International Journal of Information and Computation Technology.
ISSN 0974-2239 Volume 3, Number 4 (2013), pp. 323-328

http://www. irphouse.com /ijict.htm

Recapitulation on Transformations in Neural Network Back Propagation Algorithm

Devika Chhachhiya, Amita Sharma and Manish Gupta

IIS University, Rajasthan, India.
IIS University, Rajasthan, India.
Raj. Council for Elementary Education, Rajasthan, India.

Abstract

Back propagation algorithm is the most accepted and the primal training procedure for feed forward Artificial Neural Network Architecture (ANN). ANN is a mathematical model based on the human nervous system which employs back propagation algorithm for training-testing phase. Back propagation (BP) algorithm minimizes the error function iteratively based on gradient descent technique. Although it has its own precincts, it is applied in an extensive range of practical problems and generally works well. The major limitations of the algorithm are low rate of convergence, learning rate dependency and getting stuck in local minima resulting in sub-optimal solutions. Due to these constraints, the back propagation algorithm has been transformed repeatedly. Over the years, many improvements and modifications of the BP learning algorithm have been reported. The paper categorizes the transformed algorithm and discusses different modification and new evolved optimization techniques which are capable of overcoming the limitation of back propagation algorithm. The uniqueness of this study is to catalog the untouched combination of optimization technique to transform algorithm.

Keywords: ANN, BP, Quassi-newton, Conjugate gradient.

1. Introduction

Neural network is a mathematical model based on the human nervous system. It is getting popular due to its fast response and learning ability from the examples. Its

characteristics like robustness, self organizing, adaptive, and high degree of fault tolerance distinguish it from other traditional paradigms.
The neural network model can be broadly divided into the following three types:

1.1 Feed-forward network: In this network output from one layer of neurons feeds forward into the next layer of neurons. There are never any backward connections, and connections never skip a layer.

1.2 Recurrent network: This type of network have at least one feedback loop. Mainly used for associative memory and optimization calculation.

1.3 Self-organization networks: This network is based on unsupervised learning. In this network target output is not known to the network. Mainly used for cluster analysis.

Back propagation algorithm is a one of the most primitive training procedure used for feed forward neural network (MLP) that minimizes the error function iteratively based on gradient descent technique. It is easy to implement and is capable to solve all kinds of problems. But still there are some limitations associated with which need to be removed so that neural networks strongest features can be completely utilized. Over the years, many enhancements and modifications of the BP learning algorithm have been recommended to improve the rate of convergence and solve the problem of local minima.

Motive of this paper is to analyze and categorize all the alterations of the backpropagation algorithm in different categories. Section1 gives brief introduction neural network and backpropagation algorithm. Further section 2 backpropagation algorithm is explained in detail. Section 3 consists of the approaches and the algorithm proposed to overcome the limitations of the BP algorithm. Finally in section 4 conclusion is discussed and section 5 for references.

2. Backpropagation (Learning Algorithm)

The back propagation is a learning algorithm used for feed forward neural network, which follows the supervised learning method.. During the learning process of the network, one of the input patterns is applied to the network input layer whose output is already known to the network. Each layer computes the weighted sum of the neurons and passes it through an activation function. The result obtained by it is used an input to the other layer. Finally output layer generates an output pattern which is then compared to the target pattern. Depending on the difference between output obtained and the target, a mean squared error is obtained. This difference or error indicates the network's learning rate and its efficiency.

The purpose of the back propagation algorithm is to reduce this error which is achieved by back propagating the weights and adjusting them such that the error may be reduced.

2.1 Algorithm

1. Initialize weights with random values and set other parameters (learning rate, momentum and error precision).
2. Read in the input vector and the desired output associated with it.
3. Compute the output of the network by calculating a weighted sum of the input neurons and passing it through an activation function, and propagating through the Layers.
4. Compute the mean square error i.e. difference between the output obtained and the target output.
5. Adjust the weights by back propagating them from the output layer.

3. Approaches Adopted to Modify Backpropagation Algorithm

Standard back propagation algorithm highly depends on the learning rate parameter and its rate of convergence is also very slow. Several adaptations of the backpropagation algorithm have been proposed till now few of them have been discussed below. Adaptations have been classified into two broad categories:

3.1 Based on Numerical optimization procedures using second-order information.

3.1.1 Conjugate gradient

CG uses the hessian matrix H of second-order derivatives and its inverse to calculate the step size and the linear combinations of the previous and present gradient directions to find new search directions.

$$W_{k+1}=w_k+\eta_k\, d_k$$

$$d_{k+1}=g_{k+1}+\beta_k\, d_k$$

Different formulas exist to calculate βk in order consecutive directions to be conjugate

Table 1: Algorithms for conjugate gradient Method.

NAME	PROPOSED	FORMULA	DESCRIPTION
The Fletcher-Reeves Method	By Fletcher and Reeves	$\beta_k=\frac{g^T_{k+1}g_{k+1}}{\square^{\square}_{\square}\ \square_{\square}}$	Next search direction β is found such that it is conjugate to previous search directions

The polak-Ribiere Method	By Polak and Ribere	$\beta_k = \frac{g_{k+1}^T[g_{k+1}-g_k]}{\square_{\square}^{\square}\ \square_{\square}}$	β is found by computing the product of previous gradient with current gradient and divided by square of previous gradient
The Hestenes-stiefel Method	By Hestenes and stiefel	$\beta_k = \frac{g_{k+1}^T[g_{k+1}-g_k]}{\square_{\square}^{\square}\ [g_{k+1}-g_k]}$	β is found by computing the product of previous gradient with current gradient and divided by d_k and the previous gradient

3.1.2 Quasi-Newton

In Quasi Newton method the modeling of the incremental change of the objective functions between iterations is done with Taylor series expansion. Use of Hessian matrix H provides direction as well as step size for the optimization step. There are several proposed methods to derive H_k few of them are discussed below.

$$\Delta w = -H^{-1} g$$

Table 2: Algorithms for Quassi-Newton Method.

NAME	PROPOSED	FORMULA	DESCRIPTION
Davidon -Fletcher-Powell (DFP) Method	By Davidon , Fletcher and Powell	$\mathbf{B}_{k+1} = \mathbf{B}_k + \frac{\mathbf{p}_k\mathbf{p}_k^T}{\mathbf{p}_k^T\mathbf{q}_k} - \frac{\mathbf{B}_k\mathbf{q}_k\mathbf{q}_k^T\mathbf{B}_k}{\mathbf{q}_k^T\mathbf{B}_k\mathbf{q}_k}$	Bk is hessian matrix
Broyden-Fletcher-Goldfarb-Shanno (BFGS) Method	By Broyden-Fletcher-Goldfarb-Shanno	$\mathbf{B}_{k+1} = \mathbf{B}_k + \frac{\mathbf{p}_k\mathbf{p}_k^T}{\mathbf{p}_k^T\mathbf{q}_k} - \frac{\mathbf{B}_k\mathbf{q}_k\mathbf{q}_k^T\mathbf{B}_k}{\mathbf{q}_k^T\mathbf{B}_k\mathbf{q}_k}$	H is the Hessian matrix
Levenberg-Marquardt (LM) algorithm	By levenberg and Marquardt	$\Delta w_k = -H_k' g_k$ $H' = j'j$ $g_k = j'e$	J is the jacobian matrix and e is the error vector containing output errors for each input vectors.

3.2 Methods based on Stochastic Optimization

Evolutionary algorithms

Swarm intelligence models are referred to as computational models inspired by the behavior natural swarm systems (i.e. insect's bugs and animals). Several swarm intelligence models based on different natural swarm systems have been proposed and successfully applied in many real - life applications few of them are discussed below.

Table 3: Algorithms for Quassi-Newton Method.

NAME	PROPOSED	FORMULA	DESCRIPTION
Ant Colony Optimization	By M. Dorigo et al	$P_{ij}^{k}(t) = \begin{cases} \frac{[\tau_{ij}(t)]^{\propto}(\eta_{ij})^{\beta}}{\Sigma_{l\in N_k^i}[\tau_{il}(t)]^{\propto}(\eta_{ij})^{\beta}} & j\in N_k^i \\ 0 & j\notin N_k^i \end{cases}$	This algorithm is based on behavior of Ants $P_{ij}^{k}(t)$ is the probability of the *kth* ant to move from node *i* to node *j* at the *tth* iteration/time step. N_k^i is the set of nodes in the neighborhood of the *kth* ant in the *ith* node.
Particle Swarm Optimization	By Kennedy and Eberhart	*vid(t+1)= vid(t) + c1* ***R****1(pid(t) – xid(t)) + c2* ***R****2 (pgd(t) – xid(t))* *xid(t+1) = xid(t) + vid(t+1)*	Based on social flocking behaviour of birds *vid* represents the rate of the position change (velocity) of the *i*th particle in the *d*th dimension, and *t* denotes the iteration counter. *xid* represents the position of the *i*th particle in the *d*th dimension.
Bacterial Foraging algorithm	By Passino	$J(i,j,k,l) = j(i,j,k,l) + J_{cc}(\theta^{i}(j,k,l), P(j,k,l))$ $\theta^{i}(j+1,k,l) = \theta^{i}(j,k,l) + C(i)\frac{\Delta(i)}{\sqrt{\Delta^{T}(i)\Delta(i)}}$	Based on the behavior of the bacteria $J(i,j,k,l)$ is the fitness function
Cuckoo search Algorithm	By Yang and Deb	$x_i^{t+1} = x_i^{t} + \alpha \oplus Levy(\lambda)$	Based on obligate brood parasitic behavior of some cuckoo species in combination with the Le ´vy flight behavior of some birds and fruit flies The product $\oplus$ means entry-wise multiplications. Lévy flights essentially provide a random walk

4. Conclusion and Discussion

Backpropagation algorithm is the milestone of the feed forward neural network. It is the most primitive algorithm, although it has some limitations but it still gives better results than many other proposed algorithms for general problems. We have discussed only few well known modifications proposed for the BP which are capable to reduce the limitations of BP algorithm and gives better accuracy and consistency. A Hybrid of two and more than two algorithms is also one good solution to combine the best features of all the algorithms and make it one algorithm such that it can lead to fast convergence and problem of getting stuck in local minima can also be removed. Several swarm intelligence algorithms are combined with backpropagation algorithm to form an hybrid such that limitations of standard BP algorithm can be eliminated. Some of the hybrids are Ant Colony Optimization + BP, Particle Swarm Optimization + BP, Bacterial Foraging algorithm + BP, Cuckoo search Algorithm + BP.

References

[1] S. P. Deshpande and Dr. V. M. Thakare ," DATA MINING SYSTEM AND APPLICATIONS: A REVIEW", *International Journal of Distributed and Parallel systems (IJDPS),* Vol.1, Issue1, 2010

[2] K. Amarendra, K.V. Lakshmi & K.V. Ramani," Research on Data Mining Using Neural Networks", *International Journal of Computer Science & Informatics (IJCSI),* Vol.- II, Issue-1, 2011

[3] Y. Singh and A. S. Chauhan," NEURAL NETWORKS IN DATA MINING", *Journal of Theoretical and Applied Information Technology* 2005 - 2009 JATIT.

[4] A. Raj, Bincy G, T.Mathu," Survey on Common Data Mining Classification techniques", *International Journal of Wisdom Based Computing,* Vol. 2, Issue 1, April 2012.

[5] Hazem Ahmed and Janice Glasgow ," Swarm Intelligence: Concepts, Models and Applications ",*Technical Report 2012-585* , School of Computing Queen's University, Ontario, Canada 2012.

[6] Salim Lahmiri, "A COMPARATIVE S TUDY OF BACKPROPAGATION ALGORITHMS IN FINANCIAL PREDICTION ", *International Journal of Computer Science, Engineering and Applications (IJCSEA)* Vol.1, No.4, August 2011

[7] M.Moreira and E. Fiesler, "Neural Networks with adaptive Learning Rate and Momentum Terms", *IDIAP Technical Report*, Oct, 1995.