STUDY OF COST FUNCTIONS IN THREE TERM BACKPROPAGATION FOR CLASSIFICATION PROBLEMS

PUSPADEVI A/P KUPPUSAMY

A project report submitted in partial fulfillment of the
requirements for the award of the degree of
Master of Science (Computer Science)

Faculty of Computer Science and Information System
Universiti Teknologi Malaysia

OCTOBER 2008

# ABSTRACT

Three Term Backpropagation was proposed in 2003 by Zweiri, and has outperformed standard Two Term Backpropagation. However, further studies on Three Term Backpropagation in 2007 indicated that the network only surpassed standard BP for small scale datasets (below 100 instances) but not for medium and large scale datasets (above 100 instances). It has also been observed that by using Mean Square Error (MSE) as a cost function in Three Term Backpropagation network, has some drawbacks such as incorrect saturation and tend to trap in local minima, resulting in slow convergence and poor performance. In this study, substantial experiments on implementing various cost functions on Three Term BP are executed to probe the effectiveness of this network. The performance is measured in terms of convergence time and accuracy. The costs functions involve in this study include Mean Square Error, Bernoulli function, Modified cost function and Improved cost function. These cost functions were introduced by previous researchers. The outcome indicates that MSE is not an ideal cost function to be used for Three Term BP. Besides that, the results have also illustrated that improve cost function's converges faster, while modified cost function produces high accuracy in classification

## ABSTRAK

Algoritma rambatan balik dengan tiga terma telah diperkenalkan oleh Zweiri pada 2003, dan telah berjaya mengatasi prestasi rangkaian rambatan balik tradisi iaitu rangkaian rambatan balik dua terma. Walaubagaimanapun, kajian yang telah dilaksanakan pada 2007 telah mendapati bahawa rangkaian rambatan balik tiga terma hanya dapat mengatasi prestasi rangkaian rambatan balik tradisi pada data yang bersaiz kecil (kurang daripada 100 data) dan bukan pada data yang bersaiz sederhana atau besar(besar dari 100 data). Oleh yang demikian, boleh dinyatakan bahawa fungsi ralat piawai iaitu Ralat Min Kuasa Dua mempunyai beberapa kelemahan seperti penumpuan yang amat perlahan, sering terperangkap pada minima setempat dan prestasi yang kurang baik. Kajian ini menjalankan eksperimen yang komprehensif terhadap beberapa fungsi ralat bagi rangkaian rambatan balik tiga terma bagi mencari keberkesanan fungsi kos tersebut. Prestasi rangkaian diukur dari aspek kepantasan kadar penumpuan dan ketepatan pengelasan. Fungsi kos yang terlibat adalah Ralat Min Kuasa Dua, fungsi ralat '*Bernoulli'*, fungsi ralat yang telah 'diubahsuai', dan fungsi ralat pembaikan. Hasil kajian mempamerkan bahawa fungsi Ralat Min Kuasa Dua tidak begitu sesuai untuk algoritma rambatan balik tiga terma. Hasil kajian juga telah memperlihatkan bahawa fungsi ralat pembaikan memberi kadar penumpuan yang pantas manakala fungsi ralat yang 'diubahsuai' memberikan kadar pengelasan yang lebih tepat.

## TABLE OF CONTENT

# CHAPTER 1

# INTRODUCTION

## 1.1 Introduction

Artificial Neural Network (ANN) is a model of reasoning based on the human brain. It consists of a number of simple highly interconnected processors known as neurons, which are analogous to the biological neural cells of the brain. These neurons are connected by a large number of weighted links (Ibrahim dan Al-shams, 1997). Learning is a fundamental and essential characteristic of ANN. It is capable of learning through the network experiences to improve their performance. When ANN is exposed to a sufficient number of samples, it can generalise well to other data that they have not yet encountered (Negnevitsky, 2004).

Generally, ANN can be trained using backpropagation (BP) developed by Rumelhart, Hinton and Williams in 1986. Studies have shown that BP has been proven to be very successful in many diverse applications (Hauger, 2003). ANN training usually updates the weights iteratively using the negative gradient of a Mean Squared Error (MSE) function, multiplied by the slope of a sigmoid activation function. MSE is referred to the difference between desired and actual output values. The error signal is then backpropagated to the lower layers (Zweiri *et al*., 2003).

Then an activation function will transform the input into its own value range accordingly. There are many activation functions available such as step, sign, linear and sigmoid. The most popular activation function is sigmoid function. The sigmoid function transforms the input, which can have any value between plus and minus infinity into reasonable value in the range between 0 and 1 (Hauger, 2003). BP network's neuron uses this function to produce a standard outputs.

The outputs will be compared with the targeted output and it will backpropagates to adjust the weights. There are two parameters used in controlling weight adjustment of standard backpropagation. These are learning rate (LR) and momentum factor (MF). Recently, a new term known as proportional factor is added to the formulation to speed-up the weight adjusting process by Zweiri *et al*. (2003). This formulation is known as three term BP.

The derivative of the cost function is one of the factors in the equation of weight adjustment. This is important to determine the success of the application, to train the network with an error function that resembles the objective of the problem at hand (Falas and Stafilopatis, 1999). In most practical applications, MSE is the most commonly used cost function in BP network.

## 1.2 Problem Background

Three Term Backpropagation was proposed by Zweiri *et al.* (2003). It involves Proportional Factor (PF) besides Learning Rate (LR) and Momentum Factor (MF) for error adjustment in the algorithm. According to Zweiri *et al*., it has outperformed standard Two Term Backpropagation with less complexity, low computational cost and easy tuning to suit a particular application. It is noted that the new algorithm archives efficiency while maintaining a similar computational complexity to the conventional BP algorithm. This is in contrast to other alternative BP algorithms, which requires complex and costly calculations at each iteration to archive faster rates on convergence. Moreover in contrast to the proposed algorithm, most standard acceleration techniques must be tuned to fit particular application. This new term also can be viewed as being analogous to the common three term proportional integral derivative (PID) algorithm used in feedback control. PID controller is a generic control loop feedback mechanism widely used in industrial control systems. However, further studies on Three Term Backpropagation by Shamsuddin, Darus and Saman (2007) indicated that the network only outperformed standard BP for small scale datasets (less then 100 instances) but not for medium and large scale datasets (more then 100 instances).

Meanwhile, researches have identified proper cost function is being an important factor to improve the performance of Two Term BP in terms of convergence speed (Humpert, 1994; Neelakanta, 1996; Dhiantravan, 1996; Oh and Lee, 1999; Taji *et al*., 1999; Shamsuddin *et al.*, 2001; Jiang *et al*., 2003; Wang *et al.*, 2004; Lv and Yi, 2005; Choi *et al*., 2005; Otair and Salameh, 2006; Zhang, 2007), in terms of higher accuracy (Telfer and Szu, 1994; Rimer and Martinez, 2006) and to overcome the problems of getting stuck into local minima (Telfer and Szu, 1994; Oh and Lee, 1999; Jiang *et al*., 2003; Wang *et al*., 2004; Bi *et al.*, 2004; Zhang *et al.*, 2007).

It has been observed that, Mean Square Error cost function employed has drawbacks such as incorrect saturation and tend to trap in local minima, resulting in slow convergence and poor performance (Rimer and Martinez, 2006). Besides that, it gives more emphasis on reducing the larger errors as compared to smaller errors due to the squaring that takes place. Also 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 (Falas and Stafylopatis, 1999).

The need to improve Three Term BP is foreseen, where if a better cost function is applied in the Three Term it could perform better. This is due to the successfulness of researches that claims Two Term BP performed better with their novel cost functions instead of MSE (Wang *et al.*, 2004; Lv and Yi, 2005; Choi *et al*., 2005; Otair and Salameh, 2006; Zhang, 2007; Rimer and Martinez, 2006)

## 1.3 Problem Statement

In Three Term Backpropagation, MSE is employed as its cost function. It has been observed that, MSE cost function employed has drawbacks resulting in slow convergence and poor performance. Falas and Stafilopatis (1999) studied on impact of cost function in neural network classifier. Their result showed that a cost function other than the usual mean square gives a better performance, both in terms of the number of epochs needed for training, as well as the obtained generalization ability of the trained network.

Thus, in this study Mean Square Error, Bernoulli Cost Function of Chow *et al.* (1994), Modified Cost Function of Shamsuddin *et al.* (2001) and Improved Cost Function of Zhang *et al.* (2007) are exploited in Three Term BP to probe the convergence time and accuracy. These cost function were selected because of the simplicity of the formulation that helps to incorporate easily into the Three Term BP. Besides that those cost functions has been tested on various classification problems and proven to be performed well in the Two Term BP. The classification domain was selected or this study since BP is successful in this domain.

Subsequently, the hypothesis of this study can be stated as:

Three Term BP would yield faster convergence speed and better classification accuracy with cost functions other then MSE.

## 1.4 Project Aim

The aim of this project is to study the effectiveness of exploiting novel cost functions introduced by researches in past years to improve the Two Term BP to be applied in Three Term BP to increase the convergence speed and to produce high accuracy.

## 1.5 Objectives

In order to accomplish the hypothesis of the study, few objectives have been identified.

1. To study the cost functions of previous researches especially Mean Square Error (MSE) cost function, Bernoulli (BL) cost function, Modified (MM) cost function and Improved (IC) cost function.

2. To conduct experimental comparisons of MSE cost function, BL cost function, MM cost function and IC cost function in Three Term BP for classification problems.

## 1.6 Project Scope

The scopes of this project are defined as follows:

I. Datasets that will be employed are Balloon with 16 instances, Cancer with 500 instances, Diabetes with 768 instances and Pendigits with 1000 instances.

II. Three Term BP with the following cost functions are used in this study:

   a. Three Term BP with MSE cost function
   b. Three Term BP with BL cost function of Chow *et al.* (1994)
   c. Three Term BP with MM cost function of Shamsuddin *et al.* (2001)
   d. Three Term BP with IC cost function of Zhang *et al.* (2007)

III. Develop Three Term BP with MSE cost function, Three Term BP with BL cost function, Three Term BP with MM cost function and Three Term BP with IC cost function using Microsoft Visual C++ 6.0.

IV. Experiments will be conducted for Three Term BP only. Two Term BP will not be tested.

V. The network architecture is three layers consist of one input layer, one hidden layer and one output layer to standardize the comparison criteria.

VI. Experimental setting with ‘K+10 or K+100 Increment Rule’ for the number of epochs.

## 1.7 Significance of the Project

This project studied the performance of Three Term BP with MSE cost function, Three Term BP with BL cost function, Three Term BP with MM cost function and Three Term BP with IC cost function. The outcomes of this study will contribute to verify the performance of those cost functions for Three Term BP. Furthermore, this study will spark future research in Three Term BP algorithm.

## 1.8 Organization of Report

This report consists of five chapters. The chapter 1 presents introduction to project, problem background, objective, scope and significant of this study. Chapter 2 reviews the ANN, Two Term BP, Three Term BP, Research trends of BP Learning, Research trends of cost function in BP Network, MSE cost function, BL cost function, MM cost function and IC cost function and also importance of cost functions. Chapter 3 discusses on the methodology used in this study. It also explains details of datasets being used and network architectures. Chapter 4 is the experimental result study. Chapter 5 is the conclusion and suggestion for future work.

**REFERENCE**


Abid, S., Fnaiech, F., and Najim, M. (2001). A Fast Feedforward Training Algorithm Using a Modified Form of the Standard Backpropagation Algorithm. *IEEE Transactions On Neural Networks*, 12(2):424-430.

Asuncion, A. & Newman, D.J. (2007). UCI Machine Learning Repository [http://www.ics.uci.edu/~mlearn/MLRepository.html]. Irvine, CA: University of California, School of Information and Computer Science.

Bi, W., Wang, X., Zong, Z., and Tang, Z.(2004). Modified Error Function with Added Terms for the Backpropagation Algorithm. ISNN 2004, *LNCS* 3173, pp. 338–343, 2004.

Bossan, M.C., Seixas, J.M., Caloba, L.P., Penha, R.S., and Nadal, J. (1995) A Modified Backpropagation Algorithm For Neural Classifiers. *38th IEEE Midwest Symposium on Circuits and Systems*, 1995. Rio de Janeiro, 562-565.

Brian A. Telfer and Harold H. Szu.(1994). Energy functions for minimizing misclassification error with minimum-complexity networks. Neural Networks.7(5): 809-817

Charytoniuk, W. and Chen, M.S (2000). Neural Network Design for Short-term Load Forecasting. *International Conference on Electric Utility Deregulation and Restructuring and Power Technologies 2000*. 4-7 April 2000. City University, London. 554-561.

Chen, Y.Q., Yin, T., and Babri, H.A.(1997). A Stochastic Backpropagation Algorithm for Training Neural Networks. *International Conference on Information, Communications and Signal Processing*, 1997. 9-12 September 1997. Singapore, 703-707.

Choi, S., Lee, T-W., and Hong, D.(2005).Adaptive error-constrained method for LMS algorithms and applications. *Signal Processing.* 85 (2005):1875–1897

Chow, M-Y., Menozzi, A., Teeter, J., and Thrower, J.P. (1994).Bernoulli error measure approach to train feedforward Artificial Neural Networks for Classification problems.

Dhiantravan, Y., and Priemer, R. (1996). Error phenomena of backpropagation learning. Intelligent Engineering Systems Through Artificial Neural Networks. 6:155-160.

Drago, G.P., Morando, M., and Ridella, S. (1995). An Adaptive Momentum Back Propagation (AMBP). *Neural Comput & Application.* 3: 213-221.

Edward, R. J. (2004). An Introduction to Neural Networks A White Paper. United States of America: Visual Numerics Inc.

Fadhlina Izzah Binti Saman (2007). *Three-Term Backpropagation Algorithm For Classification Problem*. Master. Thesis. Universiti Teknologi Malaysia, Skudai.

Falas, T. and Stafylopatis, A-G. (1999). The Impact of The Error Function Selection in Neural Network-based Classifiers. *International Joint Conference on Neural Network.* 3: 1799-1804.

Fukuoka, Y., Matsuki, H., Minamitani, H., and Ishida A. (1998). A Modified Backpropagation Method To Avoid False Local Minima. *Neural Networks*. 11: 1059-1072.

Guijarro-Berdinas, B., Fontenla-Romero, O., Perez-Sanchez, B., and Fraguela, P.(2007). A Linear Learning Method for Multilayer Perceptrons Using Least-Squares. *Lecture Notes in Computer Science 4881*. Berlin Heidelberg: Springer. 365–374.

Guijarro-Berdi, B., Fontenla-Romero, O., Perez-Sanchez, B., and Fraguela, P.(2007). A Linear Learning Method for Multilayer Perceptrons Using Least-Squares. *Lecture Notes In Computer Science*. 365-374.

Hahn-Ming Lee, Chih-Ming Chen, Tzong-Ching Huang. (2001). Learning science improvement of back-propagation algorithm by error saturation prevention method. *Neurocomputing.* 41 (2001) 125-143.

Hauger, S.R.B. (2003). *Ensemble Learned Neural Networks Using Error-Correcting Output Codes and Boosting*. Master Thesis. University of Surrey.

Herself's Artificial intelligence.http://herselfsai.com/2007/02/neural-networks.html. Date Accesed:18/12/2007.

Humpert, B.K.(1994). Improving Back Propagation With A New Error Function. *Neural Networks*.7(8):1191-1192.

Ibrahim, M. E. and Al-Shams, A.A.M (1997). Transient stability assessment using artificial neural networks. *Electric Power Systems Research,* 40, 7-I6.

In-Cheol Kim, Sung-Il Chien*. (2002).* Speed-up of error  ackpropagation algorithm with class-selective relevance. *Neurocomputing*. 48 (2002) 1009– 1014.

Jiang, M., Deng, B., Wang, B. and Zhong, B.(2003). A Fast Learning Algorithm Of Neural Networks By Changing Error Functions. *IEEE International Conference Neural Networks Signal Processing*. December 14-17, 2003, Nanjing. China. 249-252

Kandil N., Khorasani, Patel R.V. and Seed V.K. (1993). Optimum Learning Rate For backpropagation Neural Network. IEEE. 465-468.

Kathirvalavakumar, T., and Thangavel, P. (2006). A Modified Backpropagation Training Algorithm for Feedforward Neural Networks. *Neural Processing Letters*. 23:111-119.

Keogh, E. (2006). The UCR Time Series Data Mining Archive [*http://www.cs.ucr.edu/~eamonn/TSDMA/index.html*]. Riverside CA. University of California - Computer Science & Engineering Department

Liu, C-S., and Tseng, C-H.(1999).Quadratic optimization method for multilayer neural networks with local error-backpropagation. *International Journal of Systems Science*. 30(8):889 - 898.

Lv, J., and Yi, Z.(2005). An Improved Backpropagation Algorithm Using Absolute Error Function. ISNN 2005, LNCS 3496, pp. 585–590, 2005.Lv et al(2005)

Mandischer M. (2002). A comparison of evolution strategies and backpropagation for neural network training. *Neurocomputing*. 42 (2002) 87–117.

Matsuoka Kiyotoshi and Yi Jianqiang (2000). Backpropagation Based on the Logarithmic Error function and Elimination of Local Minima. *IEEE* . 1117-1122.

Neelakanta, P. S. (1996). Csiszar's Generalized Error Measures for Gradient-descent-based Optimizations in Neural Networks Using the Backpropagation Algorithm. Connection Science. 8(1): 79 - 114.

Neural Networks . Statistica is a trademark of StatSoft, Inc. Date accessed 8/12/2007. http://www.statsoft.com/textbook/stneunet.html#multilayere.

Ng, S.C., Leung, S.H., and Luk, A. (1999). Fast Convergent Generalized Back-Propagation Algorithm with Constant Learning Rate. *Neural Processing Letters*. 9:13-23.

Ng S. C., Cheung C. C, Leung S. H., Luk A. (2003). Fast Convergence for Back-Propagation Network with Magnified Gradient Function. *IEEE.* 1903-1908.

Ng, W.W.Y., Yeung, D.S., and Tsang, E.C.C.(2006).Pilot Study On The Localized Generalization Error Model For Single Layer Perceptron Neural Network. *Proceedings of the Fifth International Conference on Machine Learning and*

*Cybernetics*, 13-16 August 2006. Dalian. 3078-3082.

Nii O. Attoh-Okine. (1999). Analysis of learning rate and momentum term in backpropagation neural network algorithm trained to predict pavement performance.
*Advances in Engineering Software*. 30 (1999): 291–302.

Oh, S.H., and Lee, Y.(1995). A Modified Error Function to Improve the Error Back-Propagation Algorithm for Multi-Layer Perceptrons. ETRI Journal. 17(1):11-22.

Oh, S-H.(1997).Improving the Error Backpropagation Algorithm with a Modified Error Function. *IEEE Transactions On Neural Networks*.8(3): 799-803.

Oh, S.H., and Lee, S-Y. (1999). A New Error Function at Hidden Layers for Fast Training of Multilayer Perceptrons. *IEEE Transactions On Neural Networks*. 10(4): 960-964.

Otair M. A., Salameh W. A. (2006). Efficient training of backpropagation neural networks. *Neural Network World.* 16 (4):291-311.

Pernia-Espinoza, A.V., Joaquin B., Martinez-de-Pison, O-M.F.J., and Gonzalez-Marcos, A.(2005). TAO-Robust Backpropagation Learning Algorithm. *Neural Networks*. 18:191-204.

Rimer, M., and Martinez, T. (2006).CB3: An Adaptive Error Function for Backpropagation Training. *Neural Processing Letters*. 24:81–92

Rumelhart, D.E. and McClelland, J.L. (1986). *Parallel Distributed Processing: Explorations in The Microstructure of Cognition*. Vol 1. MIT press, Cambridge,MA.

Rydvan and Milan. (1999). Biquadratic error functions for the BP-networks. *Neural Network World*. 9(1):17-24**.**

Salem, M. M., Malik, O. P., Zaki, A. M., Mahgoub, O. A., and El-Zahab, E. A. (2000). On-Line Trained Neuro-Controller with a Modified Error Function. *Proceedings, Canadian Conference on Electrical and Computer Engineering,* May 5-7, 2000, Halifax, 83-87.

Saroja. Neural network. Date Accesed:18/12/2007. www.cse.iitd.ernet.in/~saroj/nnet.ppt

Shamsuddin, S.M., Sulaiman, M.N. and Darus, M. (2001). An Improved Error Signal For Bacpropagation Model For Classification Problems. *Intern. J. Computer* Mathematics. 76(1-2): 297-305.

Shamsuddin S. M., Darus M. and Saman. (2007).Three term backpropagation algorithm for classification problem. *Neural Network World* .17 (2007): 363-376

Sridhar Narayan (1997). The generalized sigmoid activation function: Competitive supervised learning. *Information Sciences, 1-2(99). 69-82*

Taji, K., Miyake, T., and Tamura, H.(1999). On error Backpropagation Algorithm Using Absolute Error Function. *IEEE International Conference, IEEE SMC '99 Conference Proceedings, 1999*. 12-15 October 1999. Tokyo, 5(1999):401-406 Verma B.K. and Mulawka J.J. (1994). A Modified Backpropagation Algorithm. IEEE, 840-844

Wang, X.G., Tang, Z., Tamura, H., Ishii, M., and Sun, W.D. (2004).An Improved Backpropagation Algorithm To Avoid The Local Minima Problem. *Neurocomputing*. 56:455 - 460.

Wang, X.G., Tang, Z.,Tamura, H., and Ishii, M.(2004). A modified error function for the backpropagation algorithm. *Neurocomputing*. 57 (2004):477 – 484

Wang, C.H., Kao, C.H. and Lee W.H. (2007).A new interactive model for improving the learning performance of back propagation neural network. *Automation in Construction* . 16(6): 745-758.

Wen, J.W., Zhao, J.L., Luo, S.W., and Han, Z. (2000). The Improvements of BP Neural Network Learning Algorithm. *Proceedings of ICSP2000*. 1647-1649

Widder, D.R., and Fiddy, M.A. (1993). High Performance Learning by Modified Error Backpropagation. *Neural Computer Application*. 1:183-187

Xu, L. (1993). Least Mean Square Error Reconstruction Principle For Self-Organizing Neural-Nets. *Neural Networks.* 6(5): 627-648. - only in html abstact

Yam Y.F. and Chow T.W.S. (1993). Extended backpropagation algorithm. *Electronics Letters*. 29(19), 1701-1702.

Yu, C-C.,and Liu, B-D.(2002). A Backpropagation Algorithm with Adaptive Learning Rate and Momentum Coefficient. *Proceedings of the International Joint Conference on Neural Networks, IJCNN 2002*. May 2002. 2:1218-1223.

Zhiqiang, Z., Zheng, T., GuoFeng, T., Vairappan, C., XuGang, W., and RunQun, X. (2007). An Improved Algorithm for Eleman Neural Network by Adding a Modified Error Function. *Lecture Notes in Computer Science 4492.* Berlin Heidelberg: Springer. 465–473.

Zweiri, Y. H., Whidborne, J. F., Althoefer, K and Seneviratne, L.D. (2002). A new

Three Term backpropagation Algorithm With Convergence Analysis. *Proceedings of the 2002 IEEE International Conference on Robotics &Automation*. May 2002. Washington, DC : IEEE, 3882-3887.

Zweiri, Y. H., Whidborne, J. F., Althoefer, K and Seneviratne, L.D. (2003). A Three-term Backpropagation Algorithm. *Neurocomputing* 50:305-318.

Zweiri, Y. H., Whidborne, J. F., Althoefer, K and Seneviratne, L.D. (2005). Stability Analysis Of A Three-Term Backpropagation Algorithm. *Neural Networks.* 18 (2005) 1341–1347.