International Journal of Scientific Research in Computer Science, Engineering and Information Technology
 | Volume 3 | Issue 7 | ISSN : 2456-3307

Comparison of Back Propagation Algorithms and Fusion Methodology Using Dempster-Shafer Rule in Medical Application

B. Sumathi

Department of Computer Science, CMS College of Science and Commerce, Coimbatore, Tamil Nadu, India

ABSTRACT

This paper presents various Back Propagation algorithms for the diagnosis of hypertension. Parameters such as learning rate and momentum coefficients are used to improve the rate of convergence and controls the feedback loop of Back Propagation algorithm. The value for learning rate and momentum factors are varied instead of using fixed value to make the learning more effectively during the training process. The primary classifiers used in this paper are Quasi-Newton (QN), Gradient Descent (GD) and Levenberg-Marquardt (LM) Back Propagation training algorithms, each using different learning function. The Dempster-Shafer's rule has been adopted to combine the output of these three Back Propagation neural networks into single one to enhance the target result. The experimental result shows that the fusion method would provide a significantly higher accuracy for the diagnosis of hypertension.

Keywords : Back Propagation, Dempster-Shafer, Accuracy, Learning Rate, Momentum

I. INTRODUCTION

An Intelligent and accurate diagnostic system is required for better diagnosis. Each patient is diagnosed as either hypertensive or non-hypertensive. The diagnosis of the disease is based on the selected symptoms of the patients. In order to overcome the conventional manual data analysis, an efficient computer based analysis is essential.

In olden days statistical analysis method was used to diagnose, later, the expert system has developed, in which there is an interaction between the doctor and the computer.

Artificial Neural Network techniques in medical applications could result in reducing cost, time, medical error and need of human expertise. In order to improve the classification rate, fusion method is used. Since the main reason for combining classifiers is to enhance the performance of the network. A variety of schemes are proposed for combining multiple classifiers. Recently, fusion methods have been applied in many areas including medical diagnosis, handwriting recognition, character recognition, speech recognition, biometrics, sales forecasting, earth quake evaluation and industrial diagnostics applications. Various techniques such as Dempster-Shafer Theory, Rough Set Theory, and Fuzzy Set Theory etc. are used to create fusion classifiers depends upon the information given by the primary classifiers. Among these techniques Dempster-Shafer Theory can enhance the target performance.

The method used in this study is to perform an investigation on the combination of Neural Network training algorithms with Dempster-Shafer theory for improving the performance of the trained network. Dempster derived the combination rule for multiple independent sources of information. Shafer extended Dempster's theory to define degree of belief, and proposes combination rules in the form of Dempster-Shafer's (DS) combination rule.

This paper is organized as follows: Section II Methods and Material, summarizes the Back Propagation algorithms, fundamentals of Dempster-Shafer Theory, and fusion method to combine the BP neural network and Dempster-Shafer theory, Section III describes the experimental results, Section IV concludes the paper.

II. METHODS AND MATERIAL

Neural Network (NN) techniques are applied on hypertension data sets to identify patterns and relationships. These techniques are useful for predicting the diseases or early diagnosis of the disease. The Back Propagation (BP) is a systematic method for training ANN and uses the method of supervised learning. BP learning consists of the forward pass and a backward pass through different layers of networks. The input values are passed to input layer nodes during forward pass and has been propagated through the network layer by layer. During the forward pass the synaptic weight of the networks are fixed. During the backward pass, the synaptic weights are adjusted with the weight update rule. The actual response of the network is subtracted from a desired target to produce an error signal. The error signal is then propagated backward through the network. The main objective of the learning process is to adjust the synaptic weights and bias value of the network, in order to minimize the error. The weight has been updated on pattern by pattern basis until it reaches the one complete presentation of the entire training set

The learning process takes place by inputting the data to be trained by the network. The information from the input layer is distributed to the hidden layer for information process. Then the output layer processes the information to obtain the results. The predicted outputs are then compared with the desired values for error computations. The error is sent to the learning system for further training until the minimum value of error is generated. Table 1 shows the process of back propagation algorithm .

Table 1. Back Propagation Algorithm

1. Create architecture consists of three layers: input layer, one hidden layer and output layer.
2. Replace the missing data
3. Present the input dataset to the network
4. Normalize the input data.
5. Initialize the network parameters such as learning rate =0.25, Momentum=0.75
6. Initialize weights by random values.
7. Train the network
8. Calculate the error of the network
9. Update the weights.
10. Repeat the above steps from 1 to 9 until the maximum epochs are reached or the desired output is identified or the minimum gradient is reached.

The output of the Back Propagation networks is considered as the primary diagnostic result and fused with the Dempster-Shafer Theory to get the final diagnosis result.

Each of these algorithms produces output as Basic Belief Assignments (BBA) and they are used as the input to the Dempster-Shafer Theory. In order to get the final diagnosis, the evidences are combined using Dempster's

Combination rule. algorithms used to train the network are:

A. Conjugate Gradient Back Propagation Algorithm
B. Quasi-Newton Back Propagation Algorithm
C. Levenberg-Marquardt Back Propagation Algorithm
D. Fusion method

A. Conjugate Gradient Back Propagation Algorithm:

This method is used for large scale optimization problems. In this algorithm, search is performed along conjugate directions to determine the step size, which produces faster convergence. The weight updating is performed in the opposite direction as:

$$P_0 = -g_0$$

A line search is then performed to determine the optimal distance to move along the current search direction:

$$\Delta w_k = \propto_k P_k$$

Where P_k is the scaling factor α_k is the length of the step performed at iteration k. The next search direction is determined so that it is conjugate to previous search directions. The successive direction vectors are computed as a linear combination of the current negative gradient vector and the previous direction vector.

B. Quasi-Newton Back Propagation Algorithm:

This algorithm requires more computation and more storage than the conjugate gradient methods, though it takes less iteration to converge. It is an efficient training algorithm for smaller network. This algorithm calculates the error gradient as the sum of the error gradients on each training case.

The first search direction is the negative of the gradient of performance. In succeeding iterations, the search direction is computed according to the formula:

$$\Delta w_k = -H'_k g_k$$

Where H_k is the Hessian matrix of the performance index at the current values of the weights and g_k is the gradient of the error function. As the second order derivatives of total error function, hessian matrix gives the proper evaluation on the change of gradient vector. When H_k is large, it is complex and time consuming to compute Δw_k.

C. Levenberg-Marquardt Back Propagation Algorithm:

This algorithm provides effectiveness, faster and stable convergence. In the artificial neural network field, this algorithm is suitable for training small and medium sized problems. It is relatively complex algorithm, since not only gradient but also requires computation of the Jacobian matrix at each iteration step and the inverse of the square matrix. When the performance function has the form of a sum of squares then the hessian matrix can be approximated as:

$$H = J^T J$$

Where J is the Jacobian matrix. It requires both forward and backward calculations. The error has been determined during forward calculation. The elements of the Jacobian matrix are obtained by propagating this error backward through the network. The backward calculation is responsible for the complexity of the algorithm. The gradient can be calculated as:

$$g = J^T e$$

Where g is the gradient vector, e is the network error which is calculated from the actual and target output, H is the Hessian matrix and J is the Jacobian matrix that contains first order derivatives of the network errors with respect to the weights and biases. The Jacobian matrix can be calculated through a standard back propagation technique with less expensive than computing the Hessian matrix.

The levenberg-Marquardt algorithm uses this approximation to the Hessian matrix to update the weight as:

$$w_{k+1} = w_k - (J_K^T J_K + \mu I)^{-1} J_K e_K$$

Where I is identity matrix of the same dimensions as H and μ is the combination coefficient (Marquardt factor) and is always positive, which plays a critical role in this algorithm. In order to secure the convergence the parameter μ is automatically adjusted at each step. When the parameter μ is zero, then the algorithm follows the Newton's method.

D. Fusion method:

Combination rules are the special types of aggregation methods in order to combine data obtained from multiple sources. These multiple sources provide different assessments. Dempster-Shafer Theory is based on these independent sources. The block diagram of a fusion system based on Back Propagation (BP) algorithms and Dempster-Shafer Theory is as:

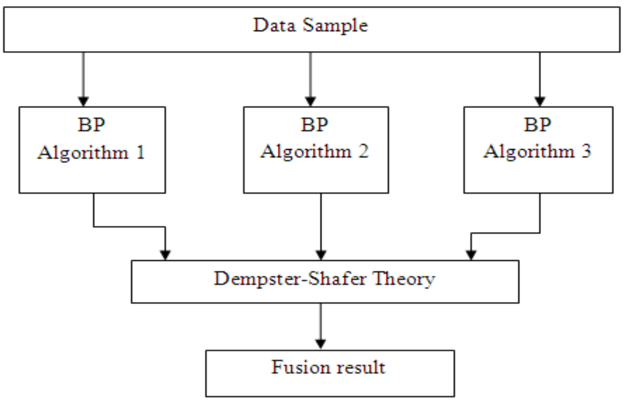

Figure 1. Fusion system

The normalized data from the data sample is taken as the input to the BP algorithm 1, BP algorithm 2 and BP algorithm 3. The outputs of the BP Networks are considered as the new Basic Belief Assignment (BBA) and then these BBAs are fused using Dempster-Shafer Theory to get the final result.

The fusion operators proposed in this research is conjunctive operator. The combination or join is calculated from the aggregation of two BBAs. The degree of conflict between the sources is defined. If the degree of conflict is close to 0 then both BBAs are not in conflict. When the degree of conflict is close to 1 then both BBA are in total conflict. Any value between 0 and 1 indicates partial conflict. The combination rule proposed in this research is Dempster's Rule.

1) Dempster's Rule:

This rule is proposed by Dempster in mathematical theory of evidence, which is used to combine evidences obtained from two or more independent sources. This rule is the agreement between multiple sources and it ignores all the conflicting evidence through a normalization factor. Dempster's rule of combination is a method for fusing belief constraints. It always produces correct results in situation of fusing belief constraints from different sources. Dempster's rule of combination is the appropriate fusion operator. This rule shows that the level of conflict between some sources of information. The degree of conflict is satisfied only with the high belief and low conflicting sources of information. The steps to evaluate Dempster's rule is as follows:

Algorithm

1. Load the data from hypertension dataset
2. Divided the data set into 80% for training, 20% for testing
3. Diagnose the hypertension using Back Propagation Neural Network training algorithms
4. Calculate the Basic Belief Assignment and uncertainty for the primary sources
5. Calculate the conjunctive rule
6. Calculate the Degree of conflict
7. Repeat step 1 and 2 for the current pattern of all n-1 sources

8. The conflicting mass k is redistributed to the sets A, B and AUB proportionally with their masses by calculate the following:
9. $m_{DS}(A) = \frac{1}{1-K} . m_{\wedge}(A)$

   $m_{DS}(B) = \frac{1}{1-K} . m_{\wedge}(B)$

   $m_{DS}(AUB) = \frac{1}{1-K} . m_{\wedge}(AUB)$
10. A Repeat step 1 to 4 for all the patterns
11. Add the value $m_{DS}(A)$ and $m_{DS}(B)$ to get the fusion result

The denominator of the Dempster's rule is the normalization factor. This has the effect of eliminating the conflicting pieces of information between the two sources. if k=1, the two evidences are in completely contradictory, and the Dempster Shafer rule does not apply and It only makes sense if k < 1. If the value of the denominator is non-zero, then the degree of conflict k is less than one. The advantages of Dempster-Shafer's rule are:

- A conjunctive behaviour
- An efficient management of the conflict between sources
- possibility to fuse a large number of sources

III. RESULTS AND DISCUSSION

Back Propagation Neural Network is able to diagnose correctly on the complete data set. In order to identify the best Back Propagation method, the accuracy of the network is evaluated. The objective of this paper is to experiment the impact of fusion algorithm, which combines the group of Back Propagation neural networks and Dempster-Shafer Theory using Dempster's rule to diagnose hypertension. The output of the different algorithms is considered as the primary diagnosis results and fused this result with the Dempster-Shafer Theory to get the final diagnosis results. Fusion method is used to enhance the target performance and reduces the uncertainty level. The real time data sets are chosen for the experiment. The experimental results are evaluated and compared the result of Back propagation algorithms with fusion method. Back Propagation algorithms namely Conjugate Gradient (CG), Levenberg-Marquardt (LM) and Quasi-Newton (QN) are trained. The algorithm is implemented in MATLAB since it increases the flexibility when computing large number of records. Network parameters such as network size and architecture of nodes, values of initial weights, learning rate and momentum were kept same, in order to compare the performance of the proposed algorithm. The performances of various Back Propagation methods are shown in figure.

It is observed that Levenberg–Marquardt algorithm used in this study provides faster convergence, consumed considerably lower time and better estimation results than other training algorithms. But Quasi-Newton method consumed a lot of time, since it required plenty of initial training. The Levenberg-Marquardt algorithm needs more iteration to converge than the Quasi-Newton algorithm. The comparison result shows that the Levenberg-Marquardt algorithm performs highest accuracy of 96.4%.

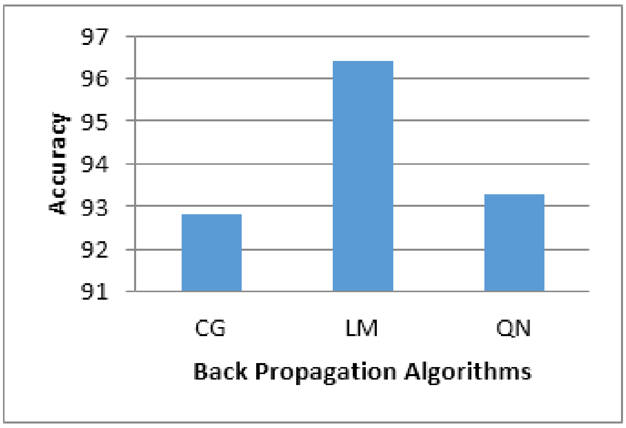

Figure 2. Comparison of Back Propagation Algorithms

The Dempster Shafer fusion algorithm is used to combine the different Back Propagation Neural Networks. The accuracy of each of network provides beliefs. The beliefs and uncertainty values are calculated from the result of three networks. The results of combination of three sources are evaluated by conjunctive rule as beliefs. Then these beliefs are combined to receive final diagnosis result using Dempster's rule. The diagnosed result of BP Algorithms is considered as three sources. The result of positive and negative to hypertension are calculated.

Table 2. Comparison of Back Propagation algorithm with Dempster-Shafer (DS) rule

Methods	Miss Diagnosis Error Rate (%)	Diagnosis Accuracy (%)
NN1	15.66	84.34
NN2	10.5	89.5
NN3	7.88	92.12
DS Rule	1.99	98.01

The Artificial Neural Networks are used to diagnose the correctly and wrongly diagnosed cases as shown in Table-2. Correct diagnosis result produced by DS Rule as 98.01 % on data set. Therefore, result produced by DS rule is the best method.

IV. CONCLUSION

In this paper, feed-forward back propagation neural network algorithms are proposed to diagnose the disease. The overall comparison shows that the Levenberg-Marquardt Back Propagation Neural Network model proves better than the other network models, by performing highest accuracy of 96.4%. In order to increase the performance of the network, different results of various algorithms are combined to produce the final result. The comparisons were made between these algorithms and the fusion method based on Dempster-Shafer Theory produces good result. The proposed algorithm is an efficient method to produce an accurate result. This algorithm was successfully tested on young and older and healthy and unhealthy people. The experimental result shows that the fusion method using Dempster's Rule produces the highest accuracy as 98.01 % and is the best method for the diagnosis of hypertension.

V. REFERENCES

[1]. Zarth. A, Ludermir. T.B, Optimization of Neural Networks Weights and Architecture: A Multimodal Methodology, IEEE International Conference on Intelligent Systems Design and Applications, 2009, pp. 209-214.

[2]. Bengio. Y, Mori Renato, Flammia. G Kompe. R, Global Optimization of A Neural Network-Hidden Markov Model Hybrid, IEEE Transaction on Neural Networks, 1992, Vol. 3, No. 2, pp. 252-259.

[3]. Behera L, Kumar S, Patnaik A, (2006) "On Adaptive Learning Rate that Guarantees Convergence in Feed Forward Networks," IEEE Transaction on Neural Networks, Vol. 17, pp. 1116-1125.

[4]. Norhamreeza Abdul Hamid, Nazri Mohd Nawi, Rozaida Ghazali, Mohd Najib Mohd Salleh, " 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, 2011, Vol. 5 No. 4, pp. 31-44.

[5]. Rehman M. Z., Nawi N. M., "Improving the Accuracy of Gradient Descent Back Propagation Algorithm (GDAM) on Classification Problems", International Journal on New Computer Architectures and Their Applications, 2011, pp. 861-870.

[6]. Yu L., Wang S., and Lai K., "An Adaptive BP Algorithm with Optimal Learning Rates and

Directional Error Correction for Foreign Exchange Market Trend Prediction", Advances in Neural Networks - Springer, Berlin Heidelberg, 2006, pp. 498-503.
[7]. Yu X.H, Chen G.A, Cheng S.X, (1993) "Acceleration of Back propagation of Learning Using Optimized Learning Rate and Momentum," Electron. Lett, 29(14) pp. 1288-1289.
[8]. Minai, A. A., and Williams, R. D., Acceleration of Back-Propagation Through Learning Rate and Momentum Adaptation, in International Joint Conference on Neural Networks, IEEE, pp 676-679, 1990.
[9]. Ampazis, N., and Perantonis, S. J., "Levenberg-Marquardt Algorithm with Adaptive Momentum for the Effiecient Training of Feedforward Networks, in International Joint Conference on Neural Networks, IEEE, 2000.
[10]. Swanston, D. J., Bishop, J. M., and Mitchell, R. J., "Simple Adaptive momentum: new algorithm for training multiplayer perceptrons", Electronics Letters, Vol. 30, pp 1498-1500, 1994.
[11]. Qui, G., Varley, M. R., and Terrell, T. J., Accelerated Training of Backpropagation Networks by Using Adaptive Momentum Step, IEE Electronics Letters, Vol. 28, No. 4, pp 377-379, 1992.
[12]. Wilson, D. Randall, and Tony R. Martinez, The Inefficiency of Batch Training for Large Training Sets, In Proceedings of the International Joint Conference on Neural Networks (IJCNN2000), Vol. II, pp. 113-117, July 2000.
[13]. Rui Quan, Shuhai Quan, Lang Huang, Changjun Xie and Qihong Chen, "Information fusion in Fault Diagnosis for Automotive Fuel Cell System Based on D-S Evidence Theory," Journal of Computational Information Systems, 2011, pp. 97-105.
[14]. Devi Parikh and Robi Polikar, "An Ensemble-Based Incremental Learning Approach to Data Fusion", IEEE Transactions on Systems, Man and Cybernetics-Part B: CYBERNETICS, 2007, Vol.37, pp. 437-450.
[15]. B. Sumathi etal., Input Feature Selection using Neural Network for the Classification of Hypertension, National Conference on Scientific Computing and Applied Mathematics NCSCAM-2011 at V.L.B. Janakiammal College of Engineering and Technology.
[16]. B. Sumathi etal., Neural Networks Evidence Combination for the Diagnosis of Hypertension, National Conference on Emerging Trends in Information and Computer Technology NCETICT-2013 at Kingston Engineering College.
[17]. B. Sumathi, Increasing the Rate of Convergence of Back Propagation Algorithm Using Dempster-Shafer Theory for the Diagnosis of Hypertension, Dr. N.G.P. College of Arts and Science, February 2016.