



KOLEJ SASTERA DAN SAINS
(College of Arts and Sciences)
Universiti Utara Malaysia

PERAKUAN KERJA KERTAS PROJEK
(Certificate of Project Paper)

Saya, yang bertandatangan, memperakukan bahawa
(I, the undersigned, certify that)

TAY SHU SHIANG
(89709)

calon untuk Ijazah
(candidate for the degree of) **MSc. (Information Technology)**


telah mengemukakan kertas projek yang bertajuk
(has presented his/her project paper of the following title)

PREDICTING EMPLOYMENT CONDITION OF TARC'S ICT GRADUATES
USING BACKPROPAGATION NEURAL NETWORK

seperti yang tercatat di muka surat tajuk dan kulit kertas projek
(as it appears on the title page and front cover of project paper)

bahawa kertas projek tersebut boleh diterima dari segi bentuk serta kandungan dan meliputi bidang ilmu dengan memuaskan.
(that the project paper acceptable in form and content, and that a satisfactory knowledge of the field is covered by the project paper).

Nama Penyelia Utama
(Name of Main Supervisor): **MR. MOHAMAD FARHAN MOHAMAD MOHSIN**

Tandatangan
(Signature) : 

Tarikh
(Date) : 18/11/09

**PREDICTING EMPLOYMENT CONDITION OF
TARC'S ICT GRADUATES
USING
BACKPROPAGATION NEURAL NETWORK**

TAY SHU SHIANG

UNIVERSITI UTARA MALAYSIA

MSc. (IT) 2009

PERMISION TO USE

This thesis presents a partial fulfillment of the requirement for a postgraduate degree from Universiti Utara Malaysia. I agree that the university library may make it freely available for inspection. I further agree that the permission for copying of this thesis in any manner, in whole or in part, for scholarly purposes may be granted by my supervisor or, in their absence by the Assistant Vice Chancellor of the College of Arts and Sciences. It is understood that any copying or publication or use of this thesis or parts thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and to Universiti Utara Malaysia for any scholarly use which may be made of any material from my thesis.

Requests for permission to copy or to make other use of materials in this thesis, in whole or in part should be addressed to:

**Assistant Vice Chancellor
College of Arts and Sciences
Universiti Utara Malaysia
06010 UUM Sintok
Kedah Darul Aman.**

ABSTRACT

This research is conducted with the purpose of classifying the employment condition of ICT students after their graduation using Backpropagation Neural Network (BPNN). To narrow down the scope of the research, ICT students from Tunku Abdul Rahman College (TARC) are targeted. The employment condition will be predicted and classified based on several macroscopic and microscopic criterion indentified. The macroscopic reasons include the social and the governmental factors while the microscopic reasons cover the college and the student factors. This paper will show the BPNN steps involved in creating a suitable multilayer-perceptron classification model for the employment condition. Detail descriptions of the BPNN methodologies applied are also included in the report. The findings of the research are expected to provide TARC's management an in-depth view on their students' marketability and adaptability in the work fields.

Keywords: *classifications, Multi-Layer Perceptron, Backpropogation, Neural Network, educational data mining (EDM), employment situation, Tunku Abdul Rahman College (TARC).*

ABSTRAK

Tujuan kajian ini dijalankan adalah untuk menklasifikasikan status kerja bekas pelajar ICT Kolej Tunku Abdul Rahman dengan menggunakan *Backpropagation Neural Network (BPNN)*. Bagi mengecilkan skope kajian, hanya pelajar ICT dari Kolej Tunku Abdul Rahman (KTAR) menjadi tumpuan. Keadaan kerja pelajar adalah mengikut ramalan dan klasifikasi berdasarkan beberapa kriteria makroskopik dan mikroskopik yang ditentukan. Ciri-ciri makroskopik meliputi faktor sosial dan faktor kerajaan sementara ciri-ciri mikroskopik meliputi faktor mahasiswa and faktor universiti. Laporan ini akan menjelaskan langkah-langkah BPNN yang terlibat dalam mencipta satu model klasifikasi perseptron berlapis (MLP) untuk menganggar keadaan kerja para pelajar. Maklumat keterangan hasil daripada metodologi yang diterapkan dalam BPNN juga disertakan dalam laporan ini. Diharap hasil penemuan kajian ini dapat memberi pandangan yang lebih mendalam kepada pengurusan KTAR tentang kemampuan dan kebolehan pelajar mereka beradaptasi dalam sektor pekerjaan.

Kata kunci: *klasifikasi perseptron berlapis (MLP), Backpropagation Neural Network, perlombongan data pendidikan (EDM), situasi pekerjaan, Kolej Tunku Abdul Rahman (KTAR).*

ACKNOWLEDGEMENTS

First of all, my most profound thankfulness goes to my final project supervisor Encik Mohamad Farhan Bin Mohamad Mohsin for his help, guidance and encouragement. I would also like to thank his continuous faith and support in me. Without his encouragement and guidance, it will not be easy for me to reach this extend in completion my report.

Secondly, I would like to thank all my dearest family members, especially to my parents, brothers who have given me their full support in my study. Their full support remains the mainstay for me in overcoming all the difficulties in completing this study. Next I would like to thank all the lecturers who have taught me before throughout the Masters Degree course because the knowledge they imparted have allowed me to be more knowledgeable and thus in a better position to complete this research.

Lastly, I would like to thank all my friends who had given me emotional support and taken care of me at times of difficulties, especially thanks to Ms. Nai Seow Hong, Mr. Ang Shao Long, Ms Khor Jia Yun, Ms. Loo Sze Phei and Ms. Lim Lai Hoon in advising and guiding me in the process of completing this report.

TABLE OF CONTENTS

TITLE	PAGE
PERMISION TO USE.....	I
ABSTRACT.....	II
ABSTRAK.....	III
ACKNOWLEDGEMENTS.....	IV
TABLE OF CONTENTS.....	V
LIST OF TABLES.....	VIII
LIST OF FIGURES.....	IX
LIST OF ABBREVIATIONS.....	X
CHAPTER 1: INTRODUCTION.....	1
1.1 Background.....	1
1.2 Problem Statement.....	2
1.3 Project's Objective.....	3
1.4 Scope and Limitation.....	3
1.5 Significance of the Study.....	4
CHAPTER 2: LITERATURE REVIEW.....	5
2.1 Unemployment.....	5
2.2 Data Mining.....	7
2.3 Neural Network (NN) and Backpropagation.....	8
2.4 Application of Data Mining and Discussion.....	8

2.4.1	<i>Education</i>	9
2.4.2	<i>Telecommunication</i>	10
2.4.3	<i>Legal</i>	11
2.4.4	<i>BioMedical</i>	11
2.4.5	<i>Natural Disaster</i>	12
2.4.6	<i>Businesses</i>	12
CHAPTER 3: METHODOLOGY		14
3.1	Data Collection.....	16
3.2	Data Cleaning	20
3.3	Data Transformation	21
3.4	Neural Network Tool	23
3.5	Data Mining & The Experiments	30
3.6	Construction of Prediction Model.....	34
CHAPTER 4: RESULTS ANALYSIS		35
4.1	To determine the most suitable number of Hidden Units.....	35
4.2	To determine the most suitable Learning Rate.....	38
4.3	To determine the most suitable Momentum Rate.....	40
4.4	To determine the best Activation Function.....	42
4.5	To determine the most suitable Number of Epoch (cycle)	43
4.6	To determine the most suitable Data Allocation.....	45
4.7	The Network Architecture.....	47

4.8	Summary.....	49
CHAPTER 5: CONCLUSIONS.....		50
5.1	Research Contributions.....	50
5.2	Future Works.....	51
REFERENCES.....		52
APPENDIX A.....		55

LIST OF TABLES

Table 3.1	Descriptions of attributes and representation
Table 3.2	How attributes values are normalized
Table 4.1(a)	Result to determine the best Hidden Unit
Table 4.1(b)	Determine the best Hidden Unit using various Weight Seeds
Table 4.2	Result to determine the best Learning Rate
Table 4.3	Result to determine the best Momentum Rate
Table 4.4	Determine the best Activation Function using various Weight Seeds
Table 4.5	Result to determine the most suitable Number of Epoch
Table 4.6	Result to determine the most suitable Data Allocation
Table 4.7	Neural Network Model and the optimum parameters

LIST OF FIGURES

Figure 3.1	Steps of methodology in predicting employment condition of TARC's ICT graduates
Figure 3.2	Samples of original data
Figure 3.3	Alumni's Employment Condition Dataset
Figure 3.4	Original Data
Figure 3.5	Normalized Data
Figure 3.6	Data allocation
Figure 3.7	Training, Validation and Testing set
Figure 3.8	General architecture of Multilayer Perceptron
Figure 3.9	Multi-Layer Perceptron Network dialog box
Figure 3.10	MLP Training Stages
Figure 3.11	Data Output settings
Figure 3.12	Output Format
Figure 3.13	Accuracy Result sample
Figure 4.1	Average Training and Testing accuracy of different Hidden Units
Figure 4.2	Training and Testing accuracy of different Learning Rates
Figure 4.3	Training and Testing accuracy of different Momentum Rates
Figure 4.4	Average Training and Testing accuracy of different Activation Functions
Figure 4.5	Training and Testing accuracy of different Number of Epoch
Figure 4.6	Average Training and Testing accuracy of different Number of Epoch
Figure 4.7	Case Network Architecture

LIST OF ABBREVIATIONS

ANN	Artificial Neural Network
BLAST	Basic Local Alignment Search Tools
BNN	Bayesian Neural Network
BPNN	Backpropagation Neural Network
CSV	Comma-separated Value
ICT	Information and Communication Technology
EDM	Educational Data Mining
MLP	Multilayer Perceptron
MOHE	Ministry of Higher Education
MCI	Multiplicative Competitive Interaction
NN	Neural Network
RMSE	Root mean square error
SAM	Spectral Angle Mapper
SPSS	Statistical Package for the Social Sciences
TARC	Tunku Abdul Rahman College
WLAN	Wide Local Area Network

CHAPTER 1

INTRODUCTION

This chapter starts with discussing the background of the study by quoting some facts obtained from the journals and local newspapers. It is followed by the problem statement, the objectives of the study and the significance of the study. The scope and the limitation of the study is also included in this chapter.

1.1 Background

In the recent years, graduates from various higher learning institutions have noticed that landing a job in the work field is becoming more challenging than the past. A lot of the graduates complained that they are always turned down during the interview or employed in the field not related to what they are majoring in. According to the statistic given by Ministry of Human Resource, 70% of public universities and institutes of higher learning graduates in the country are unemployed. This is in contrast with 26% for private institutions of higher learning and 34% for foreign graduates [28]. A survey conducted earlier has indicated that as many as 30% of the unemployed local graduates are computer science and information technology degree holders. This result is somehow surprising as the ICT skills are in obvious demand in the country.

The contents of
the thesis is for
internal user
only

REFERENCES

- [1] Ángel García-Crespo, Ricardo Colomo-Palacios, and Juan Miguel Gómez-Berbis (2009), “*IT Professionals’ Competences: High School Students’ Views*”, Journal of Information Technology Education 2009, Volume 8.
- [2] Antonio Nogueira, Paulo Salvador, and Rui Valadas (2006), “*Predicting the Quality of Service of Wireless LANs using Neural Networks*”, MSWiM '06: Proceedings of The 9th ACM International Symposium On Modeling Analysis And Simulation of Wireless And Mobile Systems November 2006.
- [3] Antonio Soriano Payá, Daniel Ruiz Fernández, David Gil Méndez, Carlos Alberto Montejo Hernández (2006), “*Development Of An Artificial Neural Network For Helping To Diagnose Diseases In Urology*”, BIONETICS '06: Proceedings of The 1st International conference On Bio Inspired Models Of Network, Information And Computing Systems December 2006.
- [4] Beale, R and Edwards, A. D. N. (1992), Recognizing postures and gestures using neural networks. in R. Beale and J. Finlay (ed.) *Neural Networks and Pattern Recognition in Human-Computer Interaction*. New York: Ellis Horwood. pp. 163–169.
- [5] C. Stergiou, and D. Siganos, “*Neural Networks*”, *SURPRISE Journal*, Imperial College, 1996, vol. 4.
- [6] DeLurgio, S.A. (2000), “*Forecasting Principle and Applications*”, McGraw-Hill International Editions.
- [7] Guangming Wang and Xiangna Zheng, “*The unemployment rate forecast model basing on BP neural network*”, International Conference on Electronic Computer Technology 2009.
- [8] Haykin, S. (1999), “*Neural Networks: A Comprehensive Foundation*”, 2nd Edition, New Jersey, Prentice Hall.
- [9] Jason T. L. Wang, Qic Heng Ma, Dennis Shasha, Cathy H. Wu (2000), “*Application of Neural Networks to Biological Data Mining: A Case Study in Protein Sequence Classification*”, KDD '00: Proceedings of The Sixth ACM SIGKDD International Conference On Knowledge Discovery And Data Mining August 2000.
- [10] Jerzy Bala, Sung Baik, Ali Hadjarian, BK Gogia, Chris Manthorne, “*Application of a Distributed Data Mining Approach to Network Intrusion Detection*”, Proceedings of The First International Joint Conference On Autonomous Agents And Multiagent Systems: Part 3, pp. 1419 – 1420.

- [11] Jiawei Han and Micheline Kamber, *"Data Mining- Concepts and Techniques"*, Academic Press, Morgan Kaufmann Publishers, 2001.
- [12] Kaastra, I. and Boyd, M. (1996), *"Designing a Neural Network for Forecasting Financial and Economic Time Series"*, *Neurocomputing*, 10, 215-236.
- [13] Kantardzic, Mehmed (2003), *"Data Mining: Concepts, Models, Methods, and Algorithms"*, John Wiley & Sons, ISBN 0471228524, OCLC 50055336.
- [14] Laurent Boheureau, DaniUe Bourcier, Paul Bourguine (1991), *"Extracting Legal Knowledge By Means Of A Multilayer Neural Network Application To Municipal Jurisprudence"*, ICAIL '91: Proceedings of The 3rd International Conference On Artificial Intelligence aAnd Law May 1991.
- [15] Lawrence, J. (1991), *"Data Preparation for a Neural Netowork"*, *AI Expert*, vol. 6, no. 11, pp. 34-41.
- [16] LiMin Fu (1994), *"Neural Networks in Computer Intelligence"*, Singapore: McGraw Hill, pp.18-19, 31, 80-82.
- [17] Lulseged Ayalew, Dietmar P.F. Möller, Gerhard Reik (2007), *"Using Artificial Neural Networks (ANN) for Real Time Flood Forecasting, the Omo River Case in Southern Ethiopia"*, SCSC: Proceedings Of The 2007 Summer Computer Simulation Conference July 2007.
- [18] Martin T. Hagan, Howard B. Demuth and Orlando De Jesus, *"An Introduction To The Use Of Neural Networks In Control System"*, School of Electrical & Computer Engineering, Oklahoma State University, Stillwater, Oklahoma, 74075, USA, Electrical & Computer Engineering Department, University of Colorado, Boulder, Colorado, 80309, USA
- [19] Qingjun Wu (2008), *"Analysis and Research on Employment Predicament of University Students"*, *Journal of Sustainable Development*, vol. 5, no. 2.
- [20] Ripley, B.D. (1996), *"Pattern Recognition and Neural Networks"*, Cambridge: Cambridge University Press.
- [21] Steve Engels, Vivek Lakshmanan, Michelle Craig (2007), *"Plagiarism Detection Using Feature-Based Neural Networks"*, SIGCSE '07: Proceedings Of The 38th SIGCSE Technical Symposium On Computer Science Education March 2007.
- [22] Sushmita Mitra, Sankar K. Pal and Pabitra Mitra, *"Data Mining in Soft Computing Framework: A Survey"*, *IEEE Transactions on Neural Netowrks*, vol. 13, no. 1, January 2002.

- [23] Thomas S. Gruca, Bruce R. Klemz, E. Ann Furr Petersen (1999), "*Mining Sales Data using a Neural Network Model of Market Response*", SIGKDD Explorations Newsletter June 1999, vol. 1, issue 1.
- [24] Tsantis, L., & Castellani, J. (2001, Fall), "*Enhancing Learning Environments Through Solution-Based Knowledge Discovery Tools: Forecasting For Self Perpetuating Systemic Reform*", JSET E Journal October 2003.
- [25] Tsoukalas, L.H., Uhrig, R.E., "*Fuzzy and Neural Approaches in Engineering*", John Wiley & Sons, New York, NY, 1997.
- [26] Warren S. Sarle (1994), "*Neural Network and Statistical Models*", Proceedings of the Nineteenth Annual SAS Users Group International Conference April 1994.
- [27] Yoon, Y. & G. Swales (1991), "*Predicting Stock Price Performance: A Neural Network Approach*", Proceedings of the IEEE 24th Annual Hawaii International Conference of System Sciences, 156-162.
- [28] The Sun, 3 Jul 2006, "*70% of grads from public institutions jobless*", B.Suresh Ram.
- [29] New Sunday Times, 2005, "*OPINION: The unemployable Malaysian graduates*", Abdul Razak Ahmad