A Survey on Backpropagation Algorithms for Feedforward Neural Networks

[1]Kuldip Vora , [2]Shruti Yagnik
[1] Mtech Scholar, [2]Lecturer
Department of Computer Engineering, Gujarat Technological University, Gujarat, India
[1]kuldipiscool@gmail.com,[2]shruya@gmail.com

***Abstract* -The Back-propagation (BP) training algorithm is a renowned representative of all iterative gradient descent algorithms used for supervised learning in neural networks. It is designed to minimize the mean square error (MSE) between the actual output of a multilayer feed-forward neural network and the desired output. BP has a great high merit of simplicity on implementation and calculation compared to other mathematically complex techniques. It is its simplicity that over period of time attracts researchers and so that, many improvements and variations of the BP learning algorithm have been reported to beat its limitations of slow convergence rate and convergence to the local minima. It is applied to a wide range of practical problems and has successfully demonstrated its power. This paper summarize the basic BP and gradual improvements over Back propagation technique used for classification in Artificial neural networks(ANN) and comparisons with new methods like genetic algorithms(GA) and showing why it is still effective and has scope to improvements.**

***Keywords :* Artificial neural network(ANN); Backpropagation algorithm(BPA); Mean square error (MSE), Multilayer feedforward neural network(MLFFNN); Classification; Cost Function; Genetic algorithm(GA)**

I. INTRODUCTION

Artificial Neural Networks (ANNs) are biologically inspired methods modeled on the learning processes of human brain. Artificial Neural Networks (ANNs) works by processing information like biological neurons in the brain and consists of small processing units known as Artificial Neurons, which can be trained to perform complex calculations. As we learn how to read, write, understand speech, recognize and distinguish pattern – all by learning from examples. In the same way, ANNs are trained rather than programmed. ANN have been successfully solved many complex real world problem such as predicting future trends based on the huge historical data of an organization. ANN have been successfully implemented in all engineering fields such as biological modeling, decision and control, health and medicine, engineering and manufacturing, marketing, ocean exploration and so on.

A multilayer feed-forward neural network (MLFFNN) consists of an input layer, hidden layer and an output layer of neurons. Every node in a layer is connected to every other node in the neighboring layer. A FFNN has no memory and the output is solely determined by the current input and weights values. A feed forward neural network consists of one or more layers of usually non-linear processing units (can use linear activation functions as well). The output of each layer serves as input to the next layer. The objective of training a NN is to produce desired output when a set of input is applied to the network The training of FNN is mainly undertaken using the back-propagation (BP) based learning. Back-Propagation Neural Network (BPNN) algorithm is the most popular and the oldest supervised learning multilayer feed-forward neural network algorithm proposed by Rumelhart, Hinton and Williams [2]. It is built on high mathematical foundation and has very good application potential such as to pattern recognition, dynamic modeling, sensitivity analysis, and the control of systems over time[23].

BPLAs use the gradient-decent search method to adjust the connection weights. The structure of a back-propagation ANN is shown in Figure 1. A multilayer feedforward network with I input neurons, m neurons in the hidden layer and n output neurons in the output layer is written as 1 - m - n. The output of each neuron is the aggregation of the numbers of neurons of the previous level multiplied by its corresponding weights. The input values are converted into output signals with the calculations of activation functions.

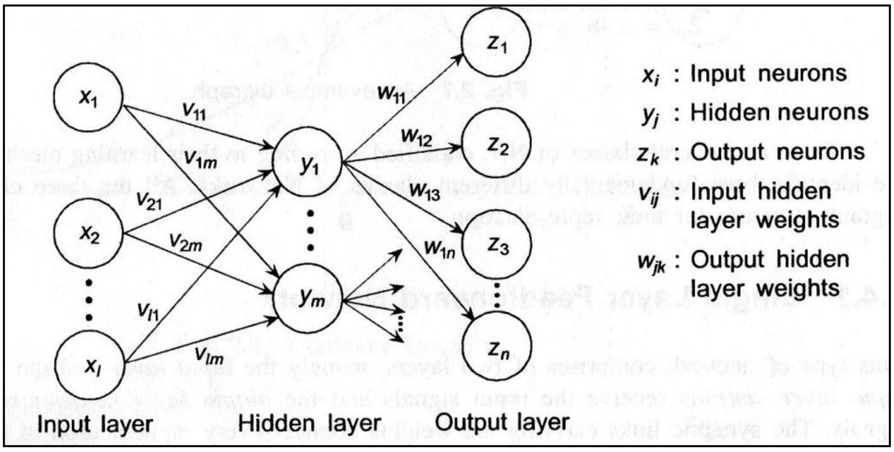

Fig. 1 : A multilayer feedforward neural network

As all techniques possesses , Backpropagation too has its pros and cons and has its problems such as slow convergence rate and problem to get stuck in local minima however, it is known for its accuracy. Backpropagation was less to be used because of its time length needed to train the network to achieve the best result possible. There has been lot of research done to improve these limitations.

In Section 2 , the detailed survey about the techniques that has evolved over the past years to overcome the limitations of backpropagation has been summarized. Through study it has been found out that , One of the parameter of a BPA is MSE cost function which has lot of limitations such as incorrect saturation and lean to stuck in local minima problem, resulting in slow convergence and poor performance [26]. So this section covers also the survey regarding different major cost functions which proved to improve the performance of the BPA.

II. LITERATURE SURVEY

This section consists a detail literature survey about the research that has been done so far about the area of BACK-PROPAGATION in Multilayer feed forward neural network (MLFFNN) and the gradual improvements that has been achieved by the researchers over its limitations that have been gathered in a tabular format followed by the descriptions of some of the major and pivotal work in these area.

A. Research done so far in Backpropagation Learning algorithm

With all the researchers and their methods and conclusions regarding the study about the gradual improvements of the back propagation techniques are described in the Table. 4.1.

Table 4.1 MAJOR CONTRIBUTION AND STUDY IN AREA OF BPNN[29]

Year	Researcher	Method/Conclusion of Work
1969	Minsky,M.L ad Papert, S[1].	Introduced the PERCEPTRON MODEL.
1986	Rumelhart et al[2].	Introduce the error back propagation (BP) method to train multilayer feedforwaed neural networks.
1987	Richard P. Lippmann[3].	Given the introduction to computing with neural nets and its different algorithms and classification problems including single layer perceptron and kohonan and multilayer perceptron and introduction of BACK PROPAGATION Algorithm.
1988	Robert A. Jacobs[4].	To achieve faster rates of convergence than steepest descent algorithms he has examined two implementations namely the momentum and delta-bar-delta and the hybrid of them also.
1988	Scott E. Falhman[5].	Compared the Quick propagation against standard back propagation algorithm using test over benchmark problems like Exclusive-Or problem, Encoder problem and found positive results.
1993	Yam and Chow[6].	According to the coefficient of correlation between the prior weight change and the downhill gradient the momentum factor and the learning rate are modified.
1994	Verma and Mulawka[7].	By solving , weight matrix for the output layer of the network using least squares method and theory of equations, adjustment is made .
1995	Drago et al[8].	An Adaptive Momentum Back Propagation (AMBP,ABP, Accelerated Learning) is

		said to attain very satisfying performance for achieving fast minimum search. The network weight update rule is chosen such that the error function is forced to behave in a certain manner that accelerates convergence. Besides the good convergence speed, a high generalization capability has been achieved,
1995	Bossan et al[9].	This technique try to decrease the time which has been spent rigorously for attaining low MSE in those dense regions of the pattern space while ignoring patterns in sparse regions of it until large number of training epochs occur.
1997	Chen et al[10].	A randomized BP algorithm in which a series of weight vectors are chosen over the learning phase is proposed.
1998	Fukuoka et al[11].	Each connecting weight in a network is multiplied by a factor in the range of (0,1] at a constant interval during a learning process. The basic idea of the method is to keep sigmoid derivative relatively large while some of the error signals are large.
1999	Ng and Leung[12].	The new proposed back-propagation algorithm is to change the derivative of the activation function so as to magnify the backward propagated error signal, thus the convergence rate can be accelerated and the local minimum can be escaped.
2000	Wen et al.	An adaptive backpropagation algorithm which can update learning rate and inertia factor automatically based on dynamical training error rate of change.
2001	Abid et al.	This approach minimizes a modified form of the criterion used in the standard backpropagation algorithm. This criterion is based on the sum of the linear and the nonlinear quadratic errors of the output neuron.
2002	Yu and Liu[13].	BPALM (Backpropagation with adaptive learning rate and momentum term) - adaptive learning rate and momentum term where the learning rate and momentum factor are adjusted at each iteration to reduce the training time.
2003	Zweiri et al[14].	Besides learning rate and momentum factor of backpropagation algorithm a new third term called proportional factor is proposed to fasten the weight adjustment process.
2004	Wang et al.	Improved BP where each training pattern has its own activation function of neurons in hidden layer to avoid local minima.
2005	Pernia-Espinoza et al.	The benfit of the non-linear regression model -estimates (introduced by Tabatabai and Argyros, 1993) is combined with the backpropagation algorithm to produce the TAO-robust learning algorithm.
2006	Kathivalavakumar and Thangavel.	A new technique and optimization criterion is proposed to train single hidden layer FFNN where it trains the hidden layer and output layer independently.
2007	Wang et al.	An Individual Inference Adjusting Learning Rate technique (IIALR) is proposed to enhance the learning performance of the BPNN.
2007	Sammy Siu et al[15].	By using Evolutionary technique Back propagation algorithm is improved.
2007	Guijarro and Fontenla[16].	An algorithm which applies linearleast-squares is proposed. It combines linear-least-squares with gradient descent. It improves the learning rate of the basic backpropagation algorithm in several orders of magnitude, while maintaining good optimization accuracy.
2007	Shamsuddin, Siti Mariyam, Maslina Darus, and Fadhlina Izzah Saman[17].	Experiments are conducted using three UCI dataset; Balloon, Iris and Cancer. The results show that the 3-Term BP outperforms standard BP for small scale data but does not work well for medium and large scale dataset.
2008	Bumghi, Ju, Deok[18].	A novel idea is proposed to solve the LOCAL MINIMA problem faced in FFNN.
2011	Kavita Burse,M.manoria,vishnu[19].	Implemented Zweiri's three -term BPLA over XOR problem and found it is very easy to solve the local minima problem in Multiplicative Neuron Model.
2011	Zhen G che,Zhen H che,Tzu[20].	Compared the BPLA with the Genetic algorithm and found in some cases it is faster than even Genetic and not much complex.
2011	Chukwuchekwa Ulumma Joy[21].	By using pattern recognition problems, comparisons are made based on the effectiveness and efficiency of both backpropagation and genetic algorithm training algorithms on the networks. The backpropagation algorithm is found to outperform the genetic algorithm in this instance.
2013	Yeremia, Hendy, et al[22].	In this study, backpropagation network algorithm is combined with genetic algorithm

		to achieve both accuracy and training swiftness for recognizing alphabets. The training time needed for backpropagation learning phase improved significantly from 03 h, 14 min and 40 sec, a standard backpropagation training time, to 02 h 18 min and 1 sec for the optimized backpropagation network. A hybrid approach proves to be improvising the network performance.

B. Cost Functions in Backpropagation Network

Simple Back propagation algorithm uses Mean square error(MSE) as a cost function. MSE gives more emphasis on reducing the larger errors as compared to smaller errors due to the squaring that takes place. It is due to the summation of the errors for all input patterns; if a class is not well presented and happens to have small errors, it may be completely ignored by the learning algorithm [26].

Table 4.2 describes other functions other than major MSE which proved to be improving the performance of the backpropagation algorithm.

Table. 4.2. MAJOR COST FUNCTIONS FOR BPNN[29]

Author	Description
Otair and Salameh[25]	They studied three algorithms and different versions of backpropagation training for stable learning and robustness to oscillations. The new modification consists of a simple change in the error signal function. These algorithms have been tested on OR problem, encoding problem and character recognition.
Rimer and Martinez[26]	Authors have proposed Classification-Based (CB) cost functions that attempted to guide the network directly to correct pattern classification. It reduces average test error. However, these functions only applicable for Classification-based (CB) problems.
Zhang et al.	He has applied a cost function of Wang for element neural network. This method can avoid the local minima problem and accelerate the speed of the convergence.
Lv andYi[27].	Presented an absolute cost function using the Lyapunov method in 2005. This algorithm is more robust and faster for learning than standard BP when target signals include some incorrect data. It also avoids local minima.
Chow et al[28].	Bernoulli error measure approach to train FeedForward Artificial NN for classification problems.
Samsuddin et al[30].	New modified cost function (MM) is proposed. It has been proved to improve convergence speed of standard backpropagation. However, for some datasets, the netwoek does not converge.

Siti Mariyam Shamsuddin, Razana Alwee, Puspadevi Kuppusamy[29] compared four cost functions in Three Term BP network. It has been tested on Balloon, Cancer, Diabetes and Pendigits datasets. Through the experiments found out that MM cost function is the best cost function compared to BL(Bernoulli Function), Modified cost function (MM) and Improved cost function (IC) cost function. It is suitable for most of the real world problems that requires high accuracy but with a moderate speed.

C. Why Backpropagation is better?

Backpropagation (BP) is a well known network that has been known for its accuracy because it allows itself to learn and improving itself thus it can achieve higher accuracy[22].

Zhen Guo che, Tzu-An Chiang and Zhen Hua Che [20] has done one of the comparison between genetic algorithm and back propagation learning algorithm over different problems such as Sin funciton, Iris plant and Diabetes datasets and found that BPLA is faster than GA in terms of training speeds and also in terms of CPU time required, the BPLA yields better results than GA. Chukwuchekwa Ulumma Joy[21] has proved using pattern recognition problem that BPA outperforms GA.

On the other hand backpropagation was less to be used because of its time length needed to train the network to achieve the best result possible[22]. When faced against the larger datasets backpropagation still stuck into local minima problem but researcher like Bumghi Choi et al.[18] has shown one of the novel idea and proposed the algorithm to avoid the local minima problem in complex problems and shown that there is still scope to fasten the BP Algorithm for larger and complex problems.

III. CONCLUSION

This survey paper is all about the evolution of the backpropagation algorithm which is a supervised learning method to train the ANN. BP Algorithm is known for its mathematical simplicity and accuracy. BPA too, has its own limitations of slow convergence rate and local minima problem which is still a big problem when dealing with large complex problems. There have been significant research done to overcome these problems and different variations of BPA has been proposed. It also surveys about the cost functions that could be important element in improving the BP algorithm performance. From this paper we can conclude that even though several variations and different techniques have been suggested to improve the performance of BPA no one guarantees to a global solution which gives rise to further research.

REFERENCES

[1] Minsky, M. L., and S. Papert. "Perceptrons: an introduction to computational geometry. Expanded." (1988).
[2] Rumelhart, D.E., Hinton, G. E., and Williams, R, J. Learning Internal Representations by Error Propagation.1986
[3] Lippmann, Richard. "An introduction to computing with neural nets." ASSP Magazine, IEEE 4.2 (1987): 4-22.
[4] Jacobs, Robert A. "Increased rates of convergence through learning rate adaptation." Neural networks 1.4 (1988): 295-307.
[5] Fahlman, S.E. (1988a), An Empirical Study of Learning Speed in Backpropagation Networks, Carnegie Mellon Report, No CMU- Cs, pp. 88-162.
[6] Yam, Y. F., and T. W. S. Chow. "Extended backpropagation algorithm." Electronics Letters 29.19 (1993): 1701-1702.
[7] Verma, B. K., and J. J. Mulawka. "A modified backpropagation algorithm." Neural Networks, 1994. IEEE World Congress on Computational Intelligence., 1994 IEEE International Conference on. Vol. 2. IEEE, 1994.
[8] Drago, Gian Paolo, Mauro Morando, and Sandro Ridella. "An adaptive momentum back propagation (AMBP)." Neural Computing & Applications 3.4 (1995): 213-221.
[9] Bossan, M. C., et al. "A modified backpropagation algorithm for neural classifiers." Circuits and Systems, 1995., Proceedings., Proceedings of the 38th Midwest Symposium on. Vol. 1. IEEE, 1995.
[10] Xiao-hu yu ,Guo-an Chen, Efficient Backpropagation learning using optimal learning rate and momentum. Neural networks Vol. 10. No. 3. pp 517-527,1997.
[11] Fukuoka, Yutaka, et al. "A modified back-propagation method to avoid false local minima." Neural networks 11.6 (1998): 1059-1072.
[12] Ng, Sin Chun, Shu Hung Leung, and Andrew Luk. "Fast convergent generalized back-propagation algorithm with constant learning rate." Neural Processing Letters 9.1 (1999): 13-23.
[13] Yu, Chien-Cheng, and Bin-Da Liu. "A backpropagation algorithm with adaptive learning rate and momentum coefficient." Neural Networks, 2002. IJCNN'02. Proceedings of the 2002 International Joint Conference on. Vol. 2. IEEE, 2002.
[14] Zweiri, Yahya H., James F. Whidborne, and Lakmal D. Seneviratne. "A three-term backpropagation algorithm." Neurocomputing 50 (2003): 305-318.
[15] Siu, Sammy, et al. "Improving the back-propagation algorithm using evolutionary strategy." Circuits and Systems II: Express Briefs, IEEE Transactions on 54.2 (2007): 171-175.
[16] Guijarro-Berdiñas, Bertha, et al. "A linear learning method for multilayer perceptrons using least-squares." Intelligent Data Engineering and Automated Learning-IDEAL 2007. Springer Berlin Heidelberg, 2007. 365-374.
[17] Shamsuddin, Siti Mariyam, Maslina Darus, and Fadhlina Izzah Saman. "Three term backpropagation algorithm for classification problem." Neural Network World 17.4 (2007): 363.
[18] Bumghi Choi,Ju-Hong Lee,Deok-Hwan Kim, Solving local minima problem with large number of hidden nodes on two-layer feed forward artificial neural networks Neurocomputing 71 (2008).
[19] Burse, Kavita, Manish Manoria, and Vishnu Pratap Singh Kirar. "Improved back propagation algorithm to avoid local minima in multiplicative neuron model."Information Technology and Mobile Communication. Springer Berlin Heidelberg, 2011. 67-73.
[20] Che, Zhen-Guo, Tzu-An Chiang, and Zhen-Hua Che. "Feed-forward neural networks training: a comparison between genetic algorithm and back-propagation learning algorithm." Int J Innov Comput Inf 7 (2011): 5839-5850.
[21] Joy, Chukwuchekwa Ulumma. "Comparing the Performance of Backpropagation Algorithm and Genetic Algorithms in Pattern Recognition Problems."International Journal of Computer Information Systems 2.5 (2011): 7-12.
[22] Yeremia, Hendy, et al. "GENETIC ALGORITHM AND NEURAL NETWORK FOR OPTICAL CHARACTER RECOGNITION." Journal of Computer Science 9.11 (2013): 1435.
[23] Rajasekaran, Sundaramoorthy, and GA Vijayalakshmi Pai. *NEURAL NETWORKS, FUZZY LOGIC AND GENETIC ALGORITHM: SYNTHESIS AND APPLICATIONS (WITH CD)*. PHI Learning Pvt. Ltd., 2003.
[24] Falas, Tasos, and A-G. Stafylopatis. "The impact of the error function selection in neural network-based classifiers." *Neural Networks, 1999. IJCNN'99. International Joint Conference on*. Vol. 3. IEEE, 1999.
[25] Otair, Mohammed A., and Walid A. Salameh. "Efficient training of backpropagation neural networks." *Neural Network World* 16.4 (2006): 291.
[26] Rimer, Michael, and Tony Martinez. "CB3: an adaptive error function for backpropagation training." *Neural processing letters* 24.1 (2006): 81-92.
[27] Lv, Jiancheng, and Zhang Yi. "An improved backpropagation algorithm using absolute error function." *Advances in Neural Networks–ISNN 2005*. Springer Berlin Heidelberg, 2005. 585-590.
[28] Chow, Mo-Yuen, et al. "Bernoulli error measure approach to train feedforward artificial neural networks for classification problems." *Neural Networks, 1994. IEEE World Congress on Computational Intelligence., 1994 IEEE International Conference on*. Vol. 1. IEEE, 1994.
[29] Shamsuddin, Siti Mariyam, et al. "Study of cost functions in Three Term Backpropagation for classification problems." *Nature & Biologically Inspired Computing, 2009. NaBIC 2009. World Congress on*. IEEE, 2009.
[30] Shamsuddin, Siti Mariyam, Md Nasir Sulaiman, and Maslina Darus. "An improved error signal for the backpropagation model for classification problems."*International Journal of Computer Mathematics* 76.3 (2001): 297-305.