International Conference on Systems, Science, Control, Communication, Engineering and Technology 2016 [ICSSCCET 2016]

ISBN	978-81-929866-6-1	VOL	02
Website	icssccet.org	eMail	icssccet@asdf.res.in
Received	25 – February – 2016	Accepted	10 - March – 2016
Article ID	ICSSCCET130	eAID	ICSSCCET.2016.130

A Survey on Different Training Algorithms for Supervised Learning of Back Propagation Artificial Neural Networks

N Mohana Sundaram[1], V Renupriya[2]

[1,2] Faculty of Computer Science and Engineering, Karpagam Institute of Technology, Coimbatore.

Abstract: *The classical back propagation algorithm is the popular algorithm used for the supervised training of Neural Networks. Further several second order optimization algorithms like the conjugate gradient, the Levenberg-Marquardt and Bayesian learning algorithms have also been developed which can be used along with back propagation algorithm for optimization of learning or accuracy. This paper explores the potential usage of the different algorithms for different applications.*

Keywords: *Neural Networks, Back propagation, Bayesian Regularization, Levenberg-Marquardt, scaled conjugate gradient.*

I. INTRODUCTION

In Artificial Neural Networks the back propagation algorithm is widely used for supervised learning. Back Propagation Algorithm is the training or learning algorithm which is an abbreviation for "backward propagation of errors" is a common method of training the artificial neural networks. Generally in supervised learning the following steps are performed.

1. The network is initialized with random weights. The weight may be small random numbers – say between –1 and +1.
2. The input pattern is applied which are multiplied by the weights added for all the neurons in the layer.
3. The output of the layer is computed by multiplying the activation function. This is called the *forward pass*.
4. The output may be completely different to the Target since all the weights are random. So calculate the *Error* of each neuron, which is essentially: *Target – Actual Output*.
5. This error is then back propagated to update the weights such that the error becomes smaller. In other words, the Output of each neuron will get closer to its Target. This process is called the reverse pass.
6. This process is repeated again and again until the error is minimized and within the prescribed tolerance.
7. Thus the network keeps training all the input patterns repeatedly until the total error falls to some pre-determined low target value and then it stops.

The goal of any supervised learning algorithm is to find a function that best maps a set of inputs to its correct output.

II. Survey

The potential of using a neural network model for estimating the lines of code was examined by K.K. Aggarwal, et al [1]. They employed multiple training algorithms listed in table 1 [1]

Cite this article as: N Mohana Sundaram, V Renupriya. "A Survey on Different Training Algorithms for Supervised Learning of Back Propagation Artificial Neural Networks". *International Conference on Systems, Science, Control, Communication, Engineering and Technology 2016*: 643-645. Print.

trg. Fnc.	Description
trainb	trains a network with weight and bias learning rules with batch updates. The weights and biases are updated at the end of an entire pass through the input data.
Trainbfg	updates weight and bias values according to the BFGS quasi-Newton method.
Trainbr	updates the weight and bias values according to Levenberg-Marquardt optimization. It minimizes a combination of squared errors and weights and then determines the correct combination so as to produce a network that generalizes well. The process is called Bayesian regularization.
Trainc	trains a network with weight and bias learning rules with incremental updates after each presentation of an input. Inputs are presented in cyclic order.
train cgb	updates weight and bias values according to the conjugate gradient backpropagation with Powell-Beale restarts.
train cgf	updates weight and bias values according to the conjugate gradient backpropagation with Fletcher-Reeves updates.
train cgp	updates weight and bias values according to the conjugate gradient backpropagation with Polak-Ribiere updates.
train gd	updates weight and bias values according to gradient descent.
train gda	updates weight and bias values according to gradient descent with adaptive learning rate.
train gdm	updates weight and bias values according to gradient descent with momentum.
train gdx	updates weight and bias values according to gradient descent momentum and an adaptive learning rate.
train lm	updates weight and bias values according to Levenberg-Marquardt optimization.
train oss	updates weight and bias values according to the one step secant method.
train rp	updates weight and bias values according to the resilient backpropagation algorithm (RPROP).
train scg	updates weight and bias values according to the scaled conjugate gradient method.

Table 1: Different learning algorithms [1].

Mean absolute Percentage Error (MAPE) was chosen as a metric by them [1].

$$MAPE = \left(\sum_{j=1}^{j=n} \left[\frac{Estimate - Actual}{Actual} \right] \right) \div n \times 100$$

They tested the performance of all the algorithms mentioned in Table 1 and confirmed that the network trained by Bayesian Regulation algorithm offers the lowest MAPE. Gradient descent and Resilient Back propagation algorithms are in the next level Ozgur and Erdal Uncuoglu conducted a comparative study on the Levenberg-Marquardt, Conjugate Gradient and Resilient Back propagation algorithms for daily stream flow forecasting on Filyos stream in Turkey and for lateral stress estimation of cohesionless soils [2]. They conclude that LM algorithm takes smaller time for training. Although the training time of RB is more than LM, the test results of RB are better. RB and CG algorithms are found to be the most robust in both cases. But they comment that it is difficult to generalize the best algorithm and the algorithm is to be chosen based on the complexity of the problem, size of data sets and the number of weights and biases in the network [2].

The Marquardt algorithm was tested on several function approximation problems, and it was compared with the conjugate gradient algorithm and with variable learning rate backpropagation by Martin T. Hagan and Mohammad B. Menhaj [3]. They conclude that the Marquardt algorithm is very efficient for training networks which have few hundred weights. The computational requirements are much higher for Marquardt algorithm. It presents a high precision. The authors also found that in many cases the Marquardt algorithm converged when the conjugate gradient and variable learning rate algorithms failed to converge [3].

S. Sapna, Dr. A. Tamilarasi and M. Pravin Kumar have used ANNs for prediction of Diabetes employing the Levenberg-Marquardt, Conjugate Gradient and Variable Learning Rate Back propagation algorithms [4]. They conclude that Levernberg-Marquardt algorithm gives the best performance in the prediction of diabetes compared to any other backpropogation algorithm [4]. They also conclude that Marquardt algorithm converged when the conjugate gradient and variable learning rate algorithms failed to converge [4].

The Scaled Conjugate gradient algorithm was used for classification of baby cry problem by José Orozco, Carlos A. Reyes García [5]. The authors found that the SCG Method avoids a time consuming line-search and thereby makes the algorithm faster than other second order Conjugate Gradient algorithms. Further the classification accuracy of 91.08% was obtained [5]

Cite this article as: N Mohana Sundaram, V Renupriya. "A Survey on Different Training Algorithms for Supervised Learning of Back Propagation Artificial Neural Networks". *International Conference on Systems, Science, Control, Communication, Engineering and Technology 2016*: 643-645. Print.

Further the Scaled Conjugate gradient algorithm was used by Haradhan Chel, Aurpan Majumder and Debashis Nandi [6] for training the neural network for handwritten text and character recognition. The algorithm offers faster training with excellent test efficiency.

Scaled Conjugate gradient algorithm and the Bayesian training algorithm were used by Tshilidzi[7] for training the neural networks for fault identification in cylinders. It was concluded that the Bayesian training algorithm is found to be more accurate. A. Payal et al. performed a comparative analysis Bayesian regularization and Levenberg-Marquardt training algorithm for localization in wireless sensor network [8]. They found that the Bayesian regularization algorithm more effective.

Sheela Tiwari, Ram Naresh and Rameshwar Jha made a comparative study on back propagation algorithms for artificial neural networks for identification of power system [9]. They declared that The Gradient Descent algorithm was very slow in converging and the average time required for training the network using the Levenberg- Marquardt algorithm was the least. The maximum time was required for training the network using the Scaled Conjugate Gradient Descent algorithm. The training algorithm Bayesian Regularization continuously modifies its performance function and hence, takes more time. Further the error computed by the LM trained neural network is the least [9].

Harwinder Kaur α & Dalwinder Singh Salaria used the three training algorithms LM, BPA and BR for effort estimation in software project development [10]. They declared that the Bayesian Regularization gives the best performance even though the training time is more.

III. Conclusion

Several applications which utilized different training algorithms are presented. It is very hard to say which the best algorithm is. Normally from the above discussions it may be conclude that for prediction or regression applications the Levenberg-Marquardt algorithm may be a better choice if the dataset is a smaller one. On the other hand the Bayesian Regularization algorithm is best suited even for large data sets. It may be a slower one but it offers the most accurate results. The Scaled Conjugate gradient algorithm is best suited for the classification applications.

References

1. K.K. Aggarwal, Yogesh Singh, Pravin Chandra and Manimala Puri , “Bayesian Regularization in a Neural Network Modelto Estimate Lines of Code Using Function Points” , Journal of Computer Sciences 1 (4): 505-509, 2005 ISSN 1549-363
2. Ozgur and Erdal Uncuoglu, “Comparison of three back propagation algorithms for two case studies”, Indian Journal of Engineering & material sciences, vol. 12, October 2005, pp 434-442.
3. Martin T. Hagan and Mohammad B. Menhaj, “Training Feedforward Networks with the Marquardt Algorithm”, IEEE Transactions on Neural Networks, Vol. 5, NO. 6, November 1994.
4. S. Sapna, Dr. A. Tamilarasi and M. Pravin Kumar, “Backpropagation Learning Algorithm Based on Levenberg Marquardtalgorithm”.
5. José Orozco, Carlos A. Reyes García, “Detecting Pathologies from Infant Cry Applying Scaled Conjugate Gradient Neural Networks”, ESANN'2003 proceedings - European Symposium on Artificial Neural NetworksBruges (Belgium), 23-25 April 2003,ISBN 2-930307-03-X, pp. 349-354.
6. Haradhan Chel, Aurpan Majumder and Debashis Nandi, “Scaled Conjugate Gradient Algorithm in Neural Network Based Approach for Handwritten Text Recognition”, Trends in Computer Science, Engineering and Information technology, Volume 204 of the series pp 196-210.
7. 7.Tshilidzi, “ Scaled Conjugate Gradient and Bayesian training of neural networks for fault identification in cylinders”, Computers and structures , 79 (2001), 2793-2803
8. Ashish Payal , “Comparative analysis of Bayesian regularization and Levenberg-Marquardt training algorithm for localization in wireless sensor network”, Advanced Communication Technology (ICACT), 2013 15th International Conference on, PyeongChang, 2013, pp. 191-194.
9. Sheela Tiwari, Ram Naresh and Rameshwar Jha . “Comparative study of Backpropagation Algorithms in Neural Network Based Identification of Power System”, International Journal of Computer Science & Information Technology (IJCSIT) Vol 5, No 4, August 2013.
10. Harwinder Kaur & Dalwinder Singh Salaria , “Bayesian Regularization Based Neural Network tool for software estimation”, Global Journal of Computer Science and Technology Neural & Artificial Intelligence, Volume 13 Issue 2 Version 1.0 Year 2013, Online ISSN: 0975-4172 & Print ISSN: 0975-4350

Cite this article as: N Mohana Sundaram, V Renupriya. “A Survey on Different Training Algorithms for Supervised Learning of Back Propagation Artificial Neural Networks”. *International Conference on Systems, Science, Control, Communication, Engineering and Technology 2016*: 643-645. Print.