (Haykin and Natural Network, 1994):

$$J(\widehat{\Theta}) = \frac{1}{2N}\sum_{t=1}^{N} \varepsilon^T(t, \widehat{\Theta})\Lambda^{-1}\varepsilon(t, \widehat{\Theta}) \tag{5}$$

$$\widehat{\Theta}(t) = \widehat{\Theta}(t-1) + \mathbf{P}(t)\Delta(t) \tag{6}$$

$$\Delta(t) = \alpha_m(t)\,\Delta(t-1) + \alpha_g(t)\psi(t)\varepsilon(t) \tag{7}$$

Where ε(t) is the prediction error, Λ is the positive symmetric matrix, $\alpha_m(t)$ is the momentum and $\alpha_g(t)$ is the rate of linearity. For improvement in convergence rate, it is necessary to do the following modifications:

$$\alpha_m(t) = \alpha_m(t-1) + a \tag{8}$$

And

$$\alpha_g(t) = \alpha_g(0)(1 - \alpha_m(t)) \tag{9}$$

Steps for the learning algorithm of MRPE are as follows:

1) Calculation of the initial weights and thresholds.
2) Adding the input to the network and calculating the outputs from formulae(1).
3) Calculation of the predicted error based on formulae(4).
4) Calculation of the $\psi(t)$ matrix.

$$\psi_k(t) = \frac{dy_k(t)}{d\theta_c} = \begin{cases} v_j^1 & \text{if}\,\theta_c = w_{jk}^2 \quad 1 \le j \le n_h \\ v_i^0 & \text{if}\ \theta_c = w_{ik}^l \quad 0 \le i \le n_i \\ v_j^1(1 - v_j^1)w_{jk}^2 & \text{if}\ \theta_c = b_j^1 \quad 1 \le j \le n_h \\ v_j^1(1 - v_j^1)w_{jk}^2 v_i^0 & \text{if}\ \theta_c = w_{ij}^1 \quad 1 \le j \le n_h,\ \ 1 \le i \le n_i \\ 0 & \text{otherwise} \end{cases} \tag{10}$$

5. calculation of $\mathrm{P}(t)$ and $\lambda(t)$ matrixes,

$$\mathbf{P}(t) = \frac{1}{\lambda(t)}[\mathbf{P}(t-1) - \mathbf{P}(t-1)\psi(t)(\lambda(t)\mathbf{I} + \psi^T(t)\mathbf{P}(t-1)\psi(t))^{-1}\psi^T(t-1)\mathbf{P}(t-1)] \tag{11}$$

$$\lambda(t) = \lambda_0\lambda(t-1) + (1 - \lambda_0) \tag{12}$$

6. calculation of $\alpha_m(t)$, $\alpha_g(t)$, $\Delta(t)$ and $\widehat{\Theta}(t)$.
7. return to step(2).

Clustered-Combined Multi-Layer Networks:

The structure of a clustered-combined multi-layer network introduced in this research is obtained from development of a combined multi-layer network depicted in fig.3.

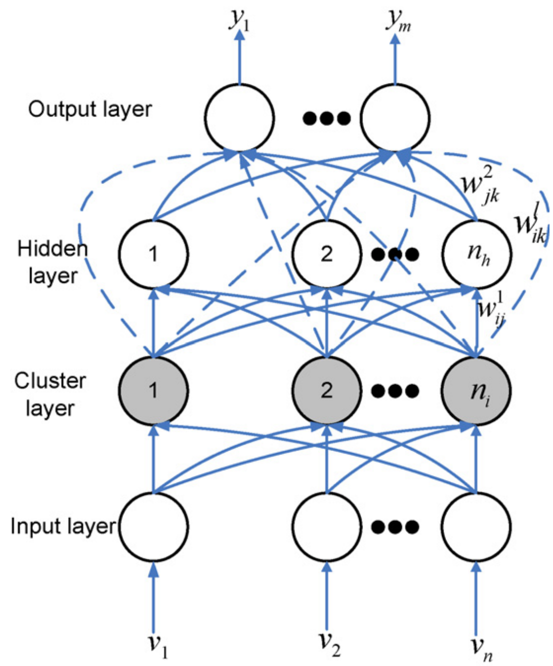

Fig. 3: Clustered-combined multi-layer network.

The output in this network is calculated as the following:

$$v_j^1(t) = F\left[\sum_{i=1}^{n_i} w_{ij}^1(\|v(t) - c_j(t)\|) + b_j^1\right]; \quad \text{for } j = 1, 2, \ldots, n_h \tag{13}$$

Where w_{ij}^1, $v(t)$, b_j^1 and $c_j(t)$, are the weights of clustered layer, hidden layer, input vector and hidden node. For getting the final output, we may use the following formulae:

$$\hat{y}_k(t) = \sum_{j=1}^{n_h} w_{jk}^2 v_j^1(t) + \sum_{i=0}^{n_i} w_{ik}^l(\|v(t) - c_j(t)\|); \text{ for } k = 1, 2, \ldots, m \tag{14}$$

For evaluating the performance from three factors of accuracy, variance and average squared mean error, we continue as following:

$$\text{Accuracy } (\%) = \frac{\text{Total number of cerrectly predicted cases (i.e. in testing phase)}}{\text{Total number of cases (i.e. in testing phase)}} \times 100\% \tag{15}$$

$$\text{Standard deviation} = \sqrt{\frac{\text{Sum of the squared deviations from the sample mean (i.e the testing phase)}}{\text{Number of observations less one (i.e testing phase)}}} \tag{16}$$

$$MSE_{\text{mean}} = \frac{\sum_{i=1}^{10} MSE_i}{10} \tag{17}$$

In the following, we will try to study and evaluate the presented algorithm.

Case Study:

In order to study the presented algorithms, it was tried to use available data from IRIS and to compare the results with some other techniques such as OCI, KS, RBF, LVQ, CN2 and ITI. The following table summarizes the obtained accuracy and standard deviation:

Table 1: Results of the accuracy and the standard deviation from different techniques.

Technique	Accuracy (%)	standard deviation
CLUSTERED-HMLP	93.32	1.17
ITI	85.44	1.98
CN2	82.35	2.11
LVQ	79.65	2.34
RBF	74.34	2.78
KS	71.11	2.99
OCI	68.96	3.07

The results of this table reveal the high accuracy of the proposed model (clustered-HMLP) which is because of clustering behavior in the algorithm since it reduces the standard deviation and leads to higher accuracy and efficiency. It is noteworthy that the table has been prepared from formulas (15) and (16).

In the following, for studying the stability of the system it is tried to set interviews with 48 experts from universities and organizations out of which about 94% of them confirmed the model and the other 6% also accepted it by some 4 revisions in the apparent structure.

Discussion:

The combined algorithm presented was evaluated using IRIS data whose results showed that the clustered-HMLP has been presented with the most accuracy and the least standard deviation. Afterward, other patterns namely ITI, CN2, LVQ, RBF, KS and OCI were chosen as the most suitable patterns respectively with the highest accuracy and the lowest standard deviation. The pattern was also investigated from the stability point of view and about 45 academic and industrial specialists from this field approved it.

These patterns are capable of being implemented in both industrial organizations and academic ones. Regarding the type of pattern and optimization algorithm in them, it is the first time that these are being considered.

Conclusions:

In this paper a combined multi-layer network is introduced in order to improve the performance of common networks, whose clustered algorithm is chosen from network architecture based on RBF and is combined in the multi-layer network architecture. Learning the multi-layer architecture is done based on the MRPE and is evaluated using IRIS data with other methods. In addition to evaluation of credibility, its stability is also studied based on structural procedures.

Further Studies:

One appropriate offer aiming to optimize the presented structure, is using genetics algorithm instead of MRPE algorithm.

REFERENCES

Aajay, P., 1991. Dual network and their pattern classification properties, in: Proceedings of the IEEE Computer Society Conference, pp: 686-687.

Billings, S.A., C.F. Fung, 1995. Recurrent radial basis function network for adaptive noise cancellation, Neural Network, 8(2): 273-290.

Billings, S.A., H.L. Wei, A. Balikhin, 2007. Generalized multiscale radial basis function networks, Neural Network, 20: 1081-1094.

Bishop, C.M., 1995. Neural Networks for Pattern Recognition, Oxford University Press, Oxford, UK.

Chathura, R.D.S., R. Surendra, Liyanage C. De Silva, 2008. Cloud basis function neural network: a modified RBF network architecture for holistic facial expression recognition, Pattern Recognition, 41(4): 1241-1253.

Chen, S., S.A. Billings, P.M. Grant, 1992. Recursive hybrid algorithm for non-linear system identification using radial basis function networks, International Journal of Control, 55: 1051-1070.

Cherkassky, V., F. Mulier, 1998. Learning From Data: Concepts, Theory, and Methods, Wiley Interscience.

Cybenko, G., 1989. Approximations by superposition of a sigmoidal function, Mathematics of Control, Signals and Systems, 2: 303-314.

Darken, C., J. Moody, 1990. Fast adaptive K-means clustering: some empirical result, International Joint Conference on Neural Networks, 2: 233-238.

Funashashi, K., 1989. On the approximation realization on continuous mappings by neural networks, Neural Network, 2: 182-192.

Gieger, H., 1990. Storing and processing information in connectionist system, in: R. Eckmiller (Ed.), Advance Neural Computers, North-Holland, Amsterdam, pp: 271-277.

Gonzalez, J., I. Rojas, J. Ortega, H. Pomares, J. Fernandez, A.F. Diaz, 2003. Multiobjective evolutionary optimization of the size, shape, and position parameters of radial basis function networks for function approximation, IEEE Transactions on Neural Networks, 14(6): 1478-1495.

Haykin, S., Neural Network, 1994. A Comprehensive Foundation, Prentice Hall, NJ.

Hoang, A., 1997. Supervised classifier performance on the UCI database, B.Sc. Thesis, University of Adelaide.

Ljung, L., T. Soderstrom, 1983. Theory and Practice of Recursive Identification, MIT Press, Cambridge, MA.

Lung, S.Y., 2007. Efficient text independent speaker recognition with wavelet feature selection based multilayered neural network using supervised learning algorithm, Pattern Recognition, 40(12): 3616-3620.

Mashor, M.Y., 1995. System identification using radial basis function network, Ph.D. Thesis, University of Sheffield, UK.

Mashor, M.Y., 1998. Supervised hybrid training algorithm for RBF network, International Journal of the Computer, the Internet and Management, 6(3): 13-27.

Mashor, M.Y., 1999. Someproperties of RBF network with applications to system identification, International Journal of the Computer, the Internet and Management, 7(1): 34-71.

Mashor, M.Y., 2000. Hybrid multilayer perceptron networks, International Journal of Systems Sciences, 31(6): 771-785.

Mashor, M.Y., 2000. Hybrid training algorithm for RBF network, International Journal of the Computer, the Internet and Management, 8(2): 50-65.

Mashor, M.Y., 2000. Performance comparison between back propagation, RPE and MRPE algorithms for training MLP networks, International Journal of the Computer, the Internet and Management, 8(3): 65-79.

Mashor, M.Y., 2003. Modified recursive prediction error algorithm for training layered neural network, International Journal of the Computer, the Internet and Management, 11(2): 29-36.

Mat-Isa, N.A., M.Y. Mahor, N.H. Othman, 2003. Comparison of segmentation performance of clustering algorithm for Pap smear images, in: Proceeding of the International Conference Robotic, Vision, Information and Signal Processing, pp: 118-125.

Ozgrid Business Applications, Neuro Classifier is Neural Network Software for Classifying Data in Microsoft Excel, Available from http://www.ozgrid.com/Services/neuro-excel-classifier.htm.

Poggio, T., F. Girosi, 1990. Network for approximation and learning, Proceeding of IEEE, 78(9): 1481-1497.

Rumelhart, D.E., J.L. McCleland, 1986. Parallels Distributed Processing: Explorations in The Microstructure of Recognition, I & I, MIT Press, Cambridge, MA.

Sadik, K., O. Mustafa, 2007. Atrial fibrillation classification with artificial neural networks, Pattern Recognition, 40(11): 2967-2973.

Sarle, W.S., 1994. Neural networks and statistical models, in: Proceedings of the Nineteenth Annual SAS Users Group International Conference, SAS Institute Inc., Cary, NC, pp: 1533-1550.

UC Irvine Machine Learning Repository, Availabe from http://www.ics.uci.edu/ ?mlearn/MLRepository.html.

Werbos, P.J., 1974. Beyond regression: new tools for prediction and analysis in the behavioral sciences, Ph.D. Thesis, Harvard University.