FEATURES REDUCTION IN CASE RETRIEVAL FOR DIABETES DATASET

ABDALLA ALI ABDALLA BALA

UNIVERSITI UTARA MALAYSIA (2007)

PUSAT PENGAJIAN SISWAZAH
(*Centre For Graduate Studies*)
Universiti Utara Malaysia

PERAKUAN KERJA KERTAS PROJEK
(Certificate of Project Paper)

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

ABDALLA ALI A. BALA

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)

FEATURES REDUCTION IN CASE RETRIEVAL FOR DIABETES DATASET

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. AZIZI AB AZIZ**

Tandatangan
(Signature) : _____

AZIZI AB AZIZ
Lecturer / Postgraduate Coordinator
Department of Computer Science
Faculty of Information Technology
Universiti Utara Malaysia

Tarikh
(Date) : 8/DEC/2007

FEATURES REDUCTION IN CASE RETRIEVAL FOR DIABETES DATASET

A Thesis Submitted To Faculty of Information Technology

In Partial Fulfillment of the Requirements for the Degree

Master of Science (Information Technology)

Universiti Utara Malaysia

By

Abdalla Ali Abdalla Bala

PERMISSION TO USE

In presenting this thesis in partial fulfillment of the requirements 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 permission for copying of this thesis in any manner, in whole or in part, for scholarly purposes may be granted by my supervisor, in her absence, by the Dean of the Faculty of Information Technology. It is understood that any copying or publication or use of this thesis or parts there of 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 material in this thesis, in whole or in part, should be addressed to:

Dean of Faculty of Information Technology

Universiti Utara Malaysia

06010 UUM Sintok

Kedah Darul Aman

ABSTRACT

In reality, the organizations often have the great quantity of data stored in the databases. The large size of data in terms of the number of attributes and objects make the analysis process becomes very difficult as the data are complex. In order to overcome this problem, the use of sufficient number of attributes and objects will contribute to get the best solution. There are many techniques which can be employed to reduce the number of attributes in the dataset. In this study, two techniques core using, namely rough set theory and Case-Based Reasoning were applied to the medical dataset.

ACKNOWLEDGEMENTS

I would like to express my thanks and gratitude to Allah, the Most Beneficent, the Most Merciful whom granted me the ability and willing to start and complete this project. I pray to his greatness to inspire and enable me to continue this work.

I am truly indebted to many people who have contributed to this thesis. First, I would like to thank my supervisor Mr. Azizi Ab Aziz, who has been the most influential person during my study, who has given me many insightful advices. The valuable guidance from him has made this project come true and success. I also would like to thank to those who has helped me, giving suggestion and encouragement to me at all time during the development of this project and writing the proceeding paper, and the report of this project. They are especially my parents, Mr. Ali Bala and Mdm. Najmiah, and all of my brothers, sisters and their husbands, for always being there for me; Dr. Fawzi Elfaidi from UKM Bangi, who had encouraged me to pursue my graduate studies abroad and very helpful to me during my study in Malaysia; and to all of my friends especially my dear friend, Norfadila Mahrom who always beside me, understand and supported me towards the completion of this project.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

Acronym	Meaning
AI	Artificial Intelligence
CBR	Case-Based Reasoning
GA	Genetic Algorithm
GGA	Generational Genetic Algorithm
IS	Importance Score
JA	Johnson Algorithm
KDD	Knowledge Discovery from Databases
MBR	Model-Based Reasoning
MMR	Multi-Modal Reasoning
PBIL	Population-Based Incremental Learning
P 2 P	Peer-To-Peer
RST	Rough Set Theory
SGA	Steady-state Genetic Algorithm

CHAPTER ONE

INTRODUCTION

This chapter presents the idea of this study and the techniques that are focused in this study. In addition, this chapter discusses the problem statement, objectives, scope and the significance of the study.

1.1 Overview Of The Study

Diabetes is formerly known as a group of diseases defined by high levels of blood glucose from deficiencies in insulin production, insulin action or both. It is a chronic disease where the body does not produce or properly use insulin. Insulin helps glucose or sugar leave the blood and get into the body's cells. If not treated, the sugar that builds up in human blood can damage the heart, eyes, kidneys and blood vessels. Referring to Hejlesen *et al.* (2000), diabetes is a disease that affects more than 100 million people in this world

As the number of people affected by diabetes keep growing each year, most of the people in this world become alert to know and understand their health status and information about diabetes. Besides, they are also generally willing to discuss about the therapies and any options they can take in order to prevent and cure this kind of disease. Diabetes can

The contents of the thesis is for internal user only

REFERENCES

Althoff, K. D., Bergmann, R., Wess, S., Manago, M., Auriol, E., Larichev, O. I., Bolotov, A., Zhuravlev, Y. I., & Gurov, S. I. (1998). Case-Based Reasoning for Medical Decision Support Tasks: The INRECA Approach. *Journal of Artificial Intelligence In Medical*, 12(1), 25-41.

Aamodt, A. (1993). Explanation-driven retrieval, ruse and learning of cases. University of Kaiserslautern SEKI Report S-93-12 (SFB 314) 279-284.

Aamodt, A, & Plaza, E. (1994) Case Based Reasoning: Foundational issues, Methodological Variations and system Approaches. AI Communications. IOS press, 7 (1), pp.39-59.

Berner, E. S., Webster, G. D., Shugerman, A. A., Jackson, J. R., Algina, J., Baker, A. L., Ball, E. V., Cobbs, C. G., Dennis, V. W., Frenkel, E. P., Hudson, L. D., Mancall, E. L., Rackley, C. E., Taunton, O. D. (1994). Performance of four computer based diagnostic systems.

Berkovsky, S. Y., & Ben-Asher, Y. (2004). UNSO: Unspecified Onntologies for Peer-to-peer Ecommerce Applications. In *Proc. Of the International Conference on Informatics, Turkey.*

Buckles, B., & Petry, F. (1982). A fuzzy model for relational databases. Int J. Fuzzy Sets Syst., Vol. 7, pp. 213–226.

Balaa, Z. E., Strauss, A., Uziel, P., Maximini, K., & Traphoner, R. (2003). FM-Ultranet: A Decision Support System using Case-based Reasoning, Applied to

Ultrasonography. *Proceedings of the International Conference on Case-Based Reasoning*, 37-44.

Campin, J., Paton, N., & Williams, M. (1997). Specifying active database systems in an object- oriented framework. Softw. Eng. Knowl. Eng. 7(1), 101–123).

Ceri, S. & Fraternali, P. (1997). Designing Applications with Objects and Rules: The IDEA Methodology. International Series on Database Systems and Applications, Addison- Wesley Longman, Reading, MA.

Chakravarthy, S. (1989). Rule management and evaluation: An active DBMS perspective. Sigmoid Rec. 18, 3, pp. 20–28.

Diaz, O. (1992). Deriving rules for constraint maintenance in an object-oriented database. In Proceedings of the International Conference on Databases and Expert Systems DEXA, I. R. A. M. Tjoa, Ed., Springer-Verlag, pp. 332–337.

Diagnostic Strategies (1999). Expert System Development Series Introduction to Case-Based Reasoning, Retrieved, 2007, from http://www. Diagnostic Strategies.com

Differences between type 1 and type 2 diabetes, (2006) Juvenile Diabetes Research Foundation. Retrieved June 18, 2007, from http://www.jdrf.org.au/publications/factsheets/differences_between_type_1_and_type_2.pdf

Eddy, D. M. (1990). The challenge. Journal of American Medical Association, 263, pp. 287–290.

Fernandez, I. B. & Aha, D. W. (1996). Case-Based problem solving for Knowledge Management system. In *Proceeding of the 12th Annual International Floroda*

Artificial Intelligence Research Symposium (FLAIRS): Knowledge Management Track. NCARAI Technical report AIC_99_005.

Gatziu, S., & Dittrich, K. (1994). Events in an active object-oriented database. In Proceedings of the First International Workshop on Rules in Database Systems, N. Paton and M. Williams, Eds., Springer-Verlag, pp. 23–39.

George, R., Srikanth, R., Petry, F. E., & Buckles, B. P. (1996). Uncertainty management issues in object-oriented database systems, IEEE Trans. Fuzzy Sys, vol. 4, pp. 179–192.

Hanson, E. N., & Widom, J. (1992). An overview of production rules in database systems. Tech. Rep., University of Florida, Department Computer and Information.

Hugo, C. H., & Tania, C. D., (2003). Analyzing the use of Dynamic Weights in Legal Case Based System. Edinburgh, Scotland, UK.

Hejlesen, Plogmann, S., & Cavan, D. (2000). DiasNet - An Internet Tool for Communication and Education in Diabetes. International Symposium on Computer and Diabetes Care.Rochester MN, September 8-10.

Kolodner, J. (1993). Case-Based Reasoning. San Mateo, CA: Morgan Kaufman Publishers.

Jaulent, M. C., Bozec, C. L., Zapletal, E., & Degoulet, P. (1997). A Case-based Reasoning Method for Computer-Assisted Diagnosis in Hisopathology. *Journals of Artificial Intelligence in Medicine*, 239-242.

Lin, T. Y., & Cercone, N. (1997). Rough Sets and Data Mining- Analysis of Imperfect Data, Kluwer Academic Publishers, Boston, London, Dordrecht, pp. 430.

Leake, D. B. (1996). Case-Based Reasoning: Experience, Lessons and future Direction. Menlo Park: AAAI Press.

Limthanmaphon, B., & Zhang, Z. (2002). Web Service Composition with Case-Based Reasoning. Department of Mathematic and Computing, University of Southern Quessnsland, Toowoomba, Australia.

Lopez, R. & Plaza, E. (1997). Case-based Reasoning: An Overview. AI Communication Journal, 10 (1), pp 21-29.

Liu, J. N. K., & Sin, D. K. Y. (1999). Evaluating case-based reasoning and evolution strategies for machine maintenance. Systems, Man, and Cybernetics, 1999. IEEE SMC '99 Conference Proceedings. 1999 IEEE International Conference on. IEEE (pp. 480 - 485 vol.2). (IEEE Document Reproduction Service No. 10.1109/ICSMC.1999.825308).

Li, K. & Liu, Y. (2002). Rough set based attribute reduction approach in data mining. Machine Learning and Cybernetics, 2002 Proceedings. 2002 International Conference on. Volume 1, 4-5 Nov. 2002 Page(s):60-63 vol.1. Digital Object Identifier 10.1109/ICMLC.2002.1176709.

Medical computing, Shortliffe EH. The science of biomedical computing. Med Inform 1984; 9:185-93 Retrieved June 10, 2007, from http://www.openclinical.org/healthinformatics.html

Montani, S., Portinale, L., Leonardi, G., & Bellazi, R. (2003). Applying Case-based Retrieval to Hemodialysis Treatment. *Proceedings of the International Conference on Case-Based Reasoning*, 53-62.

Miller, R. A. (1994). Medical Diagnosis Decision Support Systems-Past, Present, and Future. 1, 8-27. Retrieved June 20, 2007, from http://www.cpmc.columbia.edu/

Nillson, M., Funk, P., & Sollenborn, M. (2003). Complex Measurement Classification in Medical Applications Using a Case-Based Approach. *Proceedings of the International Conference on Case-Based Reasoning*, 63-72.

Oehrn, A. (1993). Rough logic control. In: (Project), Technical Report. Knowledge Systems Group, The Norwegian University of Science and Technology, Trondheim, Norway.

Pal, S. K., & Skowron, A. Fuzzy Sets, Rough Sets and Decision Making Processes. Springer-Verlag, Singapore (in preparation)

Pawlak, Z. (1992). Rough Sets: Theoretical Aspects of Reasoning about Data. Kluwer Academic Publishers, Norwell, MA, USA.

Pawlak, Z., Grzymala-Busse, J., Slowinski, R., & Ziarko, W. (1995). Rough sets. Commun. of ACM, 38, pp. 88-95.

Polkowski, L., & Skowron, A. (1998). Rough Sets in Knowledge Discovery, Physica-Verlag, l(2).

Perner, P., Gunther, T., & Perner, H. (2003). Airborne Fungi Identification by Case-based Reasoning. *Proceedings of the International Conference on Case-Based Reasoning*, 73-79.

Qiufen Qi Dalhousie University. Case-Based Reasoning (CBR) Process in Diagnosis, Retrieved June 13, 2007, from http://web.his.uvic.ca/rle/2004/QQi.ppt?

Riesbeack, C. K., & Schank, R. C. (1989). Inside CBR. Hillside, New Jersey, USA: Lawrence Erlbaum Associates.

Slowinski, R. (1992). Intelligent Decision Support. Handbook of Applications and Advances of the Rough Set Theory. Kluwer Academic Publishers, Boston, London, Dordrecht.

Stonebraker, M., & Kemnitz, G. (1991). The Progress next-generation database management system. Commun. ACM 34, Oct., pp. 78–92.

Smyth, B., & Keane, M., (1995). Experimental on Adaptation-Guided Retrieval in Case based Design. *In Topics in Case-Based Reasoning Proceedings of the International Conference on Case-Based Reasoning,* ICCBR95. LNAI series, Springer, Sesimbra, Portugal.

Tsumoto, S., Kobayashi, S., Yokomori, T., Tanaka, H., & Nakamura, A. (1996). The Fourth Internal Workshop on Rough Sets, Fuzzy Sets and Machine Discovery. The University of Tokyo.

Vaishnavi & Kuechler (2004). Design Research in information system. Retrieved June 15, 2007, from http://www.isworld.org/Researchdesign/drisISworld.htm

Widom, J. (1992). A denotational semantics forthe Starburst production rule language. Sigmod Rec. 21, 3, 4–9.

Ziarko, W. (1993). Rough Sets, Fuzzy Sets and Knowledge Discovery. Proceedings of the International Workshop on Rough Sets and Knowledge Discovery (RSKD'93), Banff, Alberta, Canada .October 12- 15, Springer-Verlag, Berlin.