MULTI OBJECTIVE GENETIC ALGORITHM FOR TRAINING THREE TERM BACKPROPAGATION NETWORK

Ashraf Osman Ibrahim[1, 2], Siti Mariyam Shamsuddin[1], Nor Bahiah Ahmad[1], and Sultan Noman Qasem[3]

1 Soft Computing Research Group, Faculty of Computing, Universiti Teknologi Malaysia, 81310 UTM Skudai Johor, Malaysia, ashrafosman2@gmail.com, mariyam@utm.my, bahiah@utm.my.
2 Faculty of Computer Science and Information Technology, Alzaeim Alazhari University (AAU), Khartoum, Sudan.ashrafosman2@gmail.com.
3 Computer Science Department, College of Computer and Information Sciences, Al-Imam Muhammad Ibn Saud Islamic University, Riyadh, Saudi Arabia, sultann.noman@ccis.imamu.edu.sa.

ABSTRACT. Multi Objective Evolutionary Algorithms has been applied for learning problem in Artificial Neural Networks to improve the generalization of the training and testing unseen data. This paper proposes the simultaneous optimization method for training Three Term Back Propagation Network (TTBPN) learning using Multi Objective Genetic Algorithm. The Non-dominated Sorting Genetic Algorithm II is applied to optimize the TTBPN structure by simultaneously reducing the error and complexity in terms of number of hidden nodes of the network for better accuracy in classification problem. This methodology is applied in two kinds of multi classes data set obtained from the University of California at Irvine repository. The results obtained for training and testing on the datasets illustrate less network error and better classification accuracy, besides having simple architecture for the TTBPN.

Keywords: Artificial Neural Networks, Multi-objective evolutionary, Three Term Back Propagation, Non-dominated Sorting Genetic AlgorithmII

INTRODUCTION

Artificial Neural Networks (ANNs) are one of the powerful machine learning methods. Recently, ANNs has been used widely in different applications (Cheok, Chin, Yusof, Talib, & Law, 2012; Khosrowshahi, 2011; Kuo & Lin, 2010). However, sometimes, the errors of the network are not good enough, and hence affect the network performance. This limitation of ANNs may lead to network complexity. Moreover, ANNs need to optimize the networks in order to achieve higher performance and accuracy.

Evolutionary Algorithms (EAs), are a good candidate for Multi objective optimization problems because of their abilities to search for multiple Pareto optimal solutions. In addition, they perform better in global search space. Equally, Pareto optimal solutions are used to evolve ANNs (Qasem and Shamsuddin, 2011), which are optimal with respect to both the classification accuracy and architectural complexity. Therefore, Multi Objective Evolutionary Algorithms (MOEAs) for the learning problem is applied to improve the generalization of the training and testing unseen data. In this paper the Non-dominated Sorting Genetic Algorithm II (NSGA-II) is applied to improve the generalization of the TTBPN, by improving the complexity in terms of the number of hidden nodes and errors of the network simultaneously. One of the most successful applications of the EAs is used for evolving ANNs, as in (Yao, 1999). The author provided a general framework for evolving

ANNs by employing Genetic Algorithms (GAs) for optimizing ANNs. Moreover, Multi-Objective Genetic Algorithm (MOGA) optimization used by (Pettersson et al., 2007) for training a feed forward neural network was able to minimize the training error and the network size using noisy data. The number of nodes, the ANNs architecture, as well as the weights, and a Pareto front was effectively constructed.

Another method used by Generalized Multi-layer Perceptrons also improve the performance of the evolutionary model. The (Delgado et al., 2008) proposed a hybrid MOGA method based on the Strength Pareto Evolutionary Algorithm-2 (SPEA2) and NSGA-II algorithms to optimize the training and the topology of the Recurrent Neural Network simultaneously. (Jin et al., 2005) used MOGA and focused on the problem of multi objective optimization for feed-forward ANNs as a solution for the regularization problems in the network’s complexity. In (Liu and Kadirkamanathan, 1999), the authors had considered both optimizing the size of neural networks in relation to addressing the benefits of multi-objective optimization for identifying nonlinear systems. Furthermore, (Abbass and Sarker, 2001) introduced a multi objective method that includes differential evolution algorithm to train the network for a single layer perceptron and to find the optimal size of hidden nodes. Thoroughly, the optimization of the structure is carried out by minimizing the number of network connections. Even though, numerous studies offered reasonable solutions for feed-forward ANNs, (Qasem and Shamsuddin, 2011)presents a new multi-objective evolutionary hybrid algorithm for the design of Radial Basis Function Networks for classification problems. Also, (Cruz-Ramírez et al., 2012) introduced a multi-objective evolutionary learning algorithm using an improved version of the NSGA-II algorithm called MPENSGA-II hybridized with a local search algorithm for training ANNs with generalized radial basis functions.

This work has developed multi objective genetic algorithm and TTBPN by optimizing the structure in terms of the number of hidden nodes and errors of TTBPN simultaneously for solving multi class pattern classification problems.

The rest of this paper is organized as follows: The next section briefly introduced the related method, followed by the complete methodology being conducted and the use of NSGA-II for training TTBPN. Further, the experimental results of the designed TTBPN for classification problem are presented and finally, the paper ends with the conclusion of the findings.

THREE TERM BACKPROPAGATION ALGORITHM (TTBP)

The Three Term Back propagation was proposed by (Zweiri et al., 2003) employs the standard architecture and procedure of the standard BP algorithm. However, the third parameter called proportional factor (PF) is introduced. This is proven to be successful in improving the convergence rate of the algorithm and speed up the weight adjusting process. Due to the success of TTBPN some studies have used this algorithm in different application (Abdulkadir et al.; Mashinchi and Shamsuddin, 2009).

MULTI-OBJECTIVE GENETIC ALGORITHMS (MOGAS)

The genetic algorithm (Gas) is suited to solve multi-objective optimization problems. Nevertheless, many optimization problems have multiple objectives. Historically, multiple objectives have been combined to form a scalar objective function, commonly through a weighted sum of the multiple objectives, or by turning objectives into constraints with associated thresholds and penalty functions.

NSGA-II Algorithm

The non-dominated sorting genetic algorithm II (NSGA-II) is one of the MOGAs proposed by (Deb et al., 2002), for it's a good performance of global searching a non-dominated sorting multi objective optimization genetic algorithm becomes a preferred method of optimization algorithm.

THE PROPOSED METHOD

The proposed algorithm is a MOGA optimization approach based on NSGA-II for TTBPN training implemented, and called MOGATTBPN. However, MOGATTBPN begins by collecting, normalizing and dividing the data into training and testing datasets. The number of hidden nodes and the maximum number of iterations are set and the individual length is computed. Subsequently, the parameters of TTBPN are determined by the traditional algorithms. Then a population of TTBPN is generated and initialized. For every iteration each individual is evaluated based on objective functions. After the maximum iterations are reached the proposed method stops and outputs a set of non-dominated TTBPNs.

To evaluate the TTBPN performance for all algorithms, two objective function will be used in this study as follows:

1. The performance of the network (Accuracy) based on the Mean Square Error (MSE) on the training set, this performance as a first objective function is given as Eq.(1):

$$f_1 = \frac{1}{N}\sum_{j=1}^{N}\left(t_j - o_j\right)^2 \tag{1}$$

Where ,

o_j Network error at output unit, t_j target value of output, N number of samples,

2. The complexity of the network based on the number of hidden nodes in the hidden layer of TTBPN, as a second objective function and it is given as Eq.(2):

$$f_2 = \sum_{h=1}^{H} \rho_h \tag{2}$$

Where,

$\rho_h \in \rho$, vector ρ is the dimension of maximum number of hidden nodes H of the network, and ρ is binary value used to refer to the hidden node if it exists in the network or not. It works as a switch to turn a hidden unit ON or OFF and is the maximum hidden nodes of TTBPN.

EXPERIMENTAL STUDY

This section presents the experimental study on MOGA and trained TTBPN. The proposed algorithm is evaluated by using 10-fold cross validation technique. In the experimental design, we considered three multi class data sets listed in Table 1. The dataset have been widely used in pattern classification. All data sets used in this study are obtained from the (A. Asuncion and D.J. Newman, 2007).

Table 1. Summary of data sets used in the experiments

Dataset	Number of features	Number of classes	Number of patterns
Iris	4	3	150
Wine	13	3	178
Yeast	8	10	1484

Results and Discussion

From Table 2, we clearly notice that the statistical results for sensitivity, specificity and accuracy of the proposed method, the result of iris data obtained an accuracy of 77.556 % , 74.292 % accuracy of wine data and the yeast data obtains 90.00 % as a highest accuracy result in used data. Equally, for the sensitivity wine data obtain 99.116% as a highest sensitivity rate. The sensitivity of the yeast data set is very difficult, due to their unbalanced data. Besides, accuracy and sensitivity Table 2 show the specificity for all datasets, we can note that the specificity rate was achieved as follows; the 100% in yeast is extremely high value in Specificity. Figure 1 shows the accuracy of the proposed method for all dataset.

Table 2. The average and standard deviations of training and testing accuracy

Dataset		Training Sensitivity	Training Specificity	Training Classification	Testing Sensitivity	Testing Specificity	Testing Classification
Iris	Mean	34.88889	99.40740667	78.17284	34	99.3333333	77.55556
	STD	27.07382	1.079102858	8.193483	24.8352597	2.10818511	7.729992
Wine	Mean	23.32268	98.65752667	73.3516	99.1161633	74.29195	74.29195
	STD	35.71279	2.393274692	10.41546	2.01756695	11.94629	11.94629
Yeast	Mean	0	100	90.00	0	100	90.01
	STD	0	0	0	0	0	0

Table 3. Two objectives optimization on the training error and error testing

Dataset		Training Error	Testing Error
Iris	Mean	0.16454	0.16536
	STD	0.023873	0.02238
Wine	Mean	0.16862	0.16818
	STD	0.039393	0.04328
Yeast	Mean	0.08157	0.08161
	STD	0.008783	0.008791

The generalization error rates of the proposed method for all datasets are shown in Table 3. It can be observed that in all datasets, the proposed method is giving promising results in the performance of both training and testing error. Furthermore, the training and testing error are considered reasonable error values with average error rates obtained in a single run of the MOGA to TTBP.

Table 4. The comparison of the accuracy and hidden nodes obtained by the proposed method and other methods

methods	MOGATTBP		MEPGANf1f2		MEPGANf1-f3	
Dataset	Accuracy	Hidden node	Accuracy	Hidden node	Accuracy	Hidden node
Iris	77.55556	**3.6**	83.78	8.2	**84.44**	6.9
Wine	**74.29195**	**4.6**	72.18	5.5	72.04	5.6
Yeast	**90.01**	**3.5**	90.00	6.5	90.01	6.0

Table 4 lists the comparison of the classification accuracy and the number of hidden nodes for MOGATTBP with other methods such as MEPGANf1f2 and MEPGANf-f3 which have been implemented in the literature review. MOGATTBPN achieved better classification accuracy than other methods in wine and yeast dataset. But, in iris data MEPGANf1f2 and MEPGANf-f3 are better than MOGATTBPN which are due to the local search algorithm that the two methods employed to enhance all individuals in the population. Regarding the complexity of the network, the MOGATTBPN achieved better network structure with the lowest complexity compared with MEPGANf1f2 and MEPGANf-f3 (showed in bold font). Moreover, Figure 2 and Figure 3 show the comparison of the accuracy and hidden nodes (complexity) with other methods respectively.

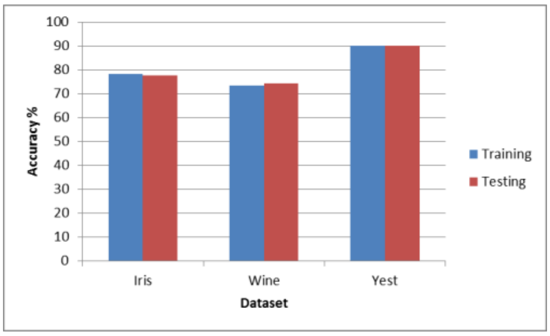

Figure 1. The average of training and testing accuracy using MOGATTBPN

Figure 2. Comparison of the accuracy for MOGATTBPN and other methods

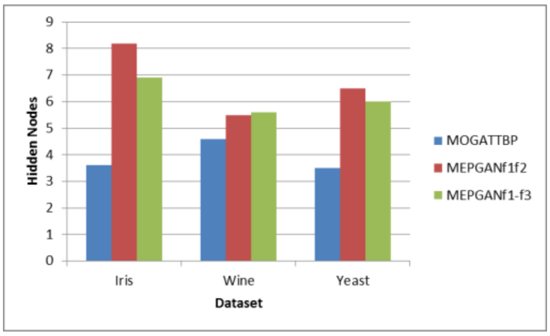

Figure 3. Comparison of the hidden nodes for MOGATTBPN and other methods

CONCLUSION

In this paper, new methods for training TTBPN has been applied on multi class datasets. The training set is used to train the TTBPN in order to get the Pareto optimal solutions, while the testing set is used to test the generalization performance of Pareto TTBPN. For future work, we will evaluate the performance of the proposed method for two-class, multi class and complex real problem pattern classification.

ACKNOWLEDGMENTS

The authors would like to thank Soft Computing Research Group, UTM and IDF for their continuous support and motivation. This work is also supported by GUP Research Grant Scheme (Q.J130000.2628.09J07).

REFERENCES

A. Asuncion and D.J. Newman. (2007). UCI Machine Learning Repository. *http://www.ics.uci.edu/$\sim$mlearn/{MLR}epository.html*.

Abbass, H. A., and Sarker, R. (2001). *Simultaneous evolution of architectures and connection weights in ANNs.* Paper presented at the Artificial Neural Networks and Expert Systems Conference, 16-21.

Abdulkadir, S. J., Shamsuddin, S. M., and Sallehuddin, R. Three Term Back Propagation Network for Moisture Prediction.

Chun-Dong, X., Wei, L., Mu-Gui, Z., and Jin-Gao, L. (2012). A BP Neural Network Activation Function Used in Exchange Rate Forecasting. In *Information Technology and Agricultural Engineering* (pp. 69-76): Springer.

Cruz-Ramírez, M., Hervás-Martínez, C., Fernández, J. C., Briceño, J., and de la Mata, M. (2012). Multi-Objective Evolutionary Algorithm for Donor-Recipient Decision System in Liver Transplants. *European Journal of Operational Research*.

Deb, K., Pratap, A., Agarwal, S., and Meyarivan, T. (2002). A fast and elitist multiobjective genetic algorithm: NSGA-II. *Evolutionary Computation, IEEE Transactions on, 6*(2), 182-197.

Delgado, M., Cuéllar, M. P., and Pegalajar, M. C. (2008). Multiobjective hybrid optimization and training of recurrent neural networks. *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on, 38*(2), 381-403.

Jin, Y., Sendhoff, B., and Körner, E. (2005). *Evolutionary multi-objective optimization for simultaneous generation of signal-type and symbol-type representations.* Paper presented at the Evolutionary Multi-Criterion Optimization, 752-766.

Liu, G., and Kadirkamanathan, V. (1999). *Multiobjective criteria for neural network structure selection and identification of nonlinear systems using genetic algorithms.* Paper presented at the Control Theory and Applications, IEE Proceedings-, 373-382.

Mashinchi, M. H., and Shamsuddin, S. M. H. (2009). Three-Term Fuzzy Back-Propagation. In *Foundations of Computational, Intelligence Volume 1* (pp. 143-158): Springer.

Pettersson, F., Chakraborti, N., and Saxén, H. (2007). A genetic algorithms based multi-objective neural net applied to noisy blast furnace data. *Applied Soft Computing, 7*(1), 387-397.

Qasem, S. N., and Shamsuddin, S. M. (2011). Memetic Elitist Pareto Differential Evolution Algorithm Based Radial Basis Function Neural Networks for Classification Problems. *Applied Soft Computing*.

Shamsuddin, S. M., Sulaiman, M. N., and Darus, M. (2001). An improved error signal for the backpropagation model for classification problems. *International Journal of Computer Mathematics*, *76*(3), 297-305.

Siti Mariyam Shamsuddin, S. Z. M. H., Maslina Darus and Ana Salwa Shafie. (2008). Improved Backpropagation Error Function for Classification Problems. *International Journal of Mathematical & Application, 1*, 9-25.

Yao, X. (1999). Evolving artificial neural networks. *Proceedings of the IEEE, 87*(9), 1423-1447.

Zweiri, Y., Whidborne, J., and Seneviratne, L. (2003). A three-term backpropagation algorithm. *Neurocomputing*, *50*, 305-318.