AGENT-BASED KNOWLEDGE MANAGEMENT APPROACH FOR CLINICAL DECISION SUPPORT SYSTEMS

By

DINA ALI ALI SHARAF-EL-DEEN

Supervised by

Prof.Dr Mohamed Essam Khalifa
Professor in Basic Science department
Faculty of Computer and Information Sciences
Ain Shams University

Dr. Ibrahim Fathy Moawad
Associate Professor in Information Systems
department
Faculty of Computer and Information Sciences
Ain Shams University

A THESIS

Submitted in partial fulfillment of the requirements for the degree of
MASTER OF INFORMATION SYSTEMS
Department of Information Systems
Faculty of Computer and Information Sciences

AIN SHAMS UNIVERSITY
2014

Abstract

Case-Based Reasoning (CBR) and Rule-Based Reasoning (RBR) are two popular approaches used in decision support systems. Due to the complexities and the diversities of this domain, most medical systems become hybrid. Besides, the case adaptation process in CBR is often a challenging issue as it is traditionally carried out manually by domain experts.

Rules in RBR usually represent general knowledge, whereas cases in CBR encompass knowledge accumulated from specific (specialized) situations. Each reasoning approach has advantages and disadvantages, which are proved to be complementary in a large degree. Therefore, it is well-justified to combine rules and cases to produce effective hybrid approaches, surpassing the disadvantages of each component method.

To take advantage of the knowledge managing and reasoning capabilities in CBR and RBR, the main research objective is to present an agent-based knowledge management approach for clinical decision support systems. This approach is presented as a web service to be easily utilized. It integrates CBR and RBR, and applies the adaptation process automatically. Both adaptation rules and reasoning rules are generated from the case-base. After solving a new case, the case-base is expanded, and both adaptation and reasoning rules are updated. Besides, a prototype was implemented and experimented to diagnose breast cancer and thyroid diseases, and to classify IRIS plant type. The evaluation results showed that this approach increases the accuracy of retrieval only CBR systems, and achieves great accuracy compared to the current mammography based breast cancer diagnosis systems, thyroid diagnosis systems, and IRIS plant type classification systems.

ملخص الرسالة

يعتبر المنطق القائم على الحالة (CBR)، والمنطق القائم على القواعد (RBR) نهجان ذو شعبية في نظم دعم اتخاذ القرار. بسبب تعقيدات وتنوع المجال الطبي، فإن معظم النظم الطبية أصبحت تتكون من أكثر من نهج . الى جانب ذلك، تعتبر عملية التكيف في حالة CBR مسألة صعبة لأنها تتم عادة يدويا من قبل خبراء المجال.

تمثل القواعد في RBR المعرفة العامة، في حين أن الحالات في CBR تشمل المعرفة المتراكمة من الحالات (المتخصصة) المحددة. كل طريقة منهما لها مزاياها وعيوبها، والتي ثبت أنها مكملة لبعضها البعض إلي درجة كبيرة. وبالتالي، فمن الجيد الجمع بينهما لانتاج نهج مجمع ، متجاوزا عيوب كل أسلوب منهما.

للاستفادة من إدارة المعلومات وقدرات حل المشكلات في CBR وRBR، فإن الهدف البحثي الرئيسي هو تقديم منهج إدارة المعرفة القائم على استخذام الوكيل البرمجي من أجل أنظمة دعم اتخاذ القرار الإكلينيكية. ويقدم هذا النهج كـ Web service لاستخدامها بسهولة. أنه يدمج CBR وRBR، ويتم تطبيق عملية التكيف تلقائيا. يتم إنشاء كل من قواعد التكيف وقواعد المنطق من قاعدة البيانات للحالات. بعد حل حالة جديدة، يتم توسيع قاعدة البيانات للحالات ، ويتم تحديث كل من قواعد التكيف وقواعد المنطق. الى جانب ذلك، تم تنفيذ برنامج أولي وتمت تجربته لتشخيص سرطان الثدي وأمراض الغدة الدرقية، وتصنيف نوع النبات IRIS . وأظهرت نتائج التقييم أن هذا النهج يزيد من دقة نظم CBR التي تقوم بالاسترجاع فقط، ويحقق دقة كبيرة بالمقارنة مع النظم الحالية لتشخيص سرطان الثدي القائم على التصوير الشعاعي للثدي ، وأنظمة تشخيص الغدة الدرقية، وأنظمة تصنيف نوع النبات IRIS.

To my father and my mother

To my husband and my daughter

Acknowledgement

I would like to express my sincere gratitude to my supervisors, for their valuable guidance and helping me with their fruitful ideas. This thesis work would not have been possible without their support and help. I am really fortunate for having the opportunity to work with them. I would like to thank my supervisor, Dr Mohamed Essam Khalifa, for his great suggestions, encouragement, and support. Also, I would like to thank my supervisor, Dr Ibrahim Fathy Moawad, for his supportive supervision, his fruitful ideas, and his great cooperation in writing the papers. Also, I want to acknowledge all members in Ain Shams University at Information Systems department for their encouragement to start my research work. Finally, and most importantly, I want to thank and acknowledge my family: my parents, my husband, my daughter, and my sisters who have inspired me and supported me always in many ways.

List of Publication

Papers published from this thesis

- **Paper 1:** Learning and Applying Range Adaptation Rules in Case-Based Reasoning Systems, **International conference(2012)**.

  Sharaf-Eldeen, D. A., Moawad, I. F., El Bahnasy, K., & Khalifa, M. E. (2012). Learning and Applying Range Adaptation Rules in Case-Based Reasoning Systems. In Advanced Machine Learning Technologies and Applications (pp. 487-495). Springer Berlin Heidelberg.

- **Paper 2:** A Breast Cancer Diagnosis System using Hybrid Case-based Approach, ***International journal (2013)****, International Journal of Computer, 0.791 (2013 impact factor).*

Sharaf-el Deen, D. A., Moawad, I. F., & Khalifa, M. E. (2013). A Breast Cancer Diagnosis System using Hybrid Case-based Approach. International Journal of Computer Applications, 72(23), 14-20.

- **Paper 3:** A New Hybrid Case-Based Reasoning Approach for Medical Diagnosis Systems. Journal of medical systems, ***International journal (2014),*** *Journal of medical systems, Indexed by SCOPUS, 1.372 (2013 impact factor).*

Sharaf-El-Deen, D. A., Moawad, I. F., & Khalifa, M. E. (2014). A New Hybrid Case-Based Reasoning Approach for Medical Diagnosis Systems. Journal of medical systems, 38(2), 1-11. (*Indexed by SCOPUS, Impact factor: 1.783*)

List of Abbreviations

AIRS	Artificial Immune Recognition System
C4.5-1	C4.5 with default learning parameters
C4.5-2	C4.5 with parameter c equal to 5
C4.5-3	C4.5 with parameter c equal to 95
CBR	Case-Based Reasoning
CSFNN	Adaptive Conic Section Function Neural Network
DIMLP	DIMLP with two hidden layers and default learning parameters
DSS	Decision Support Systems
ESTDD	Expert System for Thyroid Disease Diagnosis with NeuroFuzzy Classification
FS-PSO-SVM	Fisher Score - Particle Swarm Optimization - Support Vector Machine
GDA-WSVM	Generalized Discriminant Analysis (GDA) and Wavelet Support Vector Machine
LDA	Linear Discriminant Analysis
LVQ	Learning Vector Quantizer
MLNN with LM	Multi-Layer Perception with Levenberg-Marquardt (LM) algorithm
MLP	Multi-Layer Perception
MLP with PB	Multi-Layer Perception with Back-Propagation
PNN	Probabilistic Neural Network
PPFNN	Probabilistic Potential Function Neural Network
RBF	Radial Basis Function
RBR	Rule-Based Reasoning
RST	Rough Set Theory

List of Figures

List of Tables

Table of Contents

Chapter 1: Introduction

In this chapter, the research problem is defined. It discusses the research objective and shows the research contributions. Also, the thesis organization is presented.

1.1.1 Problem definition

Nowadays medical knowledge is expanding rapidly to the extent that even experts have difficulties in managing and following all the new results, changes, and new treatments. Decision support systems (DSS) that bear more similarities with human reasoning are often easily accepted by physicians in the medical domain (Ahmed et al. 2012). Moreover, recent DSS tend towards the hybrid integration containing two or more intelligent techniques (Sahin et al. 2012).

Case-based reasoning (CBR) systems are valuable examples of decision support systems (Bichindaritz & Montani 2011). Besides, CBR can be considered a very well suited reasoning paradigm for managing medical knowledge. It has the capability of incrementally collecting, reusing, and sharing the knowledge, which is implicitly embedded in previously situations (Montani 2011). Also, Rule-Based Reasoning (RBR) is historically one of the most successful approaches to deal with such knowledge.

CBR has been successfully applied in the medical domain (Kolodner 1993; Hunt & Miles 1994; Macura & Macura 1997; Lenz et al. 1998; Baumeister et al. 2002). However, adaptation is often a challenging issue, because it is traditionally carried out manually by domain experts (Begum et al. 2011). Moreover, most CBR systems that do not apply adaptation (retrieval only CBR systems) fail to solve some of new problems, and hence they do not provide convincing accuracy in critical domains like medical.

1.2 Research objective

CBR and RBR are two popular approaches used in decision support systems. Rules usually represent general knowledge, whereas cases encompass knowledge accumulated from specific (specialized) situations. Each approach has advantages and disadvantages, which are proved to be complementary in a large degree. Therefore, it is well-justified to combine rules and cases to produce effective hybrid approaches, surpassing the disadvantages of each component method (Prentzas & Hatzilygeroudis 2007).

To take advantage of the knowledge managing and reasoning capabilities in CBR and RBR, the main research objective of this thesis is to present an agent-based knowledge management approach for clinical decision support systems. This approach is presented as a web service to be easily utilized. It integrates CBR and RBR, and applies the adaptation process automatically. Besides, a prototype was implemented and experimented to diagnose breast cancer and thyroid diseases, and to classify IRIS plant type. The evaluation results showed that this approach increases the accuracy of the retrieval only CBR systems and achieves great accuracy compared to the current mammography based breast cancer diagnosis systems, thyroid diagnosis systems, and IRIS plant type classification systems.

1.3 Research contributions

In this research work, an agent-based knowledge management model was proposed for clinical decision support systems. This approach is presented as a web service to be easily utilized. It integrates CBR and RBR, and applies the adaptation process automatically. To achieve the integration, a new process is added to the **REUSE/ADAPT** step of the classical CBR cycle called **REASON** at which the reasoning rules are applied to infer a solution if both **REUSE** and **ADAPT** processes failed to find a solution. The proposed model contains two phases: *Knowledge Extraction Phase* and *Problem Solving Phase*. During the Knowledge Extraction Phase, both adaptation and reasoning rules are extracted from the case-base. After that, the adaptation rules and the reasoning rules are exploited and updated dynamically in the Problem Solving Phase. The adaptation rules extraction process was introduced in (Sharaf-Eldeen et al. 2012), where the reasoning rules extraction process was introduced in (Sharaf-el Deen et al. 2013). The full hybrid model and the two phases were introduced in (Sharaf-El-Deen et al. 2014).

To sum up, the main contributions of this thesis can be summarized as follow.

- Proposing a new hybrid case-based reasoning model for medical diagnosis systems by integrating CBR and RBR approaches.
- Adaptation rules are extracted from the case-base automatically. On the other hand, rough set theory is used to extract the reasoning rules. All these rules are exploited and updated automatically.

- Achieving acceptable accuracy by experimenting the proposed model on different three datasets: IRIS plant type, breast cancer disease, and thyroid disease datasets.

1.4 Thesis organization

The rest of this thesis is organized as follows.

- Chapter 2 reviews the theoretical background upon which the research is based on. It describes briefly the case-based reasoning, the rule-based reasoning, the rough set theory, and the intelligent agents.
- Chapter 3 presents an overview of the related work using two viewpoints. The first viewpoint presents the recent medical CBR systems while the second viewpoint presents the medical diagnosis systems and IRIS classification systems that we are going to compare our research work with.
- Chapter 4 introduces the design of the proposed agent-based knowledge management model and both hybrid case-based model architecture, and its phases. In addition, a diagnosis case study of the thyroid disease is presented to show how the proposed model works.
- Chapter 5 illustrates the prototype implementation and the experimental datasets. Also, it shows the evaluation of the experimental results.
- Finally, chapter 6 concludes the thesis and reviews the future work.

Chapter 2: Background

This chapter reviews the theoretical background upon which this research is based on. It describes briefly the case-based reasoning, the rule-based reasoning and the rough set theory, and the intelligent agents.

2.1 Case-based reasoning

Case-based reasoning (CBR) is a reasoning methodology that simulates human reasoning using past experiences to solve new problems (Kolodner 1993). Generally, the problem solving cycle of the classical CBR model consists of four steps (Aamodt & Plaza 1994) as shown in figure 2.1:

(1) **RETRIEVE** step that is responsible for retrieving one or more similar cases to the new case.

(2) **REUSE/ADAPT** step that is responsible for reusing the solution of the most similar case to the new case. It may include the adaptation task in which the solution of the retrieved case is adapted to fit the new case.

(3) **REVISE** step that is responsible for revising the suggested solution for confirmation.

(4) **RETAIN** step that is responsible for retaining the learned case for future use.

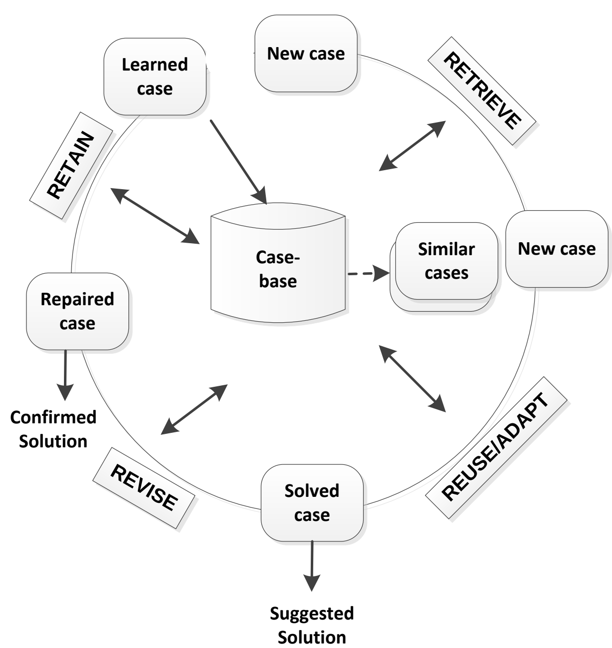

Figure 2.1: Classical case-based reasoning cycle (Aamodt & Plaza 1994).

In general, there are a number of merits of using CBR (Shiu & Pal 2004)***:***

- **Avoid repeating mistakes made in the past**. In systems that record failures as well as successes, and perhaps the reason for those failures, the information about what caused failures in the past can be used to predict potential failures in the future.
- **Provide flexibility in knowledge modeling**. Model-based systems, due to their rigidity in the problem formulation and modeling, sometimes cannot solve a problem, which is on the boundaries of their knowledge or scope, or when there is some missing or incomplete data. In contrast, case-based systems use the past experiences as the domain knowledge and often provide a reasonable solution, through appropriate adaptation, to these types of problems.
- **Reason in domains that have not been fully understood, defined, or modeled**. In situation where insufficient knowledge exists to build a causal model of a domain or to derive a set of heuristics for it, CBR system can still be developed using only a small set of cases from the domain. The underlying theory of the domain knowledge does not have to be quantified or understood entirely for a CBR system to function.
- **Make predictions of the probable success of a derived solution**. When information is stored regarding the level of success of past solutions, the CBR system may be able to predict the success of the suggested solution to a current query problem. This is done by referring to both the stored level of solutions success and the differences between the previous and current contexts of applying these solutions.
- **Learn over time**. As CBR systems are used, they encounter more problem situations and create more solutions. If solution cases are subsequently tested in the real world, and a level of success is determined for those solutions, then these cases can be added into the case-base, and then used to help solving future problems. As cases are added, a CBR system should be able to reason in a wider variety of situations, and with a higher degree of refinement and success.
- **Reason in a domain with a small body of knowledge**. While in a problem domain for which there is only a few cases available, a CBR system can start with these few known cases and incrementally build its knowledge as cases are added. The addition of new cases will cause the system to expand in directions that are determined by the cases encountered in its problem solving endeavors.