THE EFFECT OF ADAPTIVE PARAMETERS ON THE PERFORMANCE OF BACK PROPAGATION

NORHAMREEZA BINTI ABDUL HAMID

A thesis submitted in
fulfillment of the requirement for the award of the
Degree of Master of Information Technology

Faculty of Computer Science and Information Technology
Universiti Tun Hussein Onn Malaysia

APRIL, 2012

In the name of Allah, Most Gracious, Most Compassionate

I praise and thanks to Allah

Special for my beloved father and mother,
Abdul Hamid bin Utoh and Marziah binti Darkasi.

For dearest,
Nor Shamiza, Mohd Amir, Nurfirdaus, Nursuhaibah, Nur Naim and Nurrabiatul Adawiyah.
(Along, Abang Long, Fir, Kakak, Abang Chik and Adik).

For their love, support, enthusiasm, encouragement and motivation.

For my supervisor,
Assoc. Prof. Dr. Nazri bin Mohd. Nawi
For his indescribable help, patience, understanding and supportive.

For all postgraduate members, fellow friends and ummah.

This thesis is dedicated to all of you.

ACKNOWLEDGEMENT

In the name of Allah, the Most Gracious, the Most Merciful. With the deepest sense of gratitude and humility, I praise and thanks to Allah for His blessings uncounted in my life and for His willing, I was able to complete this research successfully. This dissertation would not have been possible without the guidance, help and support of many people contributed and extended their valuable assistance in the preparation and completion of this research. I take this opportunity to express my profound sense of gratitude and respects to all those people.

First and foremost, I would like to express my sincere gratitude to my supervisor, Assoc. Prof. Dr. Nazri bin Mohd Nawi for his supported in the possible way, invaluable guidance, useful advice, patience, understanding and encouragement in guiding me from initial to the final level throughout the accomplishment of this research. His enthusiasm and optimism coupled with knowledge and experience, this evidence really rewarding me. His feedback, editorial comments and suggestions were also invaluable for writing this thesis. I really appreciate it.

In preparing this research, my gratitude is extended to Universiti Tun Hussein Onn Malaysia (UTHM) for supporting this research under the Postgraduate Incentive Research Grant.

A special thanks to my beloved families, for their continuous prayer, encouragement, love, support, patience and care whenever I need during these challenging years. I dedicate this work to all of you. My overwhelming gratitude to all my friends that been together with me, thanks for love, care, concern and support.

Thanks to all staff in Faculty of Computer Science and Information Technology, Software and Multimedia Centre, Centre for Graduate Studies, and Research, Innovation, Commercialization and Consultancy Office for their support, cooperation and contribution all the way. Lastly, it is pleasure to thank all those who have helped either directly or indirectly. Thank you.

ABSTRACT

The Back Propagation algorithm or its variation on Multilayered Feedforward Networks is widely used in many applications. However, this algorithm is well-known to have difficulties with local minima problem particularly caused by neuron saturation in the hidden layer. Most existing approaches modify the learning model in order to add a random factor to the model, which overcomes the tendency to sink into local minima. However, the random perturbations of the search direction and various kinds of stochastic adjustment to the current set of weights are not effective in enabling a network to escape from local minima which cause the network fail to converge to a global minimum within a reasonable number of iterations. Thus, this research proposed a new method known as Back Propagation Gradient Descent with Adaptive Gain, Adaptive Momentum and Adaptive Learning Rate (BPGD-AGAMAL) which modifies the existing Back Propagation Gradient Descent algorithm by adaptively changing the gain, momentum coefficient and learning rate. In this method, each training pattern has its own activation functions of neurons in the hidden layer. The activation functions are adjusted by the adaptation of gain parameters together with adaptive momentum and learning rate value during the learning process. The efficiency of the proposed algorithm is compared with conventional Back Propagation Gradient Descent and Back Propagation Gradient Descent with Adaptive Gain by means of simulation on six benchmark problems namely breast cancer, card, glass, iris, soybean, and thyroid. The results show that the proposed algorithm extensively improves the learning process of conventional Back Propagation algorithm.

ABSTRAK

Algoritma *Back Propagation* atau variasinya pada *Multilayered Feedforward Networks* digunakan secara meluas dalam pelbagai aplikasi. Walau bagaimanapun, algoritma ini terkenal dengan masalah *local minima* yang disebabkan oleh *neuron saturation* dalam *hidden layer*. Kebanyakan pendekatan sedia ada, mengubahsuai model pembelajaran dengan menambah faktor rawak pada model tersebut untuk mengatasi masalah terperangkap pada *local minima*. Walau bagaimanapun, arah pencarian *random perturbations* dan pelbagai jenis *stochastic adjustment* bagi set pemberat semasa tidak efektif untuk menghindari masalah *local minima* yang menyebabkan model tersebut gagal dalam proses pembelajaran pada iterasi tertentu. Justeru itu, kajian ini mencadangkan satu kaedah baru dikenali sebagai *Back Propagation Gradient Descent with Adaptive Gain, Adaptive Momentum and Adaptive Learning Rate* (BPGD-AGAMAL) yang mengubahsuai algoritma *Back Propagation Gradient Descent* sedia ada dengan menukar *gain*, momentum dan *learning rate* secara adaptif. Dalam kaedah ini, setiap corak latihan mempunyai *activation function* tersendiri pada neuron dalam *hidden layer*. *Activation function* dilaraskan dengan penyesuaian parameter *gain* di samping mengubah nilai momentum dan *learning rate* semasa proses pembelajaran. Keberkesanan algoritma yang dicadangkan dibandingkan dengan *Back Propagation Gradient Descent* yang konvensional dan *Back Propagation Gradient Descent with Adaptive Gain* dan disahkan secara simulasi pada enam jenis masalah iaitu *breast cancer, card, glass, iris, soybean, and thyroid*. Hasil keputusan jelas menunjukkan bahawa algoritma yang dicadangkan berkeupayaan meningkatkan proses pembelajaran jika dibandingkan dengan algoritma *Back Propagation* yang konvensional.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ALGORITHMS

LIST OF SYMBOLS AND ABBREVIATIONS

θ_j	-	Bias for the j^{th} unit.
η	-	Learning rate
α	-	Momentum coefficient
$a_{net,j}$	-	Net input activation function for the j^{th} unit.
a_1	-	BPGD
a_2	-	BPGD-AG
a_3	-	BPGD-AGAMAL
c	-	Gain of the activation function
e	-	Exponent
$f(x)$	-	Function of x
f	-	The squashing or activation function of the processing unit
o_i	-	Output of the i^{th} unit
o_j	-	Output of the j^{th} unit
o_k	-	Output of the k^{th} output unit
t_k	-	Desired output of the k^{th} output unit
w_{ij}	-	Weight of the link from unit i to unit j
w_{jk}	-	Weight on the link from node j to k
$x < 0$	-	x is less than 0
$x > 1$	-	x is greater than 1
$x \geq 0$	-	x is greater than or equal to 0
$-1 \leq x \leq 1$	-	x is greater than or equal to -1 and x is less than or equal to 1
A	-	All algorithms

E	-	Error function
H	-	Performance of the BPGD-AGAMAL against BPGD on measuring criteria
J	-	Performance of the BPGD-AGAMAL against BPGD-AG on measuring criteria
K	-	Improvement ratio of the BPGD-AGAMAL against BPGD on measuring criteria
LB	-	Lower bound
M	-	Improvement ratio of the BPGD-AGAMAL against BPGD-AG on measuring criteria
N	-	Performance of BPGD on measuring criteria
Q	-	Performance of BPGD-AG on measuring criteria
R	-	Performance of BPGD on measuring criteria
UB	-	Upper bound
ADALINE	-	Adaptive Linear Element
AI	-	Artificial Intelligence
ANN	-	Artificial Neural Network
BP	-	Back Propagation
BPGD	-	Back Propagation Gradient Descent
BPGD-AG	-	Back Propagation Gradient Descent with Adaptive Gain
BPGD-AGAMAL	-	Back Propagation Gradient Descent with Adaptive Gain, Adaptive Momentum and Adaptive Learning Rate
CPU	-	Central Processing Unit
GD	-	Gradient Descent
IWS	-	Initial Weight Selection
MLFNN	-	Multilayer Feedforward Neural Network
MLP	-	Multilayer Perceptron
MSE	-	Mean Squared Error
NN	-	Neural Network
OBP	-	Optical Back Propagation
RBF	-	Radial Basis Function

SD	-	Standard Deviation
SPLNN	-	Single Layer Perceptron Neural Network
UCIMLR	-	University California Irvine Machine Learning Repository

LIST OF APPENDICES

LIST OF PUBLICATIONS

Journals:

(i) Norhamreeza Abdul Hamid, Nazri Mohd Nawi, Rozaida Ghazali, Mohd Najib Mohd Salleh (2011) "Accelerating Learning Performance of Back Propagation Algorithm by Using Adaptive Gain Together with Adaptive Momentum and Adaptive Learning Rate on Classification Problems." International Journal of Software Engineering and Its Applications. Vol. 5, No. 4, pp. 31-44.

(ii) Norhamreeza Abdul Hamid, Nazri Mohd Nawi, Rozaida Ghazali, Mohd Najib Mohd Salleh (2011) "Improvements of Back Propagation Algorithm Performance by Adaptively Changing Gain, Momentum and Learning Rate." International Journal on New Computer Architectures and Their Applications. Vol. 1, No. 4, pp. 889-901.

(iii) Norhamreeza Abdul Hamid, Nazri Mohd Nawi, Rozaida Ghazali (2011) "The Effect of Adaptive Gain and Momentum in Improving Training Time of Back Propagation algorithm on Classification problems", International Journal on Advanced Science, Engineering And Information Technology, Vol. 1, No. 2, pp. 178-184.

(iv) Nazri Mohd Nawi, Norhamreeza Abdul Hamid (2010) "BPGD-AG: A New Improvement of Back-Propagation Neural Network Learning Algorithms with Adaptive Gain", Journal of Science and Technology UTHM 2010, Vol 2, No. 2, pp. 83-102. ISSN-2229-8460.

Proceedings:

(i) Norhamreeza Abdul Hamid, Nazri Mohd Nawi, Rozaida Ghazali, Mohd Najib Mohd Salleh. “Accelerating Learning Performance of Back Propagation Algorithm by Using Adaptive Gain Together with Adaptive Momentum and Adaptive Learning Rate on Classification Problems”, In T.-h. Kim et al. 13-15 April 2011, UCMA 2011, Part II, CCIS 151, pp. 573-584, 2011. © Springer-Verlag Berlin Heidelberg 2011.

(ii) Norhamreeza Abdul Hamid, Nazri Mohd Nawi, Rozaida Ghazali, Mohd Najib Mohd Salleh. “Learning Efficiency Improvement of Back Propagation Algorithm by Adaptively Changing Gain Parameter together with Momentum and Learning Rate”. In J. M. Zain et al. ICSECS 2011, Part III, CCIS 181, pp. 812-824, 2011. © Springer-Verlag Berlin Heidelberg 2011.

(iii) Norhamreeza Abdul Hamid, Nazri Mohd Nawi, Rozaida Ghazali, Mohd Najib Mohd Salleh. (2011): “Solving Local Minima Problem In Back Propagation Algorithm Using Adaptive Gain, Adaptive Momentum and Adaptive Learning Rate On Classification Problems”, Proceeding in International Conference On Mathematical and Computational Biology 2011 (ICMCB 2011), Renaissance Melaka Hotel, Malacca, 12-14 April 2011.

(iv) Norhamreeza Abdul Hamid, Nazri Mohd Nawi, Rozaida Ghazali, Mohd Najib Mohd Salleh. “A Review on Back Propagation Algorithm.” Proceeding in The Second World Conference on Information Technology 2011 (WCIT 2011), Antalya, Turkey, 23-27 November 2011.

(v) Norhamreeza Abdul Hamid, Nazri Mohd Nawi, Rozaida Ghazali, Mohd Najib Mohd Salleh. “Accelerating Learning Performance of Back Propagation Algorithm by Using Adaptive Gain Together with Adaptive Momentum and Adaptive Learning Rate on Classification Problems.” Proceeding in The International Conference on Advanced Science, Engineering and Information Technology 2011 (ICASEIT 2011), pp. 178-184, Kuala Lumpur, 14-15 January 2011.

(vi) Norhamreeza Abdul Hamid, Nazri Mohd Nawi, Rozaida Ghazali. “The Effect of Adaptive Gain and Adaptive Momentum in Improving Training Time of Gradient Descent Back-Propagation Algorithm on Approximation Problem”. Proceeding in 4th International Conference on Postgraduate Education (ICPE-4 2010), pp. 463-466, Kuala Lumpur, 26-28 November 2010.

(vii) Norhamreeza Abdul Hamid, Nazri Mohd Nawi “The Effect of Adaptive Gain and Momentum in Improving Training Time of Back Propagation Algorithm on Approximation Problem.” Proceeding in Conference on Postgraduates Incentive Research Grant 2010 (CoGIS 2010), 15 July 2010.

(viii) Norhamreeza Abdul Hamid, Nazri Mohd Nawi (2009): ” The Effect of Gain Variation of Activation Function in Improving Training Time of Back Propagation Neural Network on Classification Problems, Proceeding in Kolokium Kebangsaan Pasca Siswazah Sains dan Matematik 2009 (KOLUPSI '09), UPSI, 21 December 2009.

(ix) Nazri Mohd Nawi, R.S. Ransing, Mohd Najib Mohd Salleh, Rozaida Ghazali, Norhamreeza Abdul Hamid: “The Effect of Gain Variation in Improving Learning Speed of Back propagation Neural Network Algorithm on Classification Problems”. Proceeding in Symposium on Progress in Information and Communication Technologies (SPICT'09), Malaysia, 7-8 December 2009.

LIST OF AWARDS

(i) **Bronze Medal in Malaysia International Technology Expo [MiTE 2010]:** Nazri Mohd Nawi, Norhamreeza Abdul Hamid, Rozaida Ghazali, Mohd Najib Mohd Salleh. "BPGD-AG: The New Improved Back Propagation Algorithm."

CHAPTER 1

INTRODUCTION

1.1 An Overview

The Artificial Neural Network (ANN) is an Artificial Intelligence (AI) methodology using computational models with architecture and operations is inspired by human knowledge on biological nervous systems, particularly the brain, to process information. This distribution of knowledge provides a property of fault tolerance and potential for massive parallel implementation (Haykin, 2009).

Over the years, the acceptance level in the applications of ANN has been growing because it is proficient in capturing process information in a black box mode. Due to its ability to solve problems with relative ease of use, robustness to noisy input data and execution speed, and due its ability to analyse complicated systems without accurate modelling in advance, ANN has successfully been implemented across an extraordinary range of problem domains, in areas as diverse as pattern recognition and classification (Nazri *et al.*, 2010b), signal and image processing (Sabeti *et al.*, 2010), robot control (Subudhi & Morris, 2009), weather prediction (Mandal *et al.*, 2009), financial forecasting (Yu *et al.*, 2009), and medical diagnosis (Nazri *et al.*, 2010a).

The Multilayer Perceptron (MLP) is a well-known and the most frequently used type of ANN (Popescu *et al.*, 2009). It is suitable for a large variety of applications (Fung *et al.*, 2005). A standard MLP consists of an input layer, one or more hidden layer(s), and an output layer. Every node in a layer, it is connected to other node in the adjacent forward layer where each connection has a weight associated with it.

Learning is a basic and essential characteristic of MLP. Learning refers to the ability to learn from experience through network examples, to generalise the captured knowledge for expectation solutions, and to self-update in order to improve its performance. During the learning phase, the network learns by adjusting the weights so it is able to predict the correct class of the input samples (Han & Kamber, 2006).

The ANN uses Back Propagation (BP) algorithm to perform parallel training to improve the efficiency of MLP's network. The BP algorithm is the most popular, effective, and easiest algorithm to produce a model for MLP's complex network. This algorithm has produced a large class of network types with many diverse topologies and training methods. The BP algorithm is a supervised learning method that involves backward error correction of the network weights. This algorithm uses a gradient descent (GD) method that attempts to minimise the error of the network by moving down the gradient of the error curve (Alsmadi *et al.*, 2009). The weights of the network are adjusted by the algorithm. Consequently, the error is reduced along a descent direction.

Although BP algorithm has been successfully applied to a wide range of practical problems (Haofei *et al.*, 2007; Lee *et al.*, 2005), it has some limitations. Since BP algorithm uses GD method, the problems include slow learning convergence and easy to get trapped at local minima (Bi *et al.*, 2005; Otair & Salameh, 2005). Moreover, the convergence behaviour of the BP algorithm depends on the selection of network topology, initial weights and biases, learning rate, momentum coefficient, activation function, and value for the gain in the activation function.

In the last decade, a significant number of methods have been produced to improve the efficiency and convergence rate (Kathirvalavakumar & Thangavel, 2006; Naimin *et al.*, 2006; Nazri *et al.*, 2010b; Nazri *et al.*, 2008; Otair & Salameh, 2005). Those studies showed that the BP performance was affected by many factors, for instances learning structure, initial weight, learning rate, momentum coefficient, and activation function.

1.2 Problem Statements

The BP algorithm is well-known for its extraordinary ability to derive meaning from complicated or imprecise data that are too complex to be noticed by either humans or other computer techniques. In some practical applications of BP, fast response to external events within an extremely short time are highly insisted and expected. However, the extensively used GD method clearly cannot satisfy large scale applications and when higher learning performances are required. Furthermore, this type of algorithm has the uncertainty in finding the global minimum of the error criterion functions. To overcome those problems, a research has been done to improve the training efficiency of conventional BP algorithm by introducing adaptive gain variation of activation function known as Back Propagation Gradient Descent With Adaptive Gain (BPGD-AG) proposed by Nazri *et al.* (2008). It has been proven that the performances of the proposed method (BPGD-AG) are better than the conventional BP.

Although the analysis results shown by Nazri *et al.* (2008) demonstrated that the method significantly increased the learning speed and outperformed the standard algorithm with constant gain in learning the target function, however during the training, it was noticed that the method only updated weights, bias and gain update expressions adaptively whereas the learning rate and momentum term were keep constant until the end of the training. The challenge of this research was to prove by simulations, that the adaptive momentum and adaptive learning rate also have the significant effects in improving the current working BPGD-AG algorithm on some classification problems.

REFERENCES

Ackley, D. H., Hinton, G. E., & Sejnowski, T. J. (1986). A learning algorithm for boltzmann machines. *Cognitive Science, 9 (1)*, pp. 147-169.

Alsmadi, M. K. S., Omar, K., & Noah, S. A. (2009). Back Propagation Algorithm: The Best Algorithm Among the Multi-layer Perceptron Algorithm. *International Journal of Computer Science and Network Security, 9 (4),* pp. 378-383.

Amari, S. (1967). A Theory of Adaptive Pattern Classifiers. *Electronic Computers, IEEE Transactions on, EC-16 (3),* pp. 299-307.

Anderson, J. A. (1972). A simple neural network generating an interactive memory. *Mathematical Biosciences, 14 (3-4),* pp. 197-220.

Bi, W., Wang, X., Tang, Z., & Tamura, H. (2005). Avoiding the Local Minima Problem in Backpropagation Algorithm with Modified Error Function. *IEICE Trans. Fundam. Electron. Commun. Comput. Sci., E88-A (12),* pp. 3645-3653.

Broomhead, D. S., & Lowe, D. (1988). Multivariable Functional Interpolation and Adaptive Networks. *Complex Systems 2,* pp. 321-355.

Castellani, M., & Rowlands, H. (2009). Evolutionary Artificial Neural Network Design and Training for Wood Veneer Classification. *Engineering Applications of Artificial Intelligence, 22 (4-5)*, pp. 732-741.

Coomans, D., Broeckaert, I., Jonckheer, M., & Massart, D. L. (1983). Comparison of Multivariate Discrimination Techniques for Clinical Data - Application to The Thyroid Functional State. *Methods of Information Medicine, 22 (2),* pp. 93 - 101.

Cybenko, G. (1989). Approximation by Superposition of a Sigmoidal Function. *Mathematics of Control, Signals and Systems, 2,* pp. 303-314.

Elhag, T. M. S., & Wang, Y.-M. (2007). Risk Assessment for Bridge Maintenance Projects: Neural Networks versus Regression Techniques. *Journal of Computing in Civil Engineering, 21 (6)*, pp. 769-775.

Engelbrecht, A. P. (2007). *Computational Intelligence: An Introduction*. 2nd ed. England: John Wiley & Sons.

Eom, K., Jung, K., & Sirisena, H. (2003). Performance improvement of backpropagation algorithm by automatic activation function gain tuning using fuzzy logic. *Neurocomputing, 50*, pp. 439-460.

Evett, I. W., & Spiehler, E. J. (1988). Rule induction in forensic science. in (Ed.). *Knowledge Based Systems*. Halsted Press. pp. 152-160.

Fisher, R. A. (1936). The use of multiple measurements in taxonomic problems. *Annals of Eugenics, 7 (2)*, pp. 179-188.

Frank, A., & Asuncion, A. (2010). *UCI Machine Learning Repository*. Retrieved on January 10, 2010 from http://archive.ics.uci.edu/ml.

Funahashi, K.-I. (1989). On the approximate realization of continuous mappings by neural networks. *Neural Networks, 2 (3)*, pp. 183-192.

Fung, C. C., Iyer, V., Brown, W., & Wong, K. W. (2005). Comparing the Performance of Different Neural Networks Architectures for the Prediction of Mineral Prospectivity. *Proceedings of the Proceedings of 2005 International Conference on Machine Learning and Cybernetics, 2005*. pp. 394-398.

Grossberg, S. (1976). Adaptive pattern classification and universal recoding: I. Parallel development and coding of neural feature detectors. *Biological Cybernetics, 23 (3)*, pp. 121-134.

Han, J., & Kamber, M. (2006). *Data Mining: Concepts and Techniques*. Morgan Kaufmann.

Haofei, Z., Guoping, X., Fangting, Y., & Han, Y. (2007). A neural network model based on the multi-stage optimization approach for short-term food price forecasting in China. *Expert Syst. Appl., 33 (2)*, pp. 347-356.

Haykin, S. S. (2009). *Neural Networks and Learning Machines*. New Jersey: Prentice Hall.

Hebb, D. (1949). *The Organization of Behavior: A Neuropsychological Theory*. Wiley.

Hopfield, J. J. (1982). Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Sciences, 79 (8)*, pp. 2554-2558.

Hornik, K., Stinchcombe, M., & White, H. (1989). Multilayer feedforward networks are universal approximators. *Neural Netw., 2 (5)*, pp. 359-366.

Hyder, M. M., Shahid, M. I., Kashem, M. A., & Islam, M. S. (2009). Initial Weight Determination of a MLP for Faster Convergence. *Journal of Electronics and Computer Science, 10*.

Jayalakshmi, T., & Santhakumaran, A. (2010). Improved Gradient Descent Back Propagation Neural Networks for Diagnoses of Type II Diabetes Mellitus. *Global Journal of Computer Science and Technology, 9 (5)*, pp. 94-97.

Kathirvalavakumar, T., & Thangavel, P. (2006). A Modified Backpropagation Training Algorithm for Feedforward Neural Networks*. *Neural Processing Letters, 23 (2)*, pp. 111-119.

Kaya, A. (2009). Slope Residual Strength Evaluation using Artificial Neural Networks. *Geotech Geol. Eng., 27 (2)*, pp. 281-289.

Kohonen, T. (1972). Correlation Matrix Memories. *Computers, IEEE Transactions on, C-21 (4)*, pp. 353-359.

Kohonen, T. (1982). Self-organized formation of topologically correct feature maps. *Biological Cybernetics, 43 (1)*, pp. 59-69.

Kolen, J. F., & Pollack, J. B. (1991). Back Propagation Is Sensitive To Initial Conditions. in R. P. Lippmann, J. E. Moody & D. S. Tpuretzky (Ed.). *Advances in Neural Information Processing Systems*. Denver: pp. 860-867.

Köppen, M., Kasabov, N., Coghill, G., Adam, S., Karras, D., & Vrahatis, M. (2009). Revisiting the Problem of Weight Initialization for Multi-Layer Perceptrons Trained with Back Propagation. in (Ed.). *Advances in Neuro-Information Processing*. Springer Berlin / Heidelberg. pp. 308-315.

Kosko, B. (1987). Adaptive bidirectional associative memories. *Appl. Opt., 26 (23)*, pp. 4947-4960.

LeCun, Y. (1985). A Learning Scheme for Asymmetric Threshold Networks. *Proceedings of the Cognitiva*. Paris, France: pp. 599-604.

Lee, C.-H., & Lin, Y.-C. (2005). An adaptive neuro-fuzzy filter design via periodic fuzzy neural network. *Signal Process., 85 (2)*, pp. 401-411.

Lee, K., Booth, D., & Alam, P. (2005). A comparison of supervised and unsupervised neural networks in predicting bankruptcy of Korean firms. *Expert Systems with Applications, 29 (1)*, pp. 1-16.

Li, S., Okada, T., Chen, X., & Tang, Z. (2006). An Individual Adaptive Gain Parameter Backpropagation Algorithm for Complex-Valued Neural Networks. in J. Wang, Z. Yi, J. Zurada, B.-L. Lu & H. Yin (Ed.). *Advances in Neural Networks - ISNN 2006*. Springer Berlin / Heidelberg. pp. 551-557.

Maier, H. R., & Dandy, G. C. (1998). The effect of internal parameters and geometry on the performance of back-propagation neural networks: an empirical study. *Environmental Modelling and Software, 13 (2)*, pp. 193-209.

Mandal, S., Sivaprasad, P. V., Venugopal, S., & Murthy, K. P. N. (2009). Artificial neural network modeling to evaluate and predict the deformation behavior of stainless steel type AISI 304L during hot torsion. *Applied Soft Computing, 9 (1)*, pp. 237-244.

Mangasarian, O. L., & Wolberg, W. H. (1990). Cancer diagnosis via linear programming. *SIAM News, 23 (5)*, pp. 1-18.

McCulloch, W., & Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *Bulletin of Mathematical Biology, 5 (4)*, pp. 115-133.

Michalski, R. S., & Chilausky, R. L. (1980). Learning by Being Told and Learning from Examples: An Experimental Comparison of the Two Methods of Knowledge Acquisition in the Context of Developing an Expert System for Soybean Disease Diagnosis. *International Journal of Policy Analysis and Information Systems, 4:2*.

Minsky, M. L., & Papert, S. (1969). *Perceptrons; An Introduction to Computational Geometry*. MIT Press.

Naimin, Z., Wei, W., & Gaofeng, Z. (2006). Convergence of gradient method with momentum for two-Layer feedforward neural networks. *Neural Networks, IEEE Transactions on, 17 (2)*, pp. 522-525.

Nazri, M. N. (2007). *Computational Issues in Process Optimsation using Historical Data*. S. University: PhD.

Nazri, M. N., Najib, M. S. M., & Rozaida, G. (2010a). The Development of Improved Back-Propagation Neural Networks Algorithm for Predicting Patients with Heart Disease. in R. Zhu, Y. Zhang, B. Liu & C. Liu (Ed.). *Information Computing and Applications*. Springer Berlin / Heidelberg. pp. 317-324.

Nazri, M. N., Ransing, R. S., Najib, M. S. M., Rozaida, G., & Norhamreeza, A. H. (2010b). An Improved Back Propagation Neural Network Algorithm on Classification Problems. in Y. Zhang, A. Cuzzocrea, J. Ma, K.-i. Chung, T. Arslan & X. Song (Ed.). *Database Theory and Application, Bio-Science and Bio-Technology*. Springer Berlin Heidelberg. pp. 177-188.

Nazri, M. N., Ransing, R. S., & Ransing, M. S. (2008). An Improved Conjugate Gradient Based Learning Algorithm for Back Propagation Neural Networks. *International Journal of Information and Mathematical Sciences, 4 (1),* pp. 46-55.

Negnevitsky, M. (2005). *Artificial Intelligence a Guide to Intelligent Systems.* 2. Harlow, England: Addison Wesley.

Ng, W. W. Y., Yeung, D. S., & Tsang, E. C. C. (2006). Pilot Study on the Localized Generalization Error Model for Single Layer Perceptron Neural Network. *Proceedings of the Machine Learning and Cybernetics, 2006 International Conference on*. pp. 3078-3082.

Norhamreeza, A. H., & Nazri, M. N. (2009). The Effect of Gain Variation of Activation Function in Improving Training Time of Back Propagation Neural Network on Classification Problems. *Proceedings of the Kolokium Kebangsaan Pasca Siswazah Sains dan Matematik 2009*. UPSI: pp.

Norhamreeza, A. H., Nazri, M. N., & Ghazali, R. (2011). The Effect of Adaptive Gain and Adaptive Momentum in Improving Training Time of Gradient Descent Back Propagation Algorithm on Classification problems. *Proceedings of the Proceeding of the International Conference on Advanced Science, Engineering and Information Technology 2011*. Hotel Equatorial Bangi-Putrajaya, Malaysia: pp. 178-184.

Otair, M. A., & Salameh, W. A. (2005). Speeding Up Back-Propagation Neural Networks. *Proceedings of the Proceeding of the 2005 Informing Science and IT Education Joint Conference*. Flagstaff, Arizona, USA: pp. 167-173.

Ozel, T., Correia, A. E., & Davim, J. P. (2009). Neural Network Process Modelling for Turning of Steel Parts using Conventional and Wiper Inserts *International Journal of Materials and Product Technology 2009, 35 (1/2)*, pp. 246-258

Parker, D. (1985). *Learning logic*. MIT: Technical Report TR-87.

Popescu, M.-C., Balas, V. E., Perescu-Popescu, L., & Mastorakis, N. (2009). Multilayer perceptron and neural networks. *WSEAS Trans. Cir. and Sys., 8 (7)*, pp. 579-588.

Quinlan, J. R. (1987). Simplifying decision trees. *Int. J. Man-Mach. Stud., 27 (3)*, pp. 221-234.

Quinlan, J. R. (1993). *C4.5: Programs for Machine Learning (Morgan Kaufmann Series in Machine Learning)*. Morgan Kaufmann.

Rosenblatt, F. (1958). The perception: a probabilistic model for information storage and organization in the brain. *Psychological Review, 65 (6)*, pp. 386-408.

Rumelhart, D. E., & McClelland, J. L. (1986). *Parallel Distributed Processing: Explorations in the Microstructure of Cognition, Vol. 1: Foundations*. MIT Press.

Sabeti, V., Samavi, S., Mahdavi, M., & Shirani, S. (2010). Steganalysis and payload estimation of embedding in pixel differences using neural networks. *Pattern Recogn., 43 (1)*, pp. 405-415.

Shepherd, A. J. (1997). *Second-Order Methods for Neural Networks: Fast and Reliable Training Methods for Multi-Layer Perceptrons*. London: Springer.

Subudhi, B., & Morris, A. S. (2009). Soft computing methods applied to the control of a flexible robot manipulator. *Applied Soft Computing, 9 (1)*, pp. 149-158.

Sun, Y.-j., Zhang, S., Miao, C.-x., & Li, J.-m. (2007). Improved BP Neural Network for Transformer Fault Diagnosis. *Journal of China University of Mining and Technology, 17 (1)*, pp. 138-142.

Thimm, G., Moerland, P., & Fiesler, E. (1996). The interchangeability of learning rate and gain in backpropagation neural networks. *Neural Comput., 8 (2)*, pp. 451-460.

Wang, X. G., Tang, Z., Tamura, H., Ishii, M., & Sun, W. D. (2004). An improved backpropagation algorithm to avoid the local minima problem. *Neurocomputing, 56* pp. 455-460.

Watkins, D. (1997). *Clementine's Neural Networks Technical Overview*. Technical Report.

Werbos, P. (1974). *Beyond regression: new tools for prediction and analysis in the behavioral sciences*. Harvard University.

Xiaoyuan, L., Bin, Q., & Lu, W. (2009). A New Improved BP Neural Network Algorithm. *Proceedings of the Proceedings of the 2009 Second International Conference on Intelligent Computation Technology and Automation - Volume 01*. IEEE Computer Society.

Ye, Y. C. (2001). *Application and Practice of the Neural Networks*. Taiwan: Scholars Publication.

Yu, L., Wang, S., & Lai, K. K. (2009). A neural-network-based nonlinear metamodeling approach to financial time series forecasting. *Appl. Soft Comput., 9 (2)*, pp. 563-574.

Yuemei, X., & Hong, Z. (2009). Study on the Improved BP Algorithm and Application. *Proceedings of the Information Processing, 2009. APCIP 2009. Asia-Pacific Conference on*. pp. 7-10.

Zweiri, Y. H. (2007). Optimization of a Three-Term Back Propagation Algorithm Used for Neural Network Learning. *International Journal of Computer Intelligence, 3 (4)*, pp. 322-327.

Zweiri, Y. H., Whidborne, J. F., & Seneviratne, L. D. (2003). A three-term backpropagation algorithm. *Neurocomputing, 50* pp. 305-318.