TIME SERIES MODELING AND DESIGNING OF ARTIFICIAL NEURAL NETWORK (ANN) FOR REVENUE FORECASTING

NORFADZLIA BINTI MOHD YUSOF

UNIVERSITI TEKNOLOGI MALAYSIA

TIME SERIES MODELING AND DESIGNING OF ARTIFICAL NEURAL NETWORK (ANN) FOR REVENUE FORECASTING

NORFADZLIA BINTI MOHD YUSOF

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

JUNE 2005

To my beloved and supportive family

ACKNOWLEDMENT

I would like to express the deepest appreciation to my major advisor, Assoc. Prof. Dr. Siti Mariyam Binti Shamsuddin for her direction, assistance, and guidance in the accomplishment of this project report. Words are inadequate to express my thanks to Dr. Siti Mariyam. I have yet to see the limits of her wisdom, patience, and selfless concern for her students. In particular, Dr. Siti Mariyam's recommendations and suggestions have been invaluable for the study and for research improvement. Many thank to my co-supervisor, Puan Roselina Binti Sallehuddin for her enthusiasm in sharing her knowledge and expertise. Without their guidance and persistent help this dissertation would not have been possible.

Many thanks to Encik Zaily Bin Ayub and Tuan Haji Md. Shah whom have always been very generous with their time, and they puts a lot more effort than necessary into teaching me about the data analysis and supplied me with the revenue data. I would also like to express thanks to my examiners for their useful comments.

Special thanks to all my family members (abah, mama, abang, adik, eleen, tok and paksu) for giving me their constant encouragement, strength, support and love I needed to complete my goals. I am very thankful to my beloved twin's sister Norfadzila Binti Mohd Yusof, who always encouraged me to do my best. I also would like to express my gratitude to the lecturers who made my experience in graduate school worthwhile.

Lastly, I would like to dedicate this project report to the memory of my beloved grandfather that passed away on 15th May 2006, as I was nearing completion of this project report. Memory nourishes the heart, and grief abates.

ABSTRAK

Rangkaian neural telah mendapat perhatian dalam teori unjuran. Walaubagaimanapun, pemilihan pelbagai parameter untuk membina model unjuran rangkaian neural bermakna proses merekanya masih melibatkan kaedah cuba jaya. Objektif kajian ini ialah untuk menyiasat kesan pengaplikasian pelbagai nilai nod input, fungsi-fungsi penggiatan dan teknik-teknik transformasi data ke atas pelaksanaan unjuran siri masa pungutan hasil oleh rangkaian neural rambatan balik. Dalam kajian ini, beberapa teknik transformasi data telah digunakan untuk mengeluarkan komponen tidak pegun di dalam siri masa data, dan kesannya ke atas proses pembelajaran dan menghasilkan unjuran menggunakan model rangkaian neural dianalisa. Kaedah cuba jaya digunakan di dalam kajian ini untuk mendapatkan jumlah nod input yang sesuai begitu juga dengan jumlah nod terselindungnya yang ditentukan melalui teorem Kolmogorov. Kajian ini juga menumpukan kepada perbandingan penggunaan fungsi logaritma dan model rangkaian neural cadangan yang menggabungkan fungsi sigmoid di nod lapisan terselindung dan fungsi logaritma di lapisan nod output, dengan fungsi sigmoid sebagai fungsi penggiatan pada nod-nod tersebut. Kaedah eksperimen validasi-silang digunakan dalam kajian ini untuk meningkatkan keupayaan pengitlakan rangkaian neural. Hasil kajian ini menunjukkan bahawa model rangkaian neural yang mempunyai jumlah nod input yang kecil sejajar dengan saiz struktur rangkaiannya yang kecil berjaya menjana unjuran yang baik walaupun ia sengsara daripada penumpuan yang lambat. Fungsi penggiatan sigmoid mengurangkan kekompleksan rangkaian neural serta menjanakan penumpuan terpantas dan keupayaan unjuran yang baik di dalam hampir keseluruhan eksperimen. Kajian ini juga menunjukkan prestasi unjuran daripada rangkaian neural dapat ditingkatkan dengan mengunakan teknik transformasi data yang sesuai.

ABSTRACT

Artificial neural networks (ANN) have found increasing consideration in forecasting theory. However, the large numbers of parameters that must be selected to develop ANN forecasting model have meant that the design process still involves much trial and error. The objective of this study is to investigate the effect of applying different number of input nodes, activation functions and pre-processing techniques on the performance of backpropagation (BP) network in time series revenue forecasting. In this study, several pre-processing techniques are presented to remove the non-stationary in the time series and their effect on ANN model learning and forecast performance are analyzed. Trial and error approach is used to find the sufficient number of input nodes as well as their corresponding number of hidden nodes which obtain using Kolmogorov theorem. This study compares the used of logarithmic function and new proposed ANN model which combines sigmoid function in hidden layer and logarithmic function in output layer, with the standard sigmoid function as the activation function in the nodes. A cross-validation experiment is employed to improve the generalization ability of ANN model. From the empirical findings, it shows that an ANN model which consists of small number of input nodes and smaller corresponding network structure produces accurate forecast result although it suffers from slow convergence. Sigmoid activation function decreases the complexity of ANN and generates fastest convergence and good forecast ability in most cases in this study. This study also shows that the forecasting performance of ANN model can considerably improve by selecting an appropriate pre-processing technique.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS

W_{ij}	-	Weight connected between node *i* and *j*
W_{jk}	-	Weight connected between node *j* and *k*
θ_j	-	Bias of node *j*
θ_k	-	Bias of node *k*
net_j	-	Summation of sum of input for node *j* multiplied by the corresponding weights between node *i* and node *j* and bias *j*
net_k	-	Summation of sum of input of node *k* multiplied by the corresponding weights between node *j* and node *k* and bias *k*
o_i	-	Output of node *i*
o_j	-	Output of node *j*
o_k	-	Output of node *k*
t_j	-	Target output value at node *j*
t_k	-	Target output value at node *k*
$W_{ij}(t)$	-	Weight from node *i* to node *j* at time t
$W_{jk}(t)$	-	Weight from node *j* to node *k* at time t
ΔW_{ij}	-	Weight adjustment between node *i* and *j*
ΔW_{jk}	-	Weight adjustment between node *j* and *k*
η	-	Learning rate
α	-	Momentum term
δ_j	-	Error signal at node *j*
δ_k	-	Error signal at node *k*

LIST OF ABBREVIATIONS

ANN	-	Artificial Neural Network
BP	-	Backpropagation
MAPE	-	Mean Absolute Percentage Error
MLP	-	Multilayer Perceptron
MSE	-	Mean Squared Error
NN	-	Neural Network
RMSE	-	Root Mean Squared Error

LIST OF TERMINOLOGIES

Backpropagation	-	A supervised learning method in which an output error signal is fed back through the network, altering connection weights so as to minimize error.
Connection	-	A link between nodes used to pass data from one node to the other. Each connection has an adjustable value called a weight.
Generalization	-	A neural network's ability to respond correctly to data not used to train it.
Global minima	-	The unique point of least error during gradient descent, metaphorically the true "bottom" of the error surface.
Gradient descent	-	A learning process that changes a neural networks weights to follow the steepest path toward the point of minimum error.
Input layer	-	A layer of nodes that forms a passive conduit of data entering a neural network.
Hidden layer	-	A layer of nodes not directly connected to a neural networks input or output.
Local minima	-	A point of regionally low error during gradient descent, a metaphorical dent in the error surface.
Over-fitting	-	A neural network's tendency to learn random details in the training data in addition to the underlying function. A network that simply memorizes may fail to generalize.
Node	-	A single neuron-like element in a neural network.
Output layer	-	The layer of nodes that produce the neural network result.
Supervised learning	-	A learning process requiring a labeled training set.

Target output	-	A “correct” result included with each input pattern in a training or testing data set.
Testing	-	A process for measuring a neural network’s performance, during which the network’s passes through an independent data set to calculate a performance index. It does not change its weights.
Training	-	A process during which a neural network passes through a data set repeatedly, changing the values of its weights to improve its performance.
Weight	-	An adjustable value associated with a connection between nodes in a neural network.

LIST OF APPENDICES

CHAPTER 1

INTRODUCTION

1.1 Introduction

Knowing future better has attracted many people for thousands of years. The forecasting methods vary greatly and will depend on the data availability, the quality of models available, and the kinds of assumptions made, amongst other things. Generally speaking, forecasting is not an easy task and therefore it has attracted many researchers to explore it.

Artificial neural network (ANN) has found increasing consideration in forecasting theory, leading to successful applications in various forecasting domains including economic (Yao, 2002), business (Crone *et al.*, 2004), financial (Lam, 2004), and many more. ANN can learn from examples (pass data), recognize a hidden pattern in historical observations and use them to forecast future values. In addition to that, they are able to deal with incomplete information or noisy data and can be very effective especially in situations where it is not possible to define the rules or steps that lead to the solution of a problem.

Despite of many satisfactory characteristics (Zhang *et al*, 1998), of ANNs, building an ANN model for a particular forecasting problem is a nontrivial task. Several authors such as Bacha and Meyer (1992), Tan and Witting (1993), Kaastra and Boyd (1996), Zhang *et al*. (1998), Plummer (2000), Xu and Chen (2001), Lam (2004) have provided an insight on issues in developing ANN model for forecasting. These modeling issues must be considered carefully because it may affect the performance of ANNs. Based on their studies, some of the discussed modeling issues in constructing ANN forecasting model are the selection of network architecture, learning parameters and data pre-processing techniques applies to the time series data.

Thus, this study uses a BP network to forecast future revenue collection for Royal Malaysian Customs Department. This study examines the effect of network parameters through trial and error approach by varying network structures based on the number of input nodes, activation functions and data pre-processing in designing of BP network forecasting model.

1.2 Problem Background

ANNs promise attractive feature to various forecasting domains: being a data driven learning machine as opposed to conventional model-based approaches, permitting universal approximation of arbitrary linear or non-linear functions, and therefore offering great flexibility in learning the generator of noisy data from examples and generalizing structure from it without priori assumptions (Zhang *et al.*, 1998). However, the nontrivial task of modeling ANNs for a particular prediction problem is still considered to be as much an art as a science (Zhang *et al.*, 1998), as the combination of choices may significantly impact on the networks ability to extrapolate results.

Due to their flexibility, neural networks lack a systematic procedure for model building. Therefore obtaining a reliable neural model involves selecting a large number of parameters experimentally through trial and error (Kaastra and Boyd, 1996). The performance of ANNs in forecasting is influenced by ANN modeling, that is the selection of the most relevant network architecture and network design. Poor selection of parameter settings can lead to slow convergence and incorrect output (Kong and Martin, 1995). One critical decision is to determine the appropriate network architecture, that is, the number of layers, the number of nodes in this layer, and the number of arcs which interconnect with the nodes. The network design decision include the selection of activation function in the hidden and output neurons, the training algorithm, data transformation or normalization method, training and test set, and performance measures.

When applying an ANN model in a real application, attention should be taken in every single step. The usage and training of ANN is an art. One successful experiment says nothing for real application. Zhang (1992), Tan and Witting (1993), Kong and Martin (1995), had focused their studies on parametric effects on building a BP network for a particular forecasting problem. Kaastra and Boyd, (1996), have provided a practical introductory guide in the design of ANN for forecasting financial and economic time series data. The issues on modeling fully-connected feed-forward networks for forecasting had been discussed by Zhang *et al.* (1998). Maier and Dandy (2000), have review the modeling issues and outlined the steps that should be followed in developing ANN model for predicting and forecasting water resource variables.

When BP algorithm was introduced in 1986, there has been much development in the use of ANNs for forecasting by a number of researchers for examples Zhang (1992), Tan and Witting (1993), Yu and Chen (1993), Kong and Martin (1995), Yu (1999), Lopes *et al.* (2000), More and Deo (2003), Crone *et al.* (2004).

BP network is characterized by its robustness, its ability to generalize, learn and to be trained (Kong and Martin, 1995). Although this algorithm is widely used and recognized as a powerful tool for training feed-forward neural network, it suffers from slow convergence process, or long training time (Nguyen *et al.*, 1999, Bilski, 2000, Xu and Chen, 2001, Kamruzzaman and Aziz, 2002, Kodogiannis and Anagnostakis, 2002). One of the identified reason of slow convergence is the used of sigmoid activation function in hidden and output layer of BP network (Bilski, 2000, Kamruzzaman and Aziz, 2002). A number of researches have been done in order to improve the convergence rate of BP learning. Therefore, several approaches have been developed in order to speed up the convergence.

Fnaiech *et al.* (2002) have summarized the approaches for increasing the BP convergence speed onto seven cases: the weight updating procedure, the choice of optimization criterion, the use of adaptive parameters, estimation of optimal initial conditions, reducing the size of problem, estimation of optimal ANN structure and application of more advanced algorithms. According to Kamruzzaman and Aziz (2002), approaches to accelerate BP learning include, selection of better cost function, dynamics variation of learning rate and momentum and selection of better activation function of the neurons. Bilski (2000); Kamruzzaman and Aziz (2002), have proposed new activation functions in order to accelerate BP learning process.

Several other approaches also have been implemented in forecasting problem. For example, Xu and Chen (2001), have employed the fast convergence algorithm the quasi-Newton method to expedite the training process in short-term load forecasting problem. Another work by Kodogiannis and Anagnostakis (2002), have adopted the adaptive learning rate BP network, which relates the learning rate with the total error function in order to accelerate the convergence speed of standard BP in short-term load forecasting.

This research attempts to design an ANN model for revenue time series forecasting. A BP network forecasting model is constructed to test its forecasting capability by implementing modification on several network parameters. The modeling issues discussed in this study are focused on network paradigms (specifically to determine sufficient number of input node and activation function) and data pre-processing. The learning ability and forecast result produce by ANN models are evaluated and examined.

1.3 Problem Statement

The problem statement of this study is as follow:

How the selection of these parameters in network modeling namely: number of input nodes, activation functions and data pre-processing techniques may affect the forecasting capability of ANN in time series revenue forecasting?

1.4 Study Aims

This study aims to provide a step by step methodology for designing ANN for revenue time series forecasting. This research also attempts to explore:

a) the effectiveness of data pre-processing technique on ANN modeling and forecasting performance.
b) the generalization capability of the ANN by varying the network structure.

1.5 Objectives

1. To design and develop ANN model which combines sigmoid activation function in hidden layer and logarithmic activation function in output layer.

2. This research attempts to understand the network parameters by varying them and observing their effect on the network. Specifically, the parametric effect of varying the:
   - data pre-processing technique
   - number of nodes in the input layer of ANN model
   - activation function in hidden and output layers of ANN model

   are monitored in an attempt to develop an understanding of their effect on building a revenue forecasting model.

1.6 Scopes of Study

The scopes of this study are as follow:

1. Real time series data of monthly revenue collection obtained from Royal Malaysian Customs Department in Putrajaya from January 1990 to December 2004 are used as input to the ANN model.

2. The MLP network with three layers (one hidden layer, an input and an output layer) are used.

3. Trial and error design procedures are employed to arrive at an acceptable structure and parameter namely: data pre-processing technique, number of input node and activation function of ANN model.

4. Different data pre-processing techniques are presented to deal with irregularity components exist in time series data and their properties are evaluated by performing one step-ahead revenue forecasting using neural network.

5. The network input nodes are varied from 1 to 11 nodes to see its effect onto the network while the number of hidden node is obtained by using Kolmogorov theorem.

6. The output of the network is the forecast of one-step-ahead revenue collection.

7. Activation functions used for observation and combination are the sigmoid function and the logarithmic function.

8. Economic and other outside factors are not considered and included in the estimation.

9. Standard BP program is developed in Windows environment using Microsoft Visual C++ 6.0.

1.7 Significance of the Study

The study examines the effectiveness of BP network model as an alternative tool for forecasting. This study provides a practical introductory guide in the design of an ANN for forecasting time series data. We use the time series corresponding to the revenue collection in Royal Malaysian Customs Department to illustrate this process. This research attempts to study the behavior of ANN models when several of its parameters are altered. The relevancy of applying difference non-linear activation functions in hidden and output layers of ANN model is also examined in this study.

1.8 Organization of the Report

This report consists of five chapters. Chapter 1 presents the introduction of the study. Chapter 2 presents an appropriate literature and review on forecasting, ANN in time series forecasting, traditional time series forecasting, performance comparison between ANN and traditional time series forecasting technique from the past researched, explanation of ANN concept including network structure, BP algorithm and the affect of activation functions to BP learning. Chapter 3 discusses on the methodology used in this study. Chapter 4 provides experimental results and analysis of the obtained results. Chapter 5 draws the conclusion and suggestions for future work.

REFERENCES

Amstrong, J.S., Collopy, F., 1992. Error measures for generalizing about forecasting methods: empirical comparisons. International Journal of Forecasting, 8, 69-80.

Athiya, A.F., 2001. Bankruptcy prediction for credit risk using neural networks: A survey and new results. IEEE Transactions on Neural Networks 12(4), 929-935.

Bacha, H. and Meyer, M., 1992. A neural network architecture for load forecasting. IEEE, 442-447.

Bilski, J., 2000. The backpropagation learning with logarithmic transfer function. In: Proceeding of The fifth Conference on Neural Network and Soft Computing. 6-10 June, 2000. Poland. Zakopane, 71-76.

Bódis, L., 2004. Financial time series using artificial neural networks. Master Thesis, .Babeş-Bolyai. University.

Carbone, Robert and Armstrong, J. S., 1982. Evaluation of extrapolative forecasting methods: Results of a survey of academicians and practitioners. Journal of Forecasting, 1, 215-217.

Chandra, P. and Singh, Y., 2004. Feedforward sigmoidal network-equicontinuity and fault-tolerance properties. IEEE Transaction on Neural Networks 15(6), 1350-1366.

Crone, S.F., Kauseh,H. and Preβmar, D., 2004. Prediction of the CATS benchmark using a business forecasting approach to multilayer perceptron modelling. In:

Wunsch, D. et al. (eds.): Proceedings of the International Joint Conference on Neural Networks, IJCNN'04, Budapest, Hungary, IEEE, New York.

Danh, N.T., Phien, H.U. and Gupta, A.S., 1999. Neural network models for river flow forecasting. Water SA 25(1). 1 January 1999. http://www.wrc.org.za.

Duch, W. and Jankowski, N., 1999. Survey of neural transfer functions. Neural Computing Surveys 2, 163-212.

Fnaiech, F., Abid, S., Ellala, N. and Cheriet, M., 2002. A comparative study of fast neural network learning algorithms. IEEE International Conference on Systems, Man and Cybernetics 2002. IEEE SMC, 24-29.

Kaastra, I. and Boyd, M., 1996. Designing a neural network for forecasting financial and economic time series. Neurocomputing 10 (3), 215-236.

Kamruzzaman, J. and Aziz, S.M., 2002. A note on activation function in multilayer feedforward learning. In: Proceeding of International Joint Conference on Neural Network 2002. IEEE, 519-523.

Kodogiannis, V.S. and Anagnostakis, E.M., 2002. Soft computing based techniques for short-term load forecasting. Fuzzy Sets and Systems 128, 413-426.

Kolobe, L., 2006. What is R-Squared? Connexions. Latest updated on February 22, 2006. http://cnx.org/content/m13447/latest/.

Kong, J. H. L. and Martin, G. P. M. D., 1995. A backpropagation neural network for sales forecasting. IEEE, 2121-2124.

Kumar, S., 2005. NEURAL NETWORKS: A CLASSROOM APPROACH, International Edition 2005. McGraw-Hill Publishing Company Limited. ISBN 007-048292-6.

Lam, M., 2004, Neural network techniques for financial performance prediction:

integrating fundamental and technical analysis. Decision Support Systems 37, 567-581.

Limsombunchai, V., Gan, C. and Lee, M., 2004. House price prediction: Hedonic price model vs. Artificial neural network. American Journal of Applied Sciences 1(3), 193-201.

Lopes, M.L.M, Minussi, C.R. and Lotufo, A.D.P., 2000. A fast electric load forecasting using neural networks. In: Proc. 43rd IEEE Midwest Symp. on Curcuits and Systems, Learning MI. 8-11 August, 2000. IEEE, 1-4.

Maier, H.R. and Dandy, G.C., 2000. Neural Networks for the prediction and forecasting of water resources variables: A review of modelling issues and applications. Environmental Modelling & Software 15, 101-124.

Moody, J., 1995. Economic forecasting: Challenges and neural network solutions. International Symposium on Artificial Neural Network, Hsinchu, Taiwan. December 1995. ftp://neural.cse.ogi.edu/pub/neurals/papers.

More, A. and Deo, M.C., 2003. Forecasting wind with neural networks. Marine Structure 16, 35-49.

Nguyen, H.H. and Chan C.W., 2004. A comparison of data preprocessing strategies for neural network modeling of oil production prediction. In: Proceeding of the Third IEEE International Conference and Cognitive Informatics (ICCI'04). IEEE Computer Science.

Osuna, R.G., 2005. Lecture 13: Validation, Intelligent Sensor System, Wright State University. http://research.cs.tamu.edu/prism/lectures/iss/iss_l13.pdf.

Palmer, A., Mantano, J.J. and Sese, A., 2005. Designing an artificial neural network for forecasting tourism time series. Science Direct.

Peres, P. and Reyes, J., 2001. Prediction of particulate air pollution using neural network techniques. Neural Computing and Application 10, 165-171.

Plummer, E.A., 2000. Time series forecasting with feed-forward neural networks: Guidelines and limitations. Master Thesis, University of Wyoming.

Ranaweera, D.K., Hubele, N.F. and Papalexopoulos A.D., 1995. Application of radial basis function neural network model for short-term load forecasting. In: IEEE Proc.-Gener. Tranm. Distrib. 142(1), 45-50.

Rumelhart, D.E., Hinton, G. E. and Willams, R.J., 1986. Learning internal representation by error propagation. Parallel Distributed Processing l, MIT Press, 318-362

Sahai, A.K., Soman, M.K. and Satyan V., 2000. All India summer monsoon rainfall prediction using an artificial neural network. Climate Dynamics 16, 291-302.

Shafie, A.S., 2005. Improved two-term backpropagation error function for classification problems. Master Thesis, University Technology of Malaysia.

Shlens, J., 1999. Time series production with artificial neural networks, 1-20.

Tal, B. and Nazareth, L., 1995. Artificial intelligence and economic forecasting. canadian business economic. Spring 1995, 69-74.

Tan, C.N.W., Wittig, G. E., 1993. A study of the parameters of a backpropagation stock price prediction model. The First New Zealand International Two Stream Conference on Artificial Neural Networks and Expert Systems (ANNES), University of Otago, Dunedin, New Zealand. November 24-26, 1993. IEEE Computer Society Press, ISBN 0-8186-4260-2, 1993a.

Thawornwong, S. and Enke, D., 2004. The adaptive selection of financial and economic variables for use with artificial neural networks. Neurocomputing 56, 205-232.

Virili, F. and Freisleben, B., 2000. Nonstationary and data preprocessing for neural network predictions of an economic time series. IEEE, 129-134.

Walczak, S., Cerpa, N., 1999. Heuristic principles for the design of artificial neural networks. Information and Software Technology 41 (2), 109-119.

Xu,Y.Y, Hsieh, R., Lu Y.L. *et al.*, 2004. Forecasting electricity market prices: A Neural network based approach. IEEE, 2789-2794.

Xu, L. and Chen, W.J., 2001. Short-term load forecasting techniques using ANN. In: Proceeeding of the 2001 IEEE International Conference of Control Applications, 157-160.

Yao, J.T., 2002. Towards a better forecasting model for economic indices. In: Proceedings of the 6th Joint Conference on Information Science, 299-303.

Yu, E.S. and Chen, R.C.Y., 1993. Traffic prediction using neural networks. IEEE, 991-995.

Yu, S.W., 1999. Forecasting and arbitrage of the nikkei stock index futures: An application of backpropagation networks, Asia-Pasific Finalcial Market 1999, 6(4), 341-354.

Zekic, M., 2005. Neural network applications in stock market predictions-A methodology analysis, 1-11.

Zhang, G., Patuwo, B.E. and Micheal, Y.H., 1998. Forecasting with artificial neural networks: The state of the art. International Journal of Forecasting 14, 35-62.

Zhang, Y., 1992. Prediction of traffic fluctuation in telephone networks with neural networks. IEEE, 909-912.