*Int. J. Advance. Soft Comput. Appl., Vol. 5, No. 2, July 2013*
*ISSN 2074-8523; Copyright©SCRG Publication, 2013*

# Weights Adjustment of Two-Term Back-Propagation Network Using Adaptive and Fixed Learning Methods

**Citra Ramadhena[1] Ashraf Osman Ibrahim[2] & Sarina Sulaiman[1]**

[1]*Soft Computing Research Group*, Faculty of Computing, UniversitiTeknologi Malaysia 81310 Skudai Johor, Malaysia
[2]Faculty of Computer Science and Information Technology, AlzaeimAlazhari University ( AAU), Khartoum, Sudan.e-mail:ashrafosman2@gmail.com
[1]*Soft Computing Research Group*, Faculty of Computing, UniversitiTeknologi Malaysia 81310 Skudai Johor, Malaysia.e-mail:sarina@utm.my

**Abstract**

***Artificial Neural Networks (ANNs) using the Back Propagation algorithm (BP) mainly depend on weights adjustment in training learning. The solutions can be faster by properly adjusting the magnitude and sign of the weights. Weight Changes is one of the solutions for a common problem faced by the two-term back-propagation network that suffers from slow convergence and trapping in a local minima. Many ways have offered to generate the proper weight magnitude and sign of the network. One of the solutions by adjusting a correct value of learning rate and momentum parameters to improve the performance of ANN by implementing an adaptive learning method. This paper implements both adaptive and fixed learning methods for two-term BP algorithms. Standard datasets of UCI machine learning are used with n-fold cross validation to train and test both methods to properly adjust and investigate the changes of weight sign accordingly. The results showed that the two-term BP using the adaptive learning method work better in producing proper weight changes with minimum time compared to two-term BP with fixed learning rate.***

**Keywords**: *Weight changes, adaptive learning, Two-Term BP, Mean square error, convergence rate, Supervised Learning .*

# 1 Introduction

The major impact towards the learning behavior of the network brings by the value assignment of the weight. If the algorithm successfully computes the correct value of the weight, it can converge faster to the solution; otherwise, the convergence might be slower or it may cause divergence. To prevent this problem happening, the step of gradient descent is controlled by a parameter called learning rate. This parameter will determine the length of step taken by the gradient to move along the error surface. Moreover, to avoid the oscillation problem that might happen around the steep valley, the fraction of the last weight update is added to the current weight update and the magnitude is adjusted by a parameter called momentum. The inclusion of these parameters is aiming to produce a correct value of weight update which later will be used to update the new weight. The correct value of weight update can be seen in two aspects such as sign and magnitude. If both aspects are properly chosen and assigned to the weight, the learning process can be optimized and the solution is not hard to reach. Due to the usefulness of two-term BP and the adaptive learning method of learning the network, this study proposing the weights sign changes with respect to gradient descent in two-term BP network with and without adaptive learning method.

Despite the general success of BP in learning, several major deficiencies are still needed to be solved. The most notable deficiencies, according to reference (Ng *et al.*, 1999) are the existence of temporary local minima due to the saturation behavior of activation function. The slow rates of convergence due to the existence of local minima and the convergence rate are relatively slow for network with more than one hidden layer. These drawbacks are also acknowledged by several scholars (Dhar *et al.*, 2010; Hongmei and Gaofeng, 2009; Yu and Liu, 2002; Zweiri *et al.*, 2003). There are many improvement approaches have been proposed to perform of BP. Such as, too early saturation avoidance, weight adjusting, normalization, interaction, learning rate changing and stimulation function changing (Dai-yuan, 2006; Eom *et al.*, 2003; Magoulas *et al.*, 2002; Trentin, 2001; Wang *et al.*, 2007; Yam and Chow, 2000; Yan, 2004; Yu and Chen, 1997; Zhang *et al.*, 2008; Zweiri *et al.*, 2003).

The rest of this paper is organized as follows. Section 2 provides related works, followed section 3 the proposed method and section 4 resultand discussion. The conclusions are also provided in section 5 and future work in section 6.

# 2 Related Works

There are many studies in literature for solving the problem of ANNs training and learning with different method. Most of them have been used two-term BP algorithm. (Yu and Liu, 2002) Proposed a backpropagation algorithm with adaptive learning rate and momentum. The modification of conventional back-propagation algorithm in the proposed algorithm that uses adaptive learning rate and momentum where the learning rate and the momentum are adjusted ateach iteration to speed up the training time. The modified back-propagation with adaptive learning rateand momentum outperforms the conventional back-propagation with fixed momentum or without momentum in term of learning speed.

On the other hand, (Duffner and Garcia, 2007) has presented a new learning algorithm for feed-forward neural network based on two-term BP method using adaptive learning rate. The adaptation is based on the error criteria where error is measured in the validation set instead of training set to dynamically adjust the global learning rate. The proposed algorithm consists of two phases. In the first phase, the learning rate is adjusted after each iteration so as the minimum error is quickly attained. While the second phase, the search algorithm is refined by repeatedly reverting to previous weight configurations and decreasing the global learning rate. The experimental result shows that the proposed method quickly converges and outperforms two-term BP in term of generalization when the size of the training set is reduced. (Shamsuddin *et al.*, 2001) has improved the convergence rates of two-term BP model with some modification in learning strategies. The experiment results show that the modified two-term BP improved with a convergence rate much better when compared with standard BP. Meanwhile, (Iranmanesh and Mahdavi, 2009) proposed a differential adaptive learning rate method for BP to speed up the learning rate. The proposed method employs the large learning rate at the beginning of training and gradually decreases the value of learning rate using the differential adaptive method. The comparison made between this method and other methods, such as two-term BP, Nguyen-Widrow weight Initialization and Optical BP shows that the proposed method outperforms the competing method in term of learning speed.

In (Hua Li and Xiangji Huang, 2012), the two-term BP is improved that can overwhelm the problems of slow learning and easy to tap into minimum by adopting an adaptive algorithm. The method divides the whole training process into many learning phases. The effects will indicate the direction of the network globally. A different range of effect values corresponds to different learning models. The next learning phase will adjust the learning model based on the evaluation effects according to previous learning phase.

To minimize the error and increase the convergence speed, (Subavathi and Kathirvalavakumar, 2011) proposed a new efficient modified back propagation algorithm with adaptive learning rate. The method eliminates initial fixing of learning rate through trial and error and replaces by adaptive learning rate. In each iteration, adaptive learning rate for output and hidden layer are determined by calculating differential linear and nonlinear errors of output layer and hidden layer separately.(Zweiri *et al.*, 2003) proposed a new approach to calculate the change of weight for the link joining the $j^{th}$ unit to the $i^{th}$ unit.

(Li *et al.*, 2009) presented the improved training algorithm of BP with self-adaptive learning rate. The functional relationship between the total quadratic training error change, the connection weight and bias change is acquired based on the Taylor formulation. By combining it with weight and bias change in the batch BP algorithm, the equations to calculate self-adaptive learning rate is obtained. The learning rate will be adaptively adjusted based on the average quadratic error and the error curve gradient. The value of the self-adaptive learning rate depends on neural network topology, training samples, average quadratic error and gradient but not artificial selection. The result of the experiment shows the effectiveness of the proposed training algorithm. Adaptive learning rate algorithm to train a single hidden layer neural network was proposed in (Kathirvalavakumar and Subavathi, 2012). The adaptive learning rate is derived by differentiating linear and nonlinear errors and functional constraints weight decay term at hidden layer and a penalty term at the output layer. Through the adaptive learning rate calculation involves first order derivative of linear and nonlinear errors and second order derivatives of functional constraints, the proposed algorithm converges quickly. (Shamsuddin *et al.*, 2013) also has studied the weight changes sign with respect to the temporal behaviour of gradient to study the learning behaviour of the network and also to measure the performance of the two-term BP algorithm.

(Xiaoyuan *et al.*, 2009) gave the proper condition for the rate of weight and the temporal behavior of the gradient. The author wrote that if the derivative has the same symbol with the previous one then the sum of the weight is increased that makes the weight increment value larger yields the increment of weight rate. While (Fukuoka *et al.*, 1998) gave a brief definition of online learning and its difference from batch learning proposed. The author defined online learning as one of the schemes of updating weight that updates weight after every input-output case while batch learning accumulates error signal over all the input-output cases before updating weight. As stated by reference (Riedmiller, 1994) , each weight update tries to minimize the error. The author also stated that the summed gradient information for the whole pattern set provides reliable information regarding the shape of the entire error function. To calculate optimum output weights (Idris *et al.*, 2009) trained BP network using learning object data,and Output-Weight-Optimizationinvolves solving a set of linear equations using the Conjugate Gradient Technique .

(Norhamreeza Abdul Hamid *et al.*, 2011) proposed a new modified back propagation learning algorithm introduced an adaptive gain with adaptive momentum together and adaptive learning rate into the weight update process, a better convergence rate and a good solution was obtained. (Hu *et al.*, 2011) a simple method and reliably proposed. The process of samples self-learning, network's generalization ability can be improved effectively. (Zhang *et al.*, 2012) introduced and adaptive learning rate adjustment factor is added; the results showed that decreased Convergence of improved BP network. A new strategy of dynamic change learning rate in BP neural network was proposed in (Guangjun *et al.*, 2008), it changes the learning rate value according to the change of system error .

From the previous works that we highlighted above and their attempts to improve the two-term BP network training and learning, there are still open issues on the enhancement of BP algorithm in training and learning the network especially in terms of weight adjustments. A lot of work needs to improve two-term BP network training and learning, anyway, the weight adjustment is very important to tuning the network learning in two-term BP algorithm.

# 3 The Proposed Method

The adaptive methods will be implemented in two-term BP with mean square error (MSE) that will be briefly discussed in the next section. The two-term BP algorithm will implement the batch earning method where the weight will be updated after the presentation of input and target to the network. Thus, the adaptive methods and weight adjustment are implemented after all data are presented to the network. The algorithm follows the standard forward and backward propagation algorithm.

## 3.1 Two-Term Back-propagation BP Network

Two-term BP algorithm is the most used neural networks training. It is also the most mature of the training algorithm, and has been widely applied in many fields. That is for its benefits of simple structure, maneuverability, less calculation, strong concurrency, can simulate an arbitrary non-linear input-output relationship.

## 3.2 Mean Square Error

Mean square error (MSE) is one way to measure the average of the squared error. The word "error" here refers to the difference between the estimator (the target value) and the true value of the amount being estimated (the calculated value). Moreover, the MSE is a quadratic function, and hence, the minimization of error is carried out by employing the gradient descent to search for the appropriate parameter that can bring smaller error. As shown in Equation (1), the MSE equation is as below:

$$E = \frac{1}{2}\sum_{k=1}^{L}\left(t_k - o_k\right)^2 \qquad (1)$$

where,

$t_k$ *is the target value for the output node* $k$ .

$o_k$ *is the network output for the output node k.*

$L$ *is the number of nodes in the output layer .*

## 3.3 Weight Sign and Adaptive Methods

Learning rate and momentum coefficient are the most commonly used parameter in the two-term BP. The use of constant value of the parameter is not always a good idea. In the case of learning rate, setting up smaller value to learning rate may decelerate the convergence speed even though it can guarantee the gradient to move in the correct direction. On the contrary, setting up larger value to learning rate may fasten the convergence speed but is prone to oscillation problem that may lead to divergence. Due to the excellent idea and performance of the algorithm as has been proven in reference (Minai and Williams, 1990), this method is proposed to assist the network in producing proper weight sign change and achieve the purpose of this paper. As shown in Fig.1. The two-term BP training with adaptive learning methods.

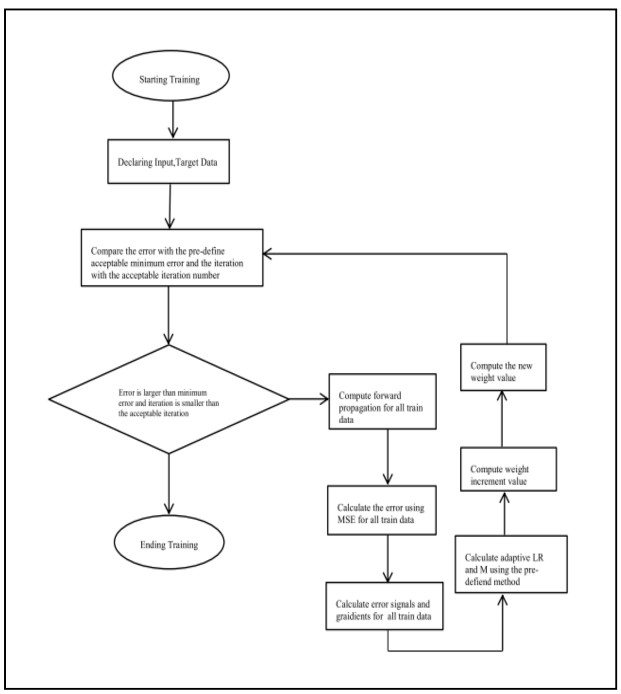


Fig.1: Flow chart of two-term BP training with adaptive learning methods

# 4 Result and Discussion

In this section we discuss the datasets that have been used and the analysis in this study is validated using K-fold cross validation method. The detailed discussion of the experiment is given below

## 4.1 Datasets

To achieve a better result from this study, two-term BP algorithm with and without adaptive learning method is applied to small and medium dataset as illustrated in Table 1. In this study, we define the small dataset contains 16 instances and 150 instances while the medium dataset contains 625 instances. The small dataset are represented by balloon and iris, while the medium dataset is represented by the balance-scale. All the datasets are obtained from the UCI Machine Learning Repository.

Table 1: Summary of datasets

| | **Balloon** | **Iris** | **Balance Scale** |
|---|---|---|---|
| **Input** | 4 | 4 | 4 |
| **Output** | 1 | 3 | 3 |
| **Instance** | 16 | 150 | 625 |

## 4.2 Cross validation

It is a common practice in machine learning and data mining to perform N-fold cross-validation to assess the performance of a classification algorithm (McLachlan *et al.*, 2004).

The cross validation is employed in the experimentation with the intention of getting better results. The cross validation consists of swapping the training set and the validation set, in the way that each one is used for the opposite purpose. This method assures that any tendency found in the results is, in fact, just tender, and not causality. Thus, the database is randomly partitioned into two sets of equal size that are in turns used as training and validation subsets (Fiszelew *et al.*, 2007).

The learning process in two-term BP aims to minimize the errors by optimizing the parameters so that it can generate the proper weight to improve the learning. At the end of the training, the network with its optimized parameters fits the training data. However, when we take an independent sample of validation data, we found out that the network does not fit the validation data better than the one in the training data. In (Sterlin, 2007) added that when the network too dependent on the training data which causes the network to learn irrelevant detail which leads to the lack of generalization in learning.

Table 2:10-fold cross validation of two-term BP with a **dynamic parameter** value for Balloon

| **Fold** | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** | **9** | **10** |
|---|---|---|---|---|---|---|---|---|---|---|
| **Accuracy** | 2 | 1 | 1 | 2 | 2 | 1 | 0 | 1 | 1 | 1 |
| **Num of Iter** | 62 | 29 | 32 | 60 | 64 | 34 | 35 | 44 | 47 | 100 |
| **Conv. Time** | 0.4543 | 0.3676 | 0.4375 | 0.4518 | 0.4861 | 0.3912 | 0.3382 | 0.3605 | 0.3941 | 0.4902 |
| **MSE** | 0.0092 | 0.0094 | 0.0095 | 0.007 | 0.0096 | 0.0096 | 0.0095 | 0.0096 | 0.0096 | 0.0078 |

Table 3: 10-fold cross validation of two-term BP with **fixed parameter** value for Balloon

| **Fold** | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** | **9** | **10** |
|---|---|---|---|---|---|---|---|---|---|---|
| **Accuracy** | 1 | 1 | 2 | 1 | 2 | 2 | 0 | 1 | 1 | 1 |
| **Num of Iter** | 296 | 1000 | 314 | 1000 | 1000 | 1000 | 303 | 1000 | 1000 | 1000 |
| **Conv. Time** | 1.2811 | 2.8698 | 0.9364 | 2.6071 | 2.5918 | 2.7047 | 0.9125 | 2.6325 | 2.57 | 3.9935 |
| **MSE** | 0.001 | 0.0636 | 0.001 | 0.0755 | 0.6999 | 0.1073 | 0.001 | 0.0548 | 0.0333 | 0.0332 |

10-fold cross validation is an example of n-fold cross validation that is widely used in estimating the generalization of performance of the network in BP. Due to its advantages of 10-fold cross validation highlight in (Govindarajan and Chandrasekaran, 2007). The authors stated that the 10-fold cross validation is a recommended method to estimate the accuracy since it is low bias and variance.

The accuracy is calculated from the result of testing at all folds. The total number of correctly classified data at all folds is divided by total of data pattern. Table 2 and Table 3 are the results for Balloon dataset with 10 fold cross validation.

We can see from Tables 2 and 3, that both algorithms are equally accurate. However, two-term BP with the dynamic parameter value still outperforms the two-term BP with fixed parameter value. It can be seen from the number of iterations, MSE and convergence time throughout the folds at Table 2 which is less than the one in Table 3. From the result of 10-folds cross validation implementation on both algorithms, the estimation of generalization performance of both algorithms show that two-term BP with adaptive learning method provides a better generalization in data learning compared to the fixed learning method.

## 4.3 Analysis on the Accuracy and Convergence Speed

In this section we applied MSE error of two-term BP with adaptive learning and fixed learning methods. Fig. 2 illustrates the MSE for adaptive learning and fixed learning methods. The behavior of the convergence speed and accuracy for the balloon datasets are given in Fig. 3.

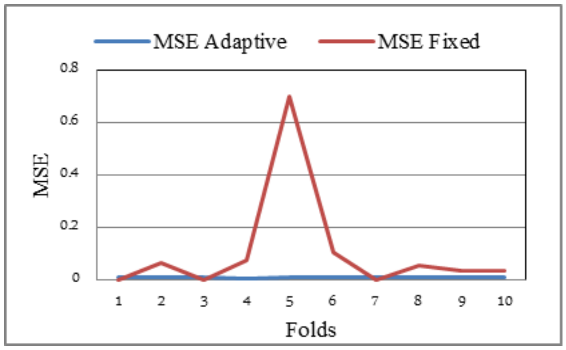


Fig.2. MSE Error in adaptive and fixed learning for Balloon Dataset

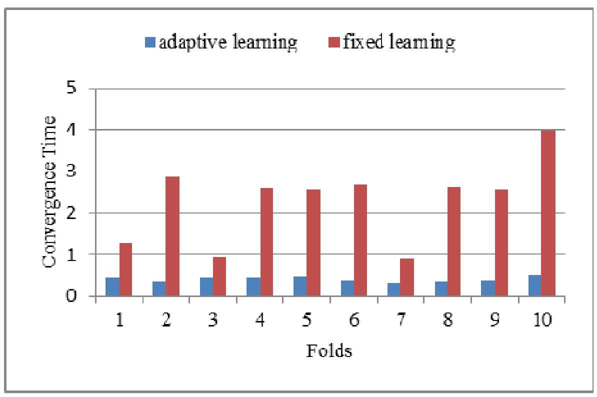


Fig.3. Convergence Time in adaptive and fixed learning for Balloon Dataset

### 4.3.1 Iris Dataset

For Iris dataset, the convergence speed is 0.3591s, with an accuracy of 95.7447% for the adaptive learning method. While for the fixed learning method, the convergence time is 0.7476s and an accuracy of 97.8723%. Fig.4 and Fig.5 illustrate the MSE error together with number of iterations for both adaptive learning and fixed learning methods on iris datasets.

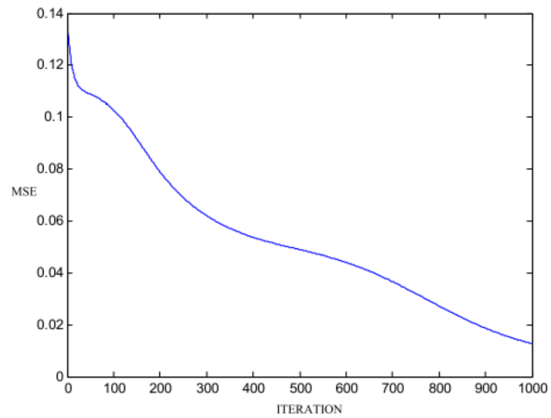


Fig.4: MSE graph of two-term BP with fixed learning method in iris dataset.

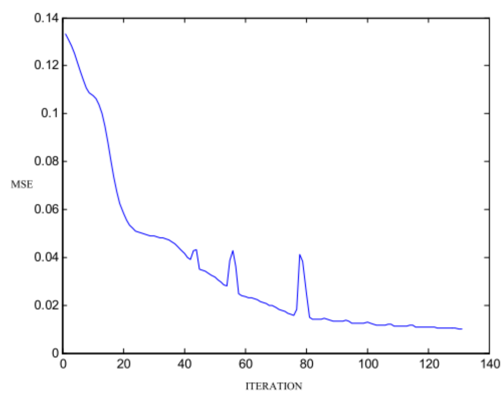


Fig.5:MSE graph of two-term BP with adaptive learning method in iris dataset.

### 4.3.2 Balloon Dataset

On the other hand, for balloon dataset, the convergence speed is recorded at 0.5774s with 34 iterations and an accuracy of 80% for the adaptive learning. While for the fixed learning method, the number of iterations is 1000 with convergence time and accuracy of 1.3423s and 80% (refer to Fig.6 and 7).

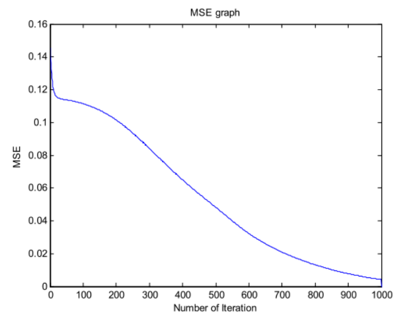


Fig.6: MSE graph of two-term BP with fixed learning in Ballon dataset.

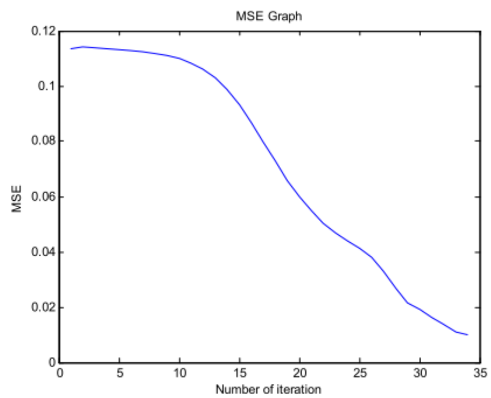


Fig.7:MSE graph of two-term BP with adaptive learning in Ballon dataset.

### 4.3.3 Balance Scale Dataset

For balance scale dataset, the number of iterations is 1000 with convergence time of 1.3242s and accuracy of 90.4255% for the adaptive learning method. For the fixed learning method, the number of iterations is 1000 with convergence time of 1.1905s and accuracy of 86.7021% (refer to Fig.8 and 9).

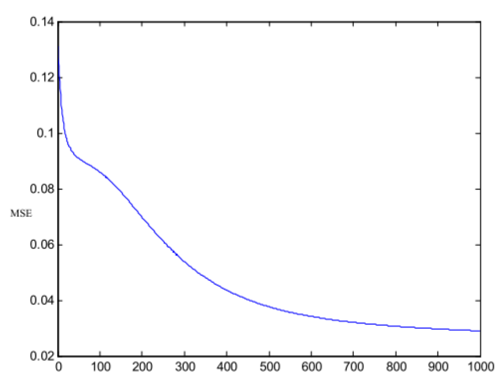


Fig.8:MSE graph of two-term BP with fixed learning method in balance scale dataset.

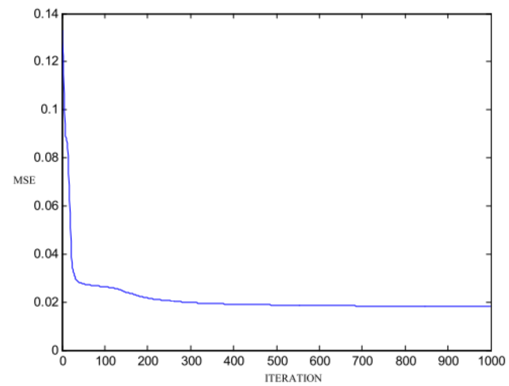


Fig.9: MSE graph of two-term BP with adaptive learning method in balance scale dataset.

# 5 Conclusion and Future Work

In this paper, we implement two methods of weight adjustment for two-term back-propagation. The results illustrate that the two-term BP using an adaptive algorithm works better in producing proper changes of weight with shorter time to converge compared to two-term BP using fixed learning method. As future work, we aim at using multi objective evolutionary algorithm methods and Pareto solutions set to further increase the accuracy of the BP network besides optimizing the network structure and weights connection.

**ACKNOWLEDGEMENTS**

This work is supported by The Universiti Teknologi Malaysia under Flagship Project (Q.J130000.2428.00G96). Authors would like to thanks Research Management Centre (RMC), Universiti Teknologi Malaysia (UTM) for the support in R & D, and Soft Computing Research Group (SCRG) for the inspiration in making this study a success.

# References

[1] Dai-yuan, Z. (2006). New algorithm for training feedforward neural networks with cubic spline weight function. *Journal of Systems Engineering and Electronics, 9*, 1434-1438.

[2] Dhar, V. K., Tickoo, A. K., Koul, R., and Dubey, B. P. (2010). Comparative performance of some popular artificial neural network algorithms on benchmark and function approximation problems. *Pramana - Journal of Physics, 74*(2), 307-324.

[3] Duffner, S., and Garcia, C. (2007). An online backpropagation algorithm with validation error-based adaptive learning rate. *Artificial Neural Networks–ICANN 2007*, 249-258.

[4] Eom, K., Jung, K., and Sirisena, H. (2003). Performance improvement of backpropagation algorithm by automatic activation function gain tuning using fuzzy logic. *Neurocomputing, 50*, 439-460.

[5] Fiszelew, A., Britos, P., Ochoa, A., Merlino, H., Fernández, E., and García-Martínez, R. (2007). Finding optimal neural network architecture using genetic algorithms. *Research in Computing Science Journal, 27*, 15-24.

[6] Fukuoka, Y., Matsuki, H., Minamitani, H., and Akimasa, I. (1998). A modified back-propagation method to avoid false local minima. *Neural Networks, 11*(6), 1059-1072.

[7] Govindarajan, M., and Chandrasekaran, R. (2007). *Classifier Based Text Mining for Neural Network.*World Acad. Sci. Eng. Technol., 21: 200-200.

[8] Guangjun, S., Jialin, Z., and Zhenlong, S. (2008). The research of dynamic change learning rate strategy in BP neural network and application in network intrusion detection. Proceedings of the 3rd International Conference on Innovative Computing Information and Control, Jun. 18-20, IEEE Xplore Press, Dalian, Liaoning, pp: 513-513.

[9] Hongmei, S., and Gaofeng, Z. (2009). *A new BP algorithm with adaptive momentum for FNNs training.*Proceedings of the WRI Global Congress on Intelligent Systems, (IS' 09), IEEE Computer Society, pp: 16-20.

[10] Hu, Y., He, Q., and Li, D. (2011). *Research on Samples Self-learning of BP Neural Network Based on Clustering*. Proceedings of the International Conference on Intelligent Computation Technology and Automation, Mar. 28-29, IEEE Xplore Press, Shenzhen, Guangdong, pp: 213-216.

[11] Hua Li, C., and Xiangji Huang, J. (2012). Spam filtering using semantic similarity approach and adaptive BPNN. *Neurocomputing,*92: 88-97.

[12] Idris, N., Yusof, N., and Saad, P. (2009). Adaptive course sequencing for personalization of learning path using neural network. *Int. J. Advance. Soft Comput. Appl, 1*(1), 49-61.

[13] Iranmanesh, S., and Mahdavi, M. A. (2009). A differential adaptive learning rate method for back-propagation neural networks. *World Academy of Science, Engineering and Technology, 38*, 289-292.

[14] Kathirvalavakumar, T., and Subavathi, S. J. (2012). *Modified backpropagation algorithm with adaptive learning rate based on*

*differential errors and differential functional constraints*, Proceedings of the International Conference on Pattern Recognition, Informatics and Medical Engineering, Mar. 21-23, IEEE Xplore Press, Salem, Tamilnadu, pp: 61-67.

[15] Li, Y., Fu, Y., Li, H., and Zhang, S.-W. (2009). *The improved training algorithm of back propagation neural network with selfadaptive learning rate.*Proceedings of the International Conference on Computational Intelligence and Natural Computing, Jun 6-7, Wuhan, China, IEEE Computer Society, pp: 73-76.

[16] Magoulas, G. D., Plagianakos, V. P., and Vrahatis, M. N. (2002). Globally convergent algorithms with local learning rates. *Neural Networks, IEEE Transactions on, 13*(3), 774-779.

[17] McLachlan, G. J., Do, K. A., and Ambroise, C. (2004). *Analyzing microarray gene expression data* (Vol. 422): John Wiley & Sons.

[18] Minai, A. A., and Williams, R. D. (1990). *Back-propagation heuristics: A study of the extended Delta-Bar-Delta algorithm.*Proceedings of the International Joint Conference on Neural Networks, Jun 17-21, IEEE Xplore Press, San Diego, CA, USA, pp: 595-600.

[19] Ng, S., Leung, S., and Luk, A. (1999). Fast convergent generalized back-propagation algorithm with constant learning rate. *Neural processing letters, 9*(1), 13-23.

[20] Norhamreeza Abdul Hamid, N. A. H., Nazri Mohd Nawi, N. M. N., Rozaida Ghazali, R. G., and Mohd Najib Mohd Salleh, M. N. M. S. (2011). Accelerating Learning Performance of Back Propagation Algorithm by Using Adaptive Gain Together with Adaptive Momentum and Adaptive Learning Rate on Classification Problems. *International Journal of Software Engineering and Its Applications, 5*(4), 31-44.

[21] Riedmiller, M. (1994). Advanced supervised learning in multi-layer perceptrons - from backpropagation to adaptive learning algorithms. *Computer Standards and Interfaces, 16*(3), 265-278.

[22] Shamsuddin, S. M., Ibrahim, A. O., and Ramadhena, C. (2013). *Weight Changes for Learning Mechanisms in Two-Term Back-Propagation Network*.

[23] 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.

[24] Sterlin, P. (2007). *Overfitting prevention with cross-validation.* Master thesis, University Pierre and Marie Curie (Paris VI): Paris, France.

[25] Subavathi, S. J., and Kathirvalavakumar, T. (2011). Adaptive modified backpropagation algorithm based on differential errors. *International Journal of Computer Science, Engineering and Applications (IJCSEA),1 (5), 21-34.*

[26] Trentin, E. (2001). Networks with trainable amplitude of activation functions. *Neural Networks, 14*(4-5), 471-493.

[27] Wang, C. H., Kao, C. H., and Lee, W. H. (2007). A new interactive model for improving the learning performance of back propagation neural network. *Automation in construction, 16*(6), 745-758.

[28] Xiaoyuan, L., Bin, Q., and Lu, W. (2009). *A new improved BP neural network algorithm.* Paper presented at the 2009 2nd International Conference on Intelligent Computing Technology and Automation, ICICTA 2009, October 10, 2009 - October 11, 2009, Changsha, Hunan, China, 19-22.

[29] Yam, J. Y. F., and Chow, T. W. S. (2000). A weight initialization method for improving training speed in feedforward neural network. *Neurocomputing, 30*(1), 219-232.

[30] Yan, W. (2004). 1, 2 and WANG Shou Jue 2, 3 1 (Department of Computer Science and Engineering, Tongji University, Shanghai 200092) 2 (Institute of Semiconductors and Information Technology, Tongji University, Shanghai 200092) 3 (Laboratory of Artificial Neural Networks, Institute of Semiconductors, Chinese Academy of Sciences, Beijing 100083); A New Algorithm to Improve the Learning Performance of Neural Network Through Result-Feedback [J]. *Journal of Computer Research and Development, 9.*

[31] Yu, C.-C., and Liu, B.-D. (2002). *A backpropagation algorithm with adaptive learning rate and momentum coefficient.*Proceedings of the International Joint Conference on Neural Networks, May 12-17, IEEE Xplore Press, Honolulu, HI, pp: 1218-1223.

[32] Yu, X. H., and Chen, G. A. (1997). Efficient backpropagation learning using optimal learning rate and momentum. *Neural Networks, 10*(3), 517-527.

[33] Zhang, C., Wu, W., Chen, X., and Xiong, Y. (2008). Convergence of BP algorithm for product unit neural networks with exponential weights. *Neurocomputing, 72*(1), 513-520.

[34] Zhang, X. H., Ren, F. J., and Jiang, Y. C. (2012). An Improved BP Algorithm Based on Steepness Factor and Adaptive Learning Rate Adjustment Factor. *Applied Mechanics and Materials, 121*, 705-709.

[35] Zweiri, Y. H., Whidborne, J. F., and Seneviratne, L. D. (2003). A three-term backpropagation algorithm. *Neurocomputing, 50*, 305-318.