Comparisons of MLP Transfer Functions for Different Classification Classes

Iza Sazanita Isa, Normasni Ad Fauzi, Juliana Md Sharif, Rohaiza Baharudin, Mohd Hussaini Abbas.
Universiti Teknologi Mara
Pulau Pinang, Malaysia
Izasazanita@ppinang.uitm.edu.my

Abstract—**This paper presents a comparison study of two different MLP transfer functions for three different classification cases of breast cancer, thyroid disease and weather classification. The transfer functions under investigation are sigmoid and hyperbolic tangent. In the study, MLP network was trained and tested to investigate the ability of the network to classify the breast cancer correctly between benign cell and malignant cell, classifying thyroid disease into normal, hyper or hypo thyroid and classifying weather conditions into four types; rain, cloudy, dry day and storm. Levenberg-Marquardt algorithm is adopted to train MLP network since it is the fastest training and ensure the best converges towards a minimum error. The performance of MLP networks was evaluated in terms of percentages for correct classification of the target outputs. Both functions are able to give accuracies up to 99% for classifying correctly. The hyperbolic tangent function had shown the capability of achieving the highest accuracy of an MLP performance with less number of hidden nodes.**

Keywords—**Levenberg-Marquardt, Multilayer Perceptron**

I. INTRODUCTION

Even though Multilayer perceptron network (MLP) has been widely used in many applications due to its robust capability and simple structure design, in some cases, it fails to provide a good solution. The reason behind this event is due to improper selection of architecture, initialization of weights or data selection. According to F. Piekniewski and L. Rybicki, another factor that affects the training process is the selection of transfer function [1]. It can be seen that most neural networks are based on either sigmoidal or Gaussian transfer functions [2]. The transfer function in MLP network is typically non-linear function. It functions to transform the weighted sum of inputs to an output value. Some of the most commonly used functions are to solve non-linear problems [3]. An activation function or a transfer function for the hidden nodes in MLP is needed to introduce nonlinearity into the network. The selection of transfer function might significantly affect the performance of a training algorithm. Many researchers have investigated to find special transfer function so that the network structure is kept simple and convergence time can be accelerating [4]. The transfer function of MLP network should have several important characteristics which are it should be continuous, differentiable and monotonically non-decreasing [5]. Research on effect of transfer function for neural network has received a lot of attention in the past literatures. In the paper by Kuan-Chieh Huang and Yau-Hwang Kuo [24] proposed a novel objective function to optimize neural networks for emotion recognition from speech patterns. The sigmoid and Gaussian transfer functions are adopted in their study to construct suitable classification boundaries. The experimental results show that the proposed model has better performance than conventional MLPs. A study conducted by S. W. Lee and C. Moraga [6] propose the function of Cosine-Modulated Gaussian for Hyper-Hill neural networks. The study has compared the Cosine-Modulated Gaussian, hyperbolic tangent, sigmoid and symsigmoid function in cascade correlation network to solve sonar benchmark problem. Research by Joarder and Aziz [7] has proved that logarithmic function is able to accelerate back propagation learning or network convergence. From the study, such classification problems have solved includes XOR problem, character recognition, machine learning database and encoder problem using MLP network with back propagation learning. The paper by Wong *et al* [8] investigated the neuronal function for network convergence and pruning performance. Transfer functions being investigated in the study are Periodic and Monotonic functions for the analyses of multilayer feed forward neural networks trained by Extended Kalman Filter (EKF) algorithm. The problems solved are multicluster classification and identification problem of XOR logic function, parity generation, handwritten digit recognition, piecewise linear function approximation and sunspot series prediction. Study by Piekniewski and Tybicki [1] employed different transfer functions in MLP networks to determine the visual comparison

978-1-4673-3143-2/12/$31.00 ©2012 IEEE

performance. It is shown that log-exponential function has been slowly accelerated. However, it improved effectively in MLP network with back propagation learning. Barycentric plotting was used as a simple projection scheme to measure the neural network performance. Some researchers have also proposed special transfer function such as logarithmic [9], Adaptive Spline [10], Type-2 Fuzzy [4], Elementary Transcendental [11], Hermite Polynomial [12], periodic and monotonic [8], Novel Adaptive [13] and Cosine-modulated Gaussian transfer function [6]. However, the most recommended transfer function among neural network researchers for MLP network is sigmoid or hyperbolic tangent [1, 5].

The objective of this study is to measure performances of MLP networks on classifying the correct classification using MLP networks and Levenberg-Marquardt (LM) as the training algorithm.

II. METHODOLOGY

The system implemented consist several parameters of meteorological data such as date information combined with solar radiation, ambient temperature, current, surface temperature, voltage, wind direction and wind speed. The recorded data were applied to train, validate and test with BP algorithm in MLP network. The breast cancer data has two output classes. The MLP network was trained and tested to investigate its network ability to classify between benign and malignant cells. Another data used is thyroid disease data from James Cook University [12] with output target normal (1), hyper (2) and hypo (3). Three classes of thyroid divisions came from five attributes with continuous value. The LM training algorithm is adopted for updating each connection weights of units and two transfer functions were applied in the learning process are sigmoid and hyperbolic tangent. In this study, LM algorithm has been used due to the reason that the training process converges quickly as the solution is approached.

A. Multilayer Perceptron Network

The MLP consists of input layer, one or more hidden layers and an output layer. The number of hidden layers can be changed depends on data under training process. The output nodes can also be changed depends on classification of target output. The common training procedure for MLP is supervised learning specifically the back propagation algorithm. The algorithm employs the method of gradient descent which tends to minimize the mean squared error between the output of MLP network and the desired output.

The output of network, *y (t)* at output layer *m*, is shown in Equation (1). A common transfer function in the nodes of the hidden or output layer is the sigmoid function. However, other functions such as hyperbolic tangent function or quadratic function can also be employed. The transfer function can be the same for all the nodes or a different function can be employed for each of the nodes. In some applications, no transfer function is employed at the output node.

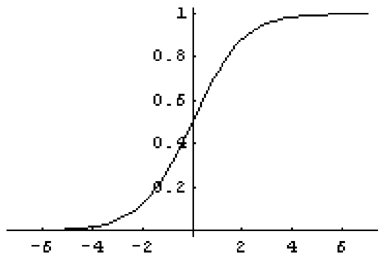

Figure 1. Sigmoid function

$$y_l(t) = h\left(\sum_{j=1}^{n_{m-1}} w_{jl}^{m} v_j^{m-1}(t)\right) \quad ; \quad 1 \le l \le n_m \qquad (1)$$

n_k : number of nodes in the *k*-th layer

n_m : number of nodes in the output layer

w's : weights

h(●): transfer function

The objective of the network is to train the MLP to achieve a balance between the ability to respond correctly to the input patterns that are used for training and the ability to provide good response to the input that is similar. Back propagation learning is one of the most important types of learning in feed forward network. It is a systematic method for training a multilayer network such as MLP.

B. Levenberg-Marquardt Training Algorithm

Training algorithm applied for this study is Levenberg-Marquardt (LM). This algorithm is typically the fastest of training algorithms [14]. The LM is basically a Hessian based algorithm for nonlinear least squares optimization. Hessian-based algorithms are used to allow ANNs to learn more suitable features of a complicated mapping [15]. The training process converges quickly as the solution is approached. The good aspect of Levenberg-Marquardt (LM) is that the determination of the new points is actually a compromise between steps in the direction of the steepest descent.

C. Transfer Functions

The transfer functions are typically a non-linear function that transforms the weighted sum of the inputs to an output value. Sometimes different transfer functions [1, 3-8, 11-23] are acquired for different networks so that it resulting in better performances. Transfer functions for the hidden nodes in MLP are needed to introduce nonlinearity into the network. Linear functions are called "linear" because they are precisely the functions which graph in the Cartesian coordinate plane in a straight line. Non linear problems are difficult to solve and are much less understandable than linear problems. The linear function is much easier as compared to non linear, the output of linear function is easier to predict. Different from perceptron, which do not have any hidden nodes, but only input and output nodes, do not have nonlinearity thus making its net less powerful. The choice of transfer function is important for performance of training algorithm. A transfer function for a backpropagation network should have several important characteristics. It should be continuous, differentiable and monotonically non-decreasing [10]. By using backpropagation learning algorithm, the transfer function used must be differentiable so that the function is bounded in certain ranges of limits.

The initial transfer function is applied to each layer of the network. Most recommended transfer function networks are sigmoid or hyperbolic tangent which commonly used transfer function used in MLP is sigmoid transfer function and it's commonly giving good results. Figure 1shows sigmoid function saturates to 0 or 1, which are the values used to indicate membership in an output class. The input for this function is varied from $-\infty$ to $+\infty$ but usually bounded to certain value. The expression of this function is given by equation (2):

$$f_{sig}(net_i) = \frac{1}{1 + e^{-S_s . net_i}} \qquad (2)$$

net_i is the net input to the *i*th neuron

S_s is the slope of the sigmoid function

Another most common transfer function used in backpropagation learning is hyperbolic tangent transfer function. Hyperbolic tangent transfer function is similar to sigmoid transfer function in transforming the net input to saturate output class between -1 and +1. Figure 2 presents the hyperbolic tangent function and expression of this function is:

$$f_{th}(net_i) = \frac{e^{net} - e^{-net}}{e^{net} + e^{-net}} \qquad (3)$$

Where net_i is the net input to the *i*th neuron.

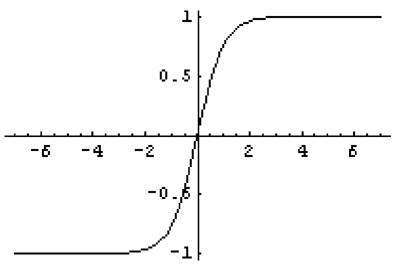

Figure 2. Hyperbolic Tangent function

The purpose of using neural network is to be able to classify the data that are too complex for the traditional statistical models. This project has chosen MLP network as the classifier. MLP network are feed forward neural networks with one or more hidden layers. This means that one hidden layer MLP is almost always sufficient to approximate any continuous function up to certain accuracy [13]. The advantages of MLP are, their abilities to learn and give the better performance especially in the case of classification are proven. In additions the construction of MLP is simple. The ability of an MLP network to classify data efficiently and make decisions based on the classification results is one of the distinguishing features that resemble human intelligence. The MLP network has to be train before it able to perform specific task with less error. Theoretically [14], a single MLP network with the best performance at the fewest number of hidden neurons will be selected as the best ANN to represent a problem. The architecture of the MLP consists of input layer, hidden layer and output layer. Number of the input layer depends to the number of attributes weather classification data (7 attributes). The number of hidden neurons decided upon training stage of the MLP networks. Four output neurons are needed to classify the class of the target output. The performances of the MLP networks will be evaluated in terms of percentages for correct classification, defined as the different between the actual and the simulated results. The performance of correct classification is defined as in Equation 4.

$$\% = \frac{Correct\ \ classification}{Actual\ \ class}\ x\ 100\% \qquad (4)$$

The four different categories defined for the ANN network of weather forecasting are raining, cloudy, dry day and storm denoted by binary number at network target output. Table 2 shows the inputs for MLP network based on the meteorological data distribution. This work uses 4109 training, 4444 for validation and 3538 data for testing with 7 attributes as the input.

TABLE I. PARAMETERS FOR METEOROLOGICAL DATA SET

Input	**Parameters**	X_{min}	X_{max}
x_1	Ambient temperature	23.5^{0}C	34.9^{0}C
x_2	Current	0A	27A
x_3	Solar irradiation	0Wm^{-2}	1350Wm^{-2}
x_4	Surface temperature	28.2^{0}C	37.1^{0}C
x_5	Voltage	0.2V	26.2V
x_6	Wind direction	0^0	360^0
x_7	Wind speed	0m/s	7 m/s

The breast cancer data uses 246 and 228 dataset of breast cancer disease obtained from the University of Wisconsin Hospital, Madison by Dr. William H. Wolberg for training and testing. This dataset has been used in the research by [15]to analyze the pattern recognition of medical diagnosis. This paper is using the same dataset to perform new analysis of MLP classification comparing two different transfer functions. All the 9 attributes were in the range between 2 and 9. Summary of data division is described in Table 2. The breast cancer data had two output classes. The predicted outputs were continuous values and converted to absolute values representation within the threshold ranges. The other data used was thyroid disease data. The original data thyroid disease came from James Cook University [12] with output target 1, 2 and 3. Three classes of thyroid divisions came from five attributes with continuous values between -0.7 to 141. The data was divided into 215 for training and 200 for testing. The two samples of data were categorized into three classes. The summary of data division is described in Table 2 and Table3. In the study, the networks with bias connection were trained and tested with the following parameters as suggested by [16]:

$$\text{Learning rate, } \eta = 0.3$$
$$\text{Momentum, } \alpha = 0.4$$
$$\text{Gain scale, } \lambda = 1.0$$
$$\text{Tolerance, } \tau = 0.1$$

TABLE II. DATA DIVISION OF BREAST CANCER DISEASE

	Training Data	Testing Data
Normal Cells	123	128
Cancer Cells	123	100
Total Data	246	228

TABLE III. DATA DIVISION OF THYROID DISEASE

	Training Data	Testing Data
Normal Thyroid	150	150
Hyper Thyroid	35	25
Hypo Thyroid	30	25
Total Data	215	200

Using the chosen initial weights, MLP networks were trained and tested until the networks were converged. Hidden node that gives the best performance for training and testing was chosen as the best node for the network of both transfer functions. This method was implemented for training and testing thyroid data set, breast cancer data set and meteorological data set. The performances of each MLP network will be evaluated in terms of percentages for correct classification, defined as the different between the actual and the simulated results.

III. RESULTS AND DISCUSSION

The main objective of this study was to compare the performance of an MLP network by changing two different transfer functions to the network. The MLP networks have been trained by using LM and the performance of the proposed system has been assessed based on the percentage of correct classification. The excellent networks, i.e. ones that give the highest percentage of correct classification during learning process will be selected as the chosen MLP network for the system. Each data has been divided into three sets with equal distinct classes for each set. This is done to make fair comparison of each network and moreover the network will learn with no bias.

The network has been trained until 20 hidden neurons. For each hidden neuron number, the MLP has been simulated 20 times to ensure that global minimum is reached. From the results, MLP network trained for weather classification using both transfer functions gives the highest percent of correct classification during testing phase. The performance of the network during training phase shows that the percent of correct classification for training using hyperbolic tangent transfer functions is achievable to 99.8%. Even though the performances of each network using both functions are almost the same, less number of hidden nodes is preferable for MLP.

It can be seen that MLP network using sigmoid function with 3 hidden nodes was able to achieve the highest testing data accuracy which was 99.3%. When MLP network is implemented the hyperbolic tangent function, the highest testing data accuracy was achievable to 99.4% at hidden nodes 5.

The performance of the network during training phase shows that the percent of correct classification for training using both functions are acceptable which achieved more than 99%. However hyperbolic tangent function gives a better performance compared to sigmoid function. In terms of number of hidden nodes, sigmoid is presentable than hyperbolic tangent. The results for both types of transfer functions are summarized in Table 4 and Table 5.

TABLE IV. MLP PERFORMANCE OF DIFFERENT CLASSIFICATION CASES FOR SIGMOID FUNCTION

Case study	No hidden neurons	Training accuracy (%)	Testing accuracy (%)
Breast cancer (two classes output)	7	97.6	95.6
Thyroid diseases (three classes output)	10	89.3	84
Weather classification (four classes output)	4	99.8	99.4

TABLE V. MLP PERFORMANCE OF DIFFERENT CLASSIFICATION CASES FOR HYPERBOLIC TANGENT FUNCTION

Case study	No hidden neurons	Training accuracy (%)	Testing accuracy (%)
Breast cancer (two classes output)	5	97.2	97
Thyroid diseases (three classes output)	20	95.8	92
Weather classification (four classes output)	3	99.8	99.3

IV. CONCLUSION

This paper investigated the network performance by using MLP networks with two different transfer functions; sigmoid and hyperbolic tangent function. These functions had been investigated in MLP network to determine the most suitable function to solve different class of classification problems. The transfer functions under investigation were sigmoid and hyperbolic tangent. Three samples of data set were used to solve classification problems. Both transfer functions were able to give accuracies of up to 99%. However hyperbolic tangent function had shown the capability of achieving the highest accuracy of an MLP performance with less number of hidden nodes. From the findings, it can be concluded that the hyperbolic tangent function is suitable for MLP network to classify data of four classes. The other transfer functions can be employed in the future to obtain higher classification accuracy. However, the success of the system in other classification tasks could further be proving with new transfer functions.

REFERENCES

[1] "Visual comparison of performance for different activation functions in MLP networks," 2004, pp. 2947-2952 vol. 4.

[2] W. Duch and N. Jankowski, "Transfer functions: hidden possibilities for better neural networks," 2001, pp. 81-94.

[3] B. Karlik and A. Vehbi, "Performance analysis of various activation functions in generalized mlp architectures of neural networks," International Journal of Artificial Intelligence and Expert Systems (IJAE), vol. 1, p. 111, 2011.

[4] M. Karakose and E. Akin, "Type-2 fuzzy activation function for multilayer feedforward neural networks," 2004, pp. 3762-3767 vol. 4.

[5] M. Rosen-Zvi, M. Biehl, and I. Kanter, "Learnability of periodic activation functions: General results," Physical Review E, vol. 58, p. 3606, 1998.

[6] S. W. Lee and C. Moraga, "A Cosine-Modulated Gaussian activation function for Hyper-Hill neural networks," 1996, pp. 1397-1400 vol. 2.

[7] J. Kamruzzaman and S. M. Aziz, "A note on activation function in multilayer feedforward learning," 2002, pp. 519-523.

[8] K. W. Wong, C. S. Leung, and S. J. Chang, "Use of periodic and monotonic activation functions in multilayer feedforward neural networks trained by extended Kalman filter algorithm," 2002, pp. 217-224.

[9] J. Bilski, "The Backpropagation learning with logarithmic transfer function," Proceeding of 5th Conf. On Neural Networks and Soft Computing, pp. 71-76.

[10] P. Campolucci, F. Capperelli, S. Guarnieri, F. Piazza, and A. Uncini, "Neural networks with adaptive spline activation function," 1996, pp. 1442-1445 vol. 3.

[11] T. Kim and T. Adali, "Complex backpropagation neural network using elementary transcendental activation functions," 2001, pp. 1281-1284 vol. 2.

[12] L. Ma and K. Khorasani, "Constructive feedforward neural networks using Hermite polynomial activation functions," Neural Networks, IEEE Transactions on, vol. 16, pp. 821-833, 2005.

[13] S. Xu and M. Zhang, "A novel adaptive activation function," 2001, pp. 2779-2782 vol. 4.

[14] A. A. Suratgar, M. B. Tavakoli, and A. Hoseinabadi, "Modified Levenberg-Marquardt method for neural networks training," 2005, pp. 46-48.

[15] X. Ying, "Role of activation function on hidden units for sample recording in three-layer neural networks," 1990, pp. 69-74 vol. 3.

[16] G. Hoffmann, "Adaptive transfer functions in radial basis function (rbf) networks," Computational Science-ICCS 2004, pp. 682-686, 2004.

[17] K. C. Huang and Y. H. Kuo, "A novel objective function to optimize neural networks for emotion recognition from speech patterns," 2010, pp. 413-417.

[18] S. Kumar, Neural networks: a classroom approach: Tata McGraw-Hill Education, 2004.

[19] O. Mangasarian, R. Setiono, and W. Wolberg, "Pattern recognition via linear programming: Theory and application to medical diagnosis," Large-scale numerical optimization, pp. 22-31, 1990.

[20] M. Y. Mashor, "Performance comparison between back propagation, RPE and MRPE algorithms for training MLP networks," School of Electrical and Electronic Engineering, Universiti Sains Malaysia, 2002.

[21] E. Shenouda, "A quantitative comparison of different MLP activation functions in classification," Advances in Neural Networks-ISNN 2006, pp. 849-857, 2006.

[22] S. Sivanandam, Introduction to Neural networks using MATLAB 6.0: Tata McGraw-Hill Education, 2006.

[23] I. C. Yeh, C. C. Chen, and K. C. Huang, "Hybrid Transfer Function Networks," 2010, pp. 37-42.

[24] H. Kuan-Chieh and K. Yau-Hwang, "A Novel Objective Function to Optimize Neural Networks for Emotion Recognition from Speech Patterns", Nature and Biologically Inspired Computing (NaBIC), 2010, pp. 413-417.