DETECTING FRAUD PATTERNS IN TELECOMMUNICATIONS USING CASE BASED REASONING

BALA MUSA SHUAIBU

UNIVERSITI UTARA MALAYSIA
2008

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)

BALA MUSA SHUAIBU

calon untuk Ijazah
(candidate for the degree of) **MSc. (Intelligent System)**

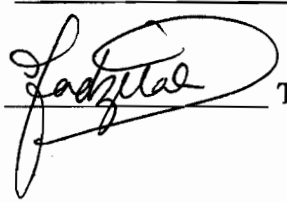
telah mengemukakan kertas projek yang bertajuk
(has presented his/ her project paper of the following title)

DETECTING FRAUD PATTERN IN TELECOMMUNICATIONS USING CASE BASED REASONING

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): **ASSOC. PROF. FADZILAH SIRAJ**

Tandatangan
(Signature) :  Tarikh (Date) : 29/04/2008

Nama Penyelia Kedua
(Name of 2nd Supervisor): **MS. NOORAINI YUSOFF**

Tandatangan
(Signature) : Tarikh (Date) : 29/04/2008

ABSTRACT

All round the world, fraud situations are significantly causing huge revenue leakage in the telecommunication companies every year. The reuse of previous cases is an important issue in dealing with fraud pattern in a data with string features. Case Based Reasoning (CBR) systems have a set of cases inform of library used to facilitate the process of validation of new cases without the direct involvement of a domain expert. The proposed detection technique in this paper is based on Case Based Reasoning used to detect the occurrence of fraud with a meaningful confidence in telecommunication data. Experimental result on the fraud data indicates that the weight for all attribute used in this study needs to be set as 0.1 in order to get 98% similarity performance.

ACKNOWLEDGEMENTS

In the Name of Allah, the Most Gracious and the Most Merciful

It is my pleasure to acknowledge the immense contribution of some people who have assisted me one way or the other towards the successful completion of this project.

First of all, I give thanks to the Allah for his guidance and mercy throughout my life. Peace and Blessing to his last Prophet Muhammad (S.A.W.), his household and his companions. My sincere appreciation goes to my beloved parents and family members for their patience, prayers and understanding over the entire period of my study.

Secondly, my grateful thanks go to my supervisors, Prof. Madya Fadzilah Siraj and Nooraini Yusuf who had given their full support and contributed immensely towards the completion of this project. They have actually spent a lot of time patiently and painstakingly giving me the necessary advice, providing valuable information and correcting errors to ensure that the best effort has been given in the completion and achievement of this project.

I also wish to convey my appreciation to Associate Professor. Dr. Norita Md. Norwawi my evaluator who has giving me support and advise for the completion of this project.

Lastly, I recognize the efforts of all my friends, Staff of Faculty of Information Technology, Universiti Utara Malaysia and those who contributed directly or indirectly towards the completion of this project. Thanks to all.

Bala Musa Shuaibu
College of Arts and Sciences
Faculty of Information Technology
Universiti Utara Malaysia
June 2008

TABLE OF CONTENTS

	Page
ABSTRACT	i
ACKNOWLEDGEMENT	ii
TABLE OF CONTENT	iii
LIST OF TABLES	vi
LIST OF FIGURES	vii

CHAPTER 1: INTRODUCTION

1.1 Background	1
1.2 Problem Statement	6
1.2 Research Questions	8
1.3 Objectives	8
1.4 Scope of the Study	9
1.5 Significance of the Study	9
1.6 Organization of the Report	10
1.7 Conclusion	11

CHAPTER 2: LITERATURE REVIEWS

2.1 Telecommunication	12
2.2 Fraud	13
2.3 Rule Based	15
2.4 Neural Network	17
2.5 Case Based	22
2.6 Other Techniques	26

2.7 Summary 30

CHAPTER 3: METHODOLOGY

3.1 Introduction 31
3.2 System Development Research Methodology 32
3.2.1 Construct a Conceptual Frame work 33
3.2.2 Development a System Architecture 34
3.2.3 Analysis and Design 35
3.2.4 Build the System 40
3.2.5 Observe and Evaluate 48
3.3 Summary 49

CHAPTER 4: FINDINGS AND RESULTS

4.1 Functionalities 50
4.2 Interface Design 51
4.3 System Test 55
4.3.1 Percentage Similarity 56
4.4 Conclusion 62

CHAPTER 5: CONCLUSION

5.1 Project's Summary 63
5.2 Problems and Limitations 64
5.3 Recommendation for Future Works 64

REFERENCES 65

APPENDICES

Appendix A	Use Case	70
Appendix B:	Use Case	72
Appendix C:	User Manual	76
Appendix D:	Test Cases	80

LIST OF TABLES

		Page
Table 1.0	Sample data set	7
Table 3.1	Data attribute	36
Table 3.2	Similarity Computation by Weighted Average	40
Table 4.1	the summary of result for test case 1,2 various attributes weight	59
Table 4.2	the summary of result for test case 3,4 various attributes weight	60
Table 4.1	the summary of result for test case 5,6 various attributes weight	61

LIST OF FIGURES

	Page
Figure 3.1: The Systems Development Research Methodology	33
Figure3.2: Fraud Detection System Architecture	34
Figure 3.3: Sample data set in Microsoft Excel	36
Figure 3.4: Sample data set in Microsoft Access.	37
Figure 3.5: The attributes of Data table	38
Figure 3.6: Rapid Application Development Architecture	43
Figure 3.7: User Design Architecture	44
Figure 3.8: Use case Diagram	44
Figure 3.9: the sequence diagram for login	45
Figure 3.10: Retrieved case sequence diagram	46
Figure 3.11: collaborative diagram	46
Figure 3.12: collaborative diagram	47
Figure 3.13: CBR Engine Welcome Page	48
Figure 4.1: Welcome Page	51
Figure 4.2: Login Page	52
Figure 4.3: Error Message	52
Figure 4.4: System Main Page	53
Figure 4.5: Similarity Retrieval Page	54
Figure 4.4: CBR output when attribute 1 Matches	56
Figure 4.5: CBR output when attribute 1 is not matched	57
Figure 4.6: CBR output when attribute 1 is not match	57
Figure 4.6: CBR output when attribute 1 is not match	58

CHAPTER 1

INTRODUCTION

This chapter gives a general overview of the project. It starts with a background review of the project which explains the motivation behind the project and the domain on which the project is based on. The chapter further describes the problem statement, the objective to be accomplished; significant to be derived from the project, the scope or coverage area and finally highlighted the way subsequent chapters will be organized.

1.1 Background

Telecommunications has brought tremendous achievement in life by providing a means to extend communication over a distance. This is done through telephony, computer networking, television, radio, and so on to transmit information in form of voice or data.. Telecommunications facilitates Internet service provision, networking of computers and telephones and cellular service for cars, fax or modem, or offers voice mail that meet specific business needs. Many telecommunications companies manufacture equipments ranging from simple pagers to mobile office packages for Internet service and high-speed connections. Therefore, telecommunications has brought revolution of new products and technologies to the market.

Examples of such products are:

The contents of the thesis is for internal user only

Reference:

Adem, K., Senay Y., Cengiz, K., & Mert, S. (2007). Fraud Detection Using an Adaptive Neuro-Fuzzy Inference System in Mobile Telecommunication. Networks Information Sciences 2007, pp1440-1446.

Amani, N., Fathi, M., Dehghan, M., (2005). A case-based reasoning method for alarm filtering and correlation in telecommunication networks Electrical and Computer Engineering, 2005. in Canadian Conference, 2005, pp2182 - 2186

Burge, P., & Shawe-Taylor, J., (2001). An Unsupervised Neural Network Approach to Profiling the Behaviour of Mobile Phone Users for Use in Fraud Detection. Journal of Parallel and Distributed Computing 61, pp915-925.

THE HINDU group of publications. Business Daily (n.d.). Retrieved January 23, 2008, from http://www.thehindubusinessline.com/2006/09/08/stories/2006090803460400.htm

Clifton, P., Vincent, L., Kate, S. & Ross, G., (2005). Comprehensive Survey of Data Mining-based Fraud Detection Research. School of Business Systems, Faculty of Information Technology, Monash University, Clayton campus Wellington Road, Clayton, Victoria 3800, Australia

CFCA press release (2007). Retrieved on 15th April 2008 from http://www.cfca.org/pdf/press/3-28-6PR.pdf,

Fair, I., (2003). Prepaid Telecommunications Fraud Techniques and Detection Retrieved on 15th April 2008 from: www.fairisaac.com/telecom.

Jaakko, H., (2000). User profiling and classification for fraud detection in mobile communications networks. Dissertation for the degree of Doctor of Science in Technology, Helsinki University of Technology Department of Computer Science and Engineering Laboratory of Computer and Information Science.

Jimmy, M., & Seán, H., (2003). Approach to Rules based Fraud Management in Emerging Converged Networks. Telecommunications Software & Systems Group, Waterford Institute of Technology, Ireland

Joshua, D., Bruce, M,. and David, W., (2004). Towards Applying Case-Based Reasoning to Composable Behavior Modeling. In the Proceedings of the 2004 Conference on Behavior Representation (BRIMS), Arlington, Virginia.

Jian-Bo Yang; Jun Liu; Jin Wang; Sii, H.S.(2003). The Evidential Reasoning approach for Inference in rule-based systems. In Systems, Man and Cybernetics, 2003. IEEE International Conference, 3, pp2461 - 2468

Michael, H., Diane, L., Jos´e C., & Don, X. (2000). Detecting Fraud in the Real world. Retrieved from http://cm.bell-labs.com/cm/ms/departments/sia/jcp/HMDS.ps

Pablo, A., Claudio, M., & Claudio, A. (2006). Subscription Fraud Prevention in Telecommunications using Fuzzy Rules and Neural Networks. Expert Systems with Applications Journal, August 2006, pp337-344.

Rupesh, K. G., and Saroj, K., (2007). Rule-based Approach for Anomaly Detection in Subscriber Usage Pattern. World Academy Of Science, Engineering And Technology Journal , 25, pp1307-6884.

Wheeler, R., Aitken, S (2000). Multiple Algorithms for Fraud Detection. Knowledge Based Systems Journal, 13 pp2-3 & 93-99.

Wikipedia, the free encyclopedia (2008). Case Based Reasoning retrieved on 24th January 2008 from http://en.wikipedia.org/wiki/Case-based_reasoning

Giulio, V., and Alessandro, R. (2004). Using Case-based Reasoning to support Operational Knowledge In Proceedings of 14th International Conference, EKAW 2004 Whittlebury Hall, UK,

Tomoharu, N., Gaku, N., and Hisao I., (2003). Credit Assignment by Fuzzy Rule-Based Systems in Fuzzy Classifier Ensembles. In a Proceedings 2003 IEEE International Symposium on Computational Intelligence in Robotics and Automation, Kobe, Japan.

Ching-Chang, W., and Chia-Chong, C., (2000). An SVD-QR-based approach to fuzzy modeling. In Systems, Man and Cybernetics, 2000. IEEE International Conference.

Pattara, W., and Peachavanish, R.,(2007). Estimating Road Traffic Congestion from Cell Dwell Time using Neural Network. In Telecommunications, 2007. ITST '07. 7th International Conference on ITS.

Munshi, K. Vempada, P., Sheila P., Sonmez, E., and Schumacher, H., (2003). Small signal and large signal modeling of HBT's using neural networks. In Telecommunications in Modern Satellite, Cable and Broadcasting Service, 2003. TELSIKS 2003. 6th International Conference.

Zhi-Wei, N., Shan-Lin, Y., Long-Shu, L., and Rui-Yu J., (2003). Integrated case-based reasoning. In machine Learning and Cybernetics, 2003 International Conference

G. Kamp, S. Lange and C. Globig, (1998) Case-based reasoning technology: related areas. In: M. Lenz Editor, *Case-based Reasoning Technology: from Foundations to ApplicationLNAI no. 1400* Springer, Berlin, pp. 325

Intec Telecom Systems PLC (2002). The Boom of New Technologies and the Rise of Telecom Fraud in Malaysia. Retrieved on 15th April 2008 from http://www.cybersecurity.org.my/bahasa/knowledge_bank/news/2002/main/detail/937/index.html

Mingyang Gu, Xin Tong, and Agnar Aamodt, (2005). Comparing Similarity Calculation Methods in Conversational CBR. Department of Computer and Information Science, Norwegian University of Science

Mingyang Gu1 and Xin Tong2(2004) An Intelligent Component Retrieval System Using Conversational CBR. a Proceedings of the 28th Annual International Computer Software and Applications Conference (COMPSAC'04) 0730-3157/04

Mehmet S. Aktas, Marlon Pierce, Geoffrey C. Fox, David Leake (2004). A Web based Conversational Case-Based Recommender System for Ontology aided Metadata Discovery. In proceedings of the Fifth IEEE/ACM International Workshop on Grid Computing (GRID'04) 1550-5510/04

Aamodt, A., & Plaza, E. (1994). Case-based reasoning: Foundational issues, methodological variations, and system approaches. *AI Communications*, 7, pp39-59.

Nenad S. Kojić, Irini S. Reljin, Branimir D. Reljin (2007). Different Wavelength Assignment Techniques in Al l-Optical Networks Controlled by Neural Network.

O. Dehzangi, M. J. Zolghadri, S. Taheri and S.M. Fakhrahmad (2007).Efficient fuzzy rule generation: A new approach using data mining principles and rule weighting. IN Fourth International Conference on Fuzzy Systems and Knowledge Discovery (FSKD 2007).

J. Shawe-Taylor, K. Howker, P. Gosset, M. Hyland, H. Verrelst, Y. Moreau, C. Stoermann, and P. Burge (2000). In Business Applications of Neural Networks, chapter Novel techniques for prolong and fraud detection in mobile telecommunications, pp113-139.

Wheeler, R. and Aitken, S.(2000) Multiple Algorithms for Fraud Detection. In Journal Knowledge-Based Systems, 13,pp2-39 & 3-99.

David G. Messerschmitt (1996).Convergence of telecommunications with computing Department of Electrical Engineering and Computer Sciences University of California. In Proceeding of the IEEE, 84,8, pp1167-1186

John Beacon (2004). The impact of Mobile telephoning in Australia. In a conference Australian Mobile telecommunication Association conference.

Bolton R. J. And Hand D. J. (2002). Statistical Fraud Detection: A Review. In journal Statistical Science 17, 3, pp235–255.

Clifton P., Damminda A., and Vincent L. (2004). Minority Report in Fraud Detection:Classification of Skewed Data. In ACM SIGKDD Explorations Newsletter, 6 , 1. pp50-59.

Abhinav S., Amlan K., Shamik S., and Arun K. M. (2008) Credit Card Fraud Detection Using Hidden Markov Model. In proceeding Ieee Transactions On Dependable And Secure Computing 5, 1, pp37-48

Marius M., and Anders K. (2004). Representing and Reasoning about Context in a Mobile Environment Proceedings Workshop on Modelling and Retrieval of Context, 2004

Janet L. K. (2004). An introduction to case-based reasoning, Netherlands: Springer.

Olusola A. (2005). Data Mining, Fraud Detection and Mobile Telecommunications: Call Pattern Analysis with Unsupervised Neural Networks. Master of Science thesis, Computer Science, University of the Western Cape..

Frank R. J., Hunt S. P., and Davey N. (2000). Applications of Neural Networks to Telecommunications Systems.

Bonchi F., Giannotti F., Mainetto G., Pedreschi D. (1999). A classification-based methodology for planning audit strategies in fraud detection. In Proceedings of the fifth ACM SIGKDD international conference on Knowledge discovery and data mining, pp175 – 184.

Bailey, K. D. (1982), ***Methods of Social Research,*** The Free Press,.

Blake, S. P.,(1978) ***Managing for Responsive Research and Development,*** W. H. Freeman and Company,

Nunamaker J.F., Chen M. and Purdin T.D.M. (1991). Systems development in information systems research. In Journal of Management Information Systems. 7, 3, pp. 89-106, 1991.

Blalock, A. B. and Blalock, H. M., Jr (1982). Introduction to Social Research, Prentice-Hall, second edition,.

Konsynski, B. R. and Nunamaker, J. F., Jr., (1982). Plexsys: A System Development System," in Couger, Colter, and Knapp (eds.), Advanced Systems Development/Feasibility Techniques, John Wiley & Sons, Inc.,.