International Journal of Scientific & Engineering Research, Volume 6, Issue 3, March-2015
ISSN 2229-5518

Comparison between Resilient and Standard Back Propagation Algorithms Efficiency in Pattern Recognition

Hanaa M. Mushgil [1] Dr. Haithem A. Alani [2] Dr. Loay E. George [3]

Abstract— Pattern recognition systems are systems that automatically identify objects based on features derived from its properties and according this the Neural Network (NN) could be pattern recognition system, so we made this study to compare the performance of the neural network in pattern recognition using learning algorithms: basic Back propagation (BP) with momentum (in both modes pattern and batch) and Resilient BP (Rprop), these algorithms are tested in two different classification tasks, the first one considered to be simple data set and the second one , which is noisy, considered to be difficult data set , the Rprop solves the first problem with less time and number of iterations than basic BP ,although Rprop lessens or avoids some disadvantages of standard BP; but with increasing the problem complexity standard BP (in pattern mode) gives the best results .

Index Terms— Pattern Recognition, Neural Network, Back Propagation, Resilient BP, Pattern mode, Batch mode, Local minimum.

---

1 INTRODUCTION

Classification is one of the most frequently encountered decision making tasks of human activity. A classification problem occurs when an object needs to be assigned into a predefined group or class based on a number of observed attributes related to that object. Many problems in business, science, industry, and medicine can be treated as classification problems [12], since we are living in a world full of data every day people encounter a large amount of information and store or represent it as data, for further analysis and management. One of the vital means in dealing with these data is to classify or group them into a set of categories or clusters. Neural network is one of the intelligence method based decision making and prediction systems where these methods are seemed to be successful to solve difficult and diverse problems by supervised training methods such as back-propagation algorithm [9]. BP training can be done in either a batch or continuous manner. Claims have frequently been made that batch training is faster and/or more "correct" than continuous training. [5], but Randall Wilson in his research show that these claims are untrue and they are often supported by empirical evidence on very small data sets.

Multilayer networks typically use sigmoid transfer functions in the hidden layers. These functions are often called "squashing" functions, because they compress an infinite input range into a finite output range. Sigmoid functions are characterized by the fact that their slopes must approach zero as the input gets large. This causes a problem when you use steepest descent to train a multilayer network with sigmoid functions, because the gradient can have a very small magnitude and, therefore, cause small changes in the weights and biases, even though the weights and biases are far from their optimal values [10]. The Rprop try to lessen the disadvantage of this problem by using adaptively computed parameters which change in every iteration. In fact, these parameters are adjusted during the learning process based on the direction of convergence. This is based on the sign of the respective partial derivative at the current

1 University of Al-Nahrain, College of Science, Computer Science Department, Iraq, E-mai:Hanaa_Musgjil@yahoo.com.
2 University of Al-Nahrain, College of Science, Computer Science Department, Iraq.
3 University of Baghdad, College of Science, Computer Science Department, Iraq.

and the previous epoch [6], so it is an improved adaptation of the batch back propagation algorithm, **and for numerous problems it converges very quickly** [8].

4 BACK PROPAGATION ALGORITHM

The back-propagation training algorithm is an iterative gradient designed to minimize the mean square error between the actual output of multi-layer feed forward perceptron and the desired output [1].

There are two types of BP algorithm in order to learn the ANN which are the batch mode learning algorithm and the incremental mode learning algorithm. In the batch mode, the weights values are modified after all patterns are presented, while in the incremental mode, the weights values are updated at every iteration after input pattern is presented [11].

Training the BP network requires the steps that follow:

Step1. Select the training pair from the training set; apply the input vector to the network input.

Step2. Calculate the output of the network.

Step3. Calculate the error between the network output and the desired output (the target vector from the training pair).

Step4. Adjust the weights of the network in a way that minimizes error.

Step5. Repeat steps 1 through 4 for each vector in the training set until the error for the entire set is acceptably low [4].

For input vector: $x_0, x_1 \ldots x_{N-1}$ and specify the desired output $d_0, d_1 \ldots d_{N-1}$, BP is a recursive algorithm starting at the output nodes and working back to the first hidden layer. Adjust weights by

$$w_{ij}^{(t+1)} = w_{ij}^{t} + \eta\, \delta_j\, x_i \quad \ldots\ldots\ldots\ldots\ldots\ldots (1)$$

In this equation is w_{ij}^{t} the weight from hidden node i, or from an input to node j at time t, x_i is either the output node i or is an input , η is a gain term, and δ_j is an error term for node j , if node j is an output node, then

$$\delta_j = y_j(1-y_j)\,(d_j-y_j) \quad \ldots\ldots\ldots\ldots\ldots\ldots (2)$$

Where d_j is the desired output of node j and y_j is the actual output. If node j is an internal hidden node, then

$$\delta_j = x_j(1-x_j)\Sigma\, \delta_{jm} w_{jk} \quad \ldots\ldots\ldots\ldots\ldots (3)$$

Where k is over all nodes in the layers above node j. [7]

5 Rprop ALGORITHM

The Resilient propagation is a first-order algorithm performing supervised batch learning in multi-layered perceptrons. The basic principle of Rprop is to eliminate the harmful influence of the size of the partial error derivative on the weight step. As a consequence, only the sign of the derivative is considered to indicate the direction of weight update. The size of the weight change is exclusively determined by a weight-specific, so called "update-value" Δ.

$$\Delta w_{ij}^{t} = \begin{cases} -\Delta_{ij}^{t}\ ;\ \text{If} \dfrac{\partial E}{\partial w_{ij}}^{t} > 0 \\ +\Delta_{ij}^{t}\ ;\ \text{If} \dfrac{\partial E}{\partial w_{ij}}^{t} < 0 \\ 0;\ \text{otherwise} \end{cases}$$

Where $\dfrac{\partial E}{\partial w_{ij}}^{t}$ denotes the summed gradient information over the patterns of the pattern set ("batch learning"). The second step of Rprop learning is to determine the new update values Δ_{ij}^{t}. This is based on a sign-dependent adaptation process [4] [6] [10].

$$\Delta_{ij}^{t} = \begin{cases} \eta^{+} \times \Delta_{ij}^{t}\ ;\ If \dfrac{\partial E^{t-1}}{\partial w_{ij}} \times \dfrac{\partial E^{t}}{\partial w_{ij}} > 0 \\ \eta^{-} \times \Delta_{ij}^{t}\ ;\ If \dfrac{\partial E^{t-1}}{\partial w_{ij}} \times \dfrac{\partial E^{t}}{\partial w_{ij}} < 0 \Delta_{ij}^{t}\ ; \\ \Delta_{ij}^{t}\ ;\ \text{otherwise.} \\ \text{Where } 0 < \eta- < 1 < \eta+. \end{cases}$$

6 Flat-Spot Problem

Flat spot problem is one of the main reasons of convergence difficulties [2]. The gradient descent weight changes depend on the gradient of the error surface. Consequently, if the error surface has flat spots, the learning algorithm would take a long time to pass through them. A particular problem with the sigmoid activation functions is that the derivative tends to zero as it saturates (i.e. gets towards 0 or 1) [3].

Many ways have been proposed to deal with the premature saturation problem by adding an offset of 0.1 to the derivative of the sigmoid function. This makes the derivative of the sigmoid prime function never to zero [3].

7 Experimental Work

In this work a simulator has been developed and utilized data sets that have controllable statistical behaviors with various degree of complexity. The used neural network composed of three layers; input layer contain n nodes, where n is the number of class features, hidden layer contained p nodes, where:

$p = n/2 + 1$, and the output layer contains m nodes, where m is the number of classes, and because Rprop training algorithm is susceptible to the Flat Spot Problem for the used activation function , we treated this problem by adding an offset of 0.1 to the derivative of the sigmoid function.

Next the results of two data sets classification, first one the simple data set (data set with stable features values), figure 1 shows the features behavior in class 1 of experiment 1, then the problem complexity is increased in experiment 2 by making some features had unstable behaviors while some other features were overlapped, figure 2 shows the features behavior in class 1 of experiment 2.

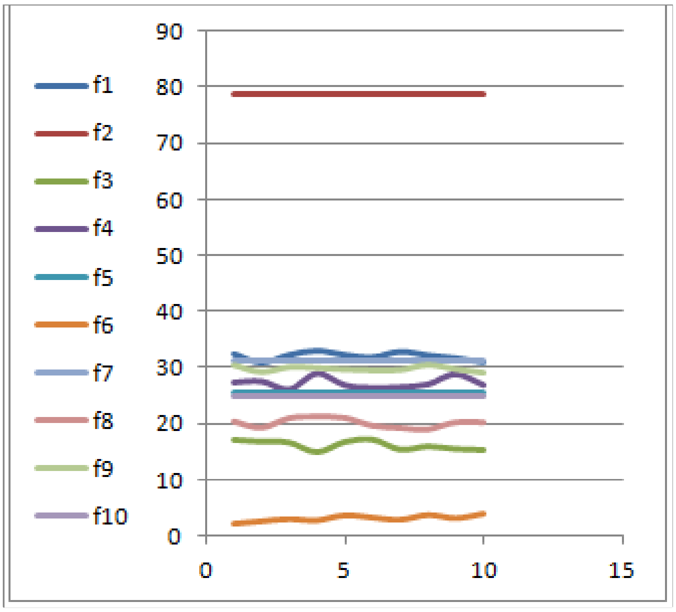

Fig.1, Features behavior in class 1_Experiment1

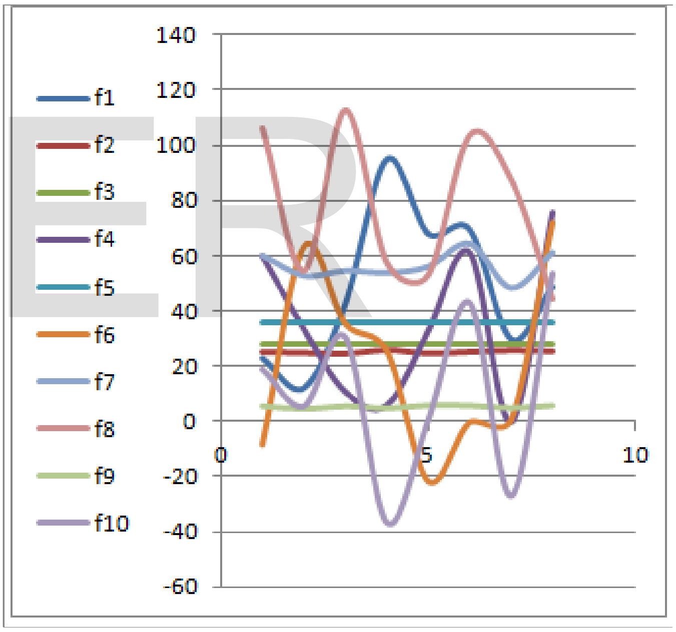

Fig.2, Features behavior in class 1_Experiment2

8 Results

Next the results of two data sets classification, each data set contains 10 classes each class contain 10 samples each sample composed of 10 features, where 30% of the data samples used for testing only (unseen from the network), and 70% of the samples used for training. And then the complete data set tested, the results of the first experiment show the efficiency and fast convergence of the Rprop in avoiding BP problems like local minimum which was met in one of our experiments as in figure (3);

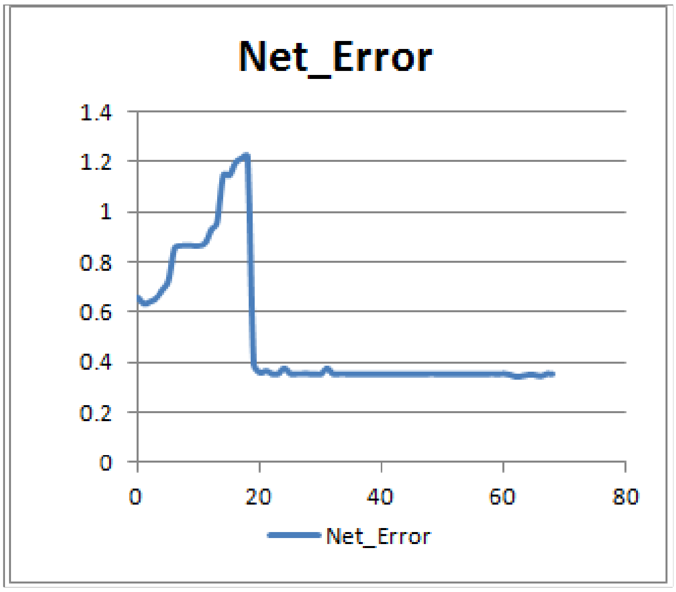

Fig.3, Network error per epoch (Local minimum in BP training)

The results of the first experiment summarized in table 1, and the network error per epoch for each used method is shown in figures (4), (5), (6).

TABLE (1): SIMPLE DATA SET TRAINING RESULTS

Method	Input nodes	Hidden nodes	Output Nodes	Right decisions	Wrong decisions	Time
BP batch mode	10	6	10	93	7	1.6 h
BP pattern mode	10	6	10	100	0	0.13 h
RPROP	10	6	10	100	0	0.21 h

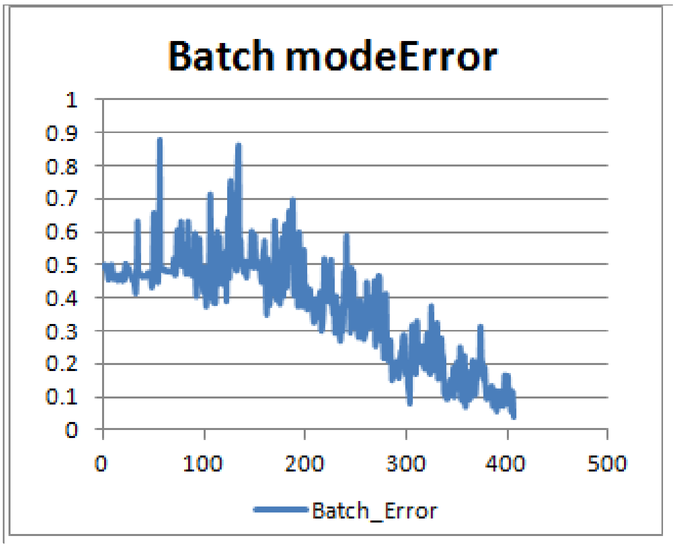

Fig.4, Network error per epoch
(BP Batch mode_ Experiment 1)

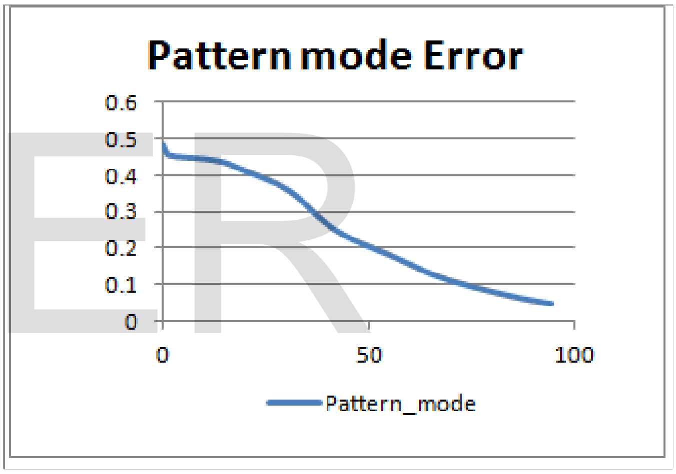

Fig.5, Network error per epoch
(BP Pattern mode_ Experiment 1)

While the results of the second experiment summarized in table (2), and the net work error per epoch for each used method are shown in figures (7), (8), (9).

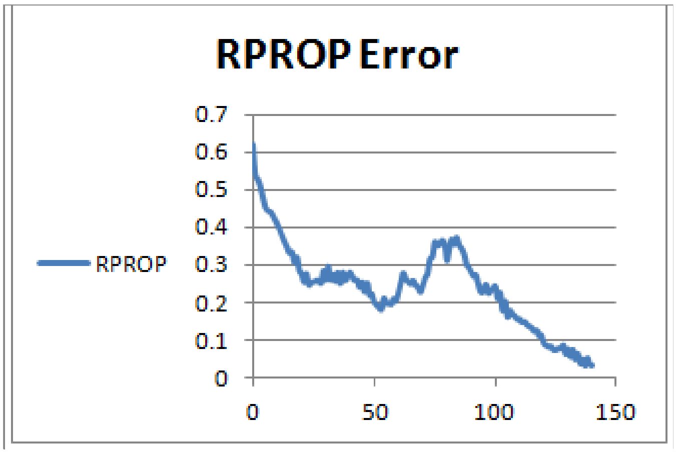

Fig 6, Network error per epoch
(Rprop_ Experiment 1)

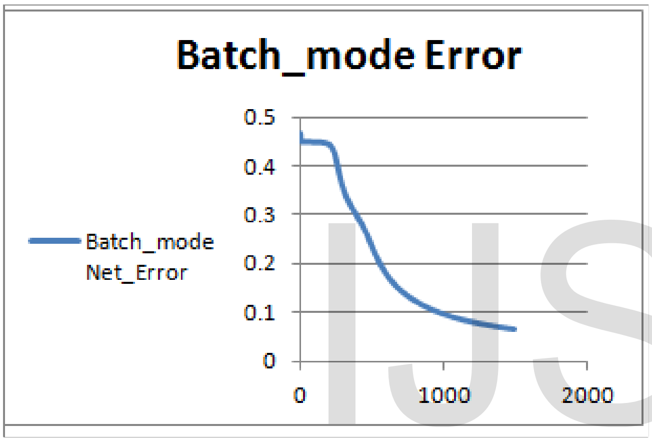

Fig 7, Network error per epoch
(BP batch mode_ Experiment 2)

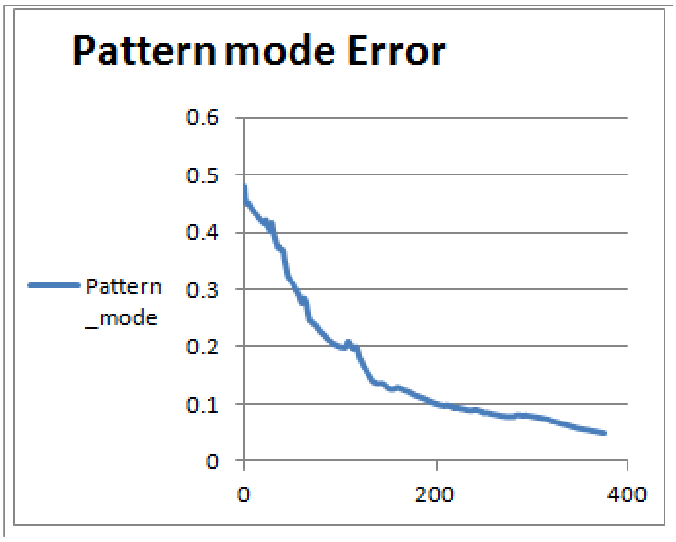

Fig 8, Network error per epoch
(BP pattern mode_ Experiment 2)

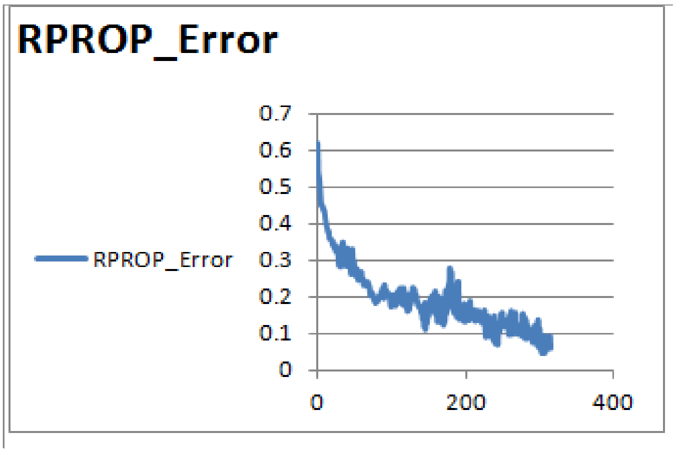

Fig 9, Network error per epoch
(Rprop_ Experiment 2)

TABLE (2): DIFFICULT DATA SET TRAINING RESULTS

Method	Input nodes	Hidden nodes	Output Nodes	Right decisions	Wrong decisions	Time
BP batch mode	10	6	10	86	14	4.3 h
BP pattern mode	10	6	10	90	10	0.28 h
RPROP	10	6	10	82	18	0.47 h

9 CONCLUSIONS

Two different classification problems were used to compare the efficiency of Rprop and standard BP in pattern recognition, although experimental results show that the Rprop algorithm avoids some problems of standard BP algorithms (like local minimum) and with simple classification problem Rprop takes short time comparing with standard BP batch mode; but with increasing the problem complexity in experiment 2, standard BP in pattern mode gives the best results in accuracy and time.

10 REFERENCES

[1] Amit G., Y. P. Kosta, Gaurang P., Chintan G.," Initial Classification Through Back Propagation In a Neural Network Following Optimization Through GA to Evaluate the Fitness of an Algorithm", *International Journal of Computer Science & Information Technology (IJCSIT)*, Vol. 3, No 1, Feb 2011.

[2] Bogdan M. W.," Neural Network Architectures and Learning ", *IEEE*, 2003.

[3] Bo Y. , Ya-Dong W., Xiao-Hong S., Lijuan W. ," Solving Flat-Spot Problem In Back-Propagation Learning Algorithm Based On Magnified Error ", *Proceedings of the Third International Conference on Machine Learning and Cybernetics*, Shanghai, 26-29 August 2004, IEEE, 2004.

[4] Devendra K. C.," Soft Computing Techniques and its Applications in Electrical Engineering", *Springer-Verlag Berlin Heidelberg*, 2008.

[5] D. Randall Wilson Tony R. Martinez ," The Inefficiency of Batch Training for Large Training Sets ",*Proceedings of the International Joint Conference on Neural Networks*, IJCNN,Vol. II, pp. 113-117, July 2000.

[6] Hamed A., Saeid S., Karim M.," Improving the Neural Network Training for Face Recognition using Adaptive Learning Rate, Resilient Back Propagation and Conjugate Gradient Algorithm ", *International Journal of Computer Applications*, (0975 – 8887) Volume 34– No.2, November 2011.

[7] Hanaa Mohammed M., Dr. Loay E. George , Dr. Ban N. Dhannoon,"Agents Technology Based Cooperative Neural Networks for General Pattern Classification Model", *International Journal of Scientific & Engineering Research*, Volume 5, Issue 3, March-2014.

[8] Iftikhar A., M.A Ansari, Sajjad M.,"Performance Comparison between Backpropagation Algorithms Applied to Intrusion Detection in Computer Network Systems ", *9th WSEA International Conference on Neural Networks (NN'08)*, Sofia, Bulgaria, May 2-4, 2008.

[9] Insung J., Gi-Nam W.," Pattern Classification of Back-Propagation Algorithm Using Exclusive Connecting Network ", *International Journal of Electrical and Electronics Engineering*, 2008.

[10] Kritika G., Sandeep K.," Implementation of Resilient Backpropagation & Fuzzy Clustering Based Approach for Finding Fault Prone Modules in Open Source Software Systems ", *International Journal of Research in Engineering and Technology (IJRET)*, Vol. 1 No. 1, 2012.

[11] Norhamreeza Binti Abdul Hamid," The Effect Of Adaptive Parameters On The Performance Of Back Propagation", PhD dissertation, Faculty of Computer Science and Information Technology ,University Tun Hussein Onn Malaysia,2012.

[12] Sandhyahe S.," Neural networks from applied sciences and engineering ", *Taylor & Francis Group, 2007.*