Using Knowledge and Rule Induction Methods for Enhancing Clinical Diagnosis: Success Stories

Fariba Shadabi
Faculty of Information Sciences and Engineering
University of Canberra, ACT, 2601, Australia
Fariba.Shadabi@canberra.edu.au

Dharmendra Sharma
Faculty of Information Sciences and Engineering
University of Canberra, ACT, 2601, Australia
Dharmendra.Sharma@canberra.edu.au

Abstract— *The economic and social benefits of accurately predicting medical outcomes are very high. As a result, the problem of improving predictive models has attracted many researchers. Over the past few years there has been great interest in the use of advance knowledge discover techniques to mimic human functions. Research shows that such techniques can be applied in healthcare environments where an automated process must improve its performance based on previous data, adapt to changes and deal with uncertain and incomplete medical knowledge. The underlying purpose of this paper is to illustrate the utility of combining multi agent approach and hybrid machine learning and data mining techniques for producing predictive classifiers in clinical settings, through a few real world success stories.*

Keywords- multi agent and data mining

I. INTRODUCTION

In recent years there has been a rapid growth in the successful use of advanced information systems and knowledge-based techniques in many diverse areas such as science, medicine and commerce. The main contribution factor for the development of such systems in medicine has been the extra demands for a more powerful yet flexible and transparent technique to cope with the extensive amount of data and knowledge stored in clinical databases. Medical centres, hospitals and other health care providers are using knowledge systems and advanced information systems such as data mining and artificial intelligent to improve medical care, predict diagnosing patterns, provide reasoning for knowledge management and share information about medical diagnosis.

Advanced information systems and knowledge-based techniques have been involved in clinical decision support systems, mining medical databases, effective diagnostic and treatment methodologies and intelligent systems for hospital, home and community health care. The computational technologies for the intelligent systems have mainly been soft computing, which includes, rule based systems, fuzzy theory, neural networks, neuro-fuzzy systems, data mining and evolutionary algorithms.
Machine learning and data mining techniques are combined with prior knowledge to overcome limits of small sample size and noisy data. This research illustrates the utility of combining such techniques with multi Agents system through a few real world examples for producing simple predictive classifiers.

II. APPLICATION

A. Multi-Agents Approach to Knowledge Discovery

Agents are software entities that use Artificial Intelligent techniques to perform a set of actions to attain the goal of the users. Multi Agents system (MAS) is a set of agents collectively and collaboratively working to perform a task that cannot be done by an individual agent. agents can interact with other external systems and can be used to manage both distributed and local knowledge. This is an important feature since the E-health knowledge is usually generated by different sources and often from different places. For instance, a consultation agent can work with diagnosis agent in order to provide a better answer to the enquiry of patients. Multi agents can be used to solve problems that are too large for a centralized agent to solve because of resource limitations and/or to avoid a one point bottleneck or failure point [1], [2].

Figure 1 demonstrates the use of a real time data mining cooperative multi agents system called DMMAS. DMMAS is a multiagent system with multiple miner agents and a combination agent as agent manager. The main goal the system is to explore how data partitioning and multi agent approach can help to improve the efficiency and also if possible the accuracy of chronic diseases management and prediction tasks in real time. Our initial experimental results have shown promising results, in this case, using the UCI diabetes data set [3]. It should be noted that the application domain is not limited to diabetes data and it can be applied

978-0-7695-3591-3/09 $25.00 © 2009 IEEE
DOI 10.1109/ICFCC.2009.140

across many other domains such as cancer data for improving the quality of care.

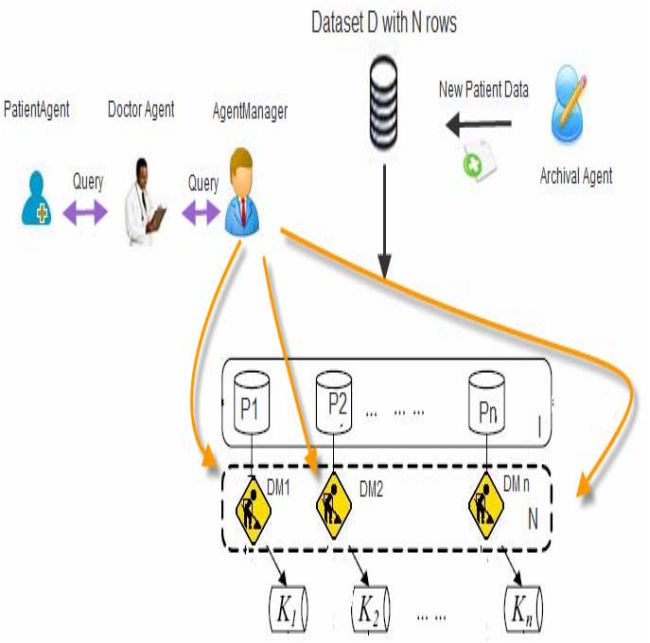

Figure 1: The proposed DMMAS approach with multiple miner agents and a combination agent as a personal agent [8]

B. Hybrid Machine learning techniques in Kidney Transplant Management System

This section demonstrate how advance machine learning technique can be use to handle the complexity of clinical data and rule extraction from individual black box classification models such neural networks by studying the data. We demonstrate this through a novel hybrid-learning model called RIDC-ANNE [5].

This novel algorithm was successfully developed with reference to the data filtering, bagging-based ensemble and the hybrid decision tree-neural networks ensemble theories [4],[5]. The RIDC-ANNE approach assists the data preparation process by configuring an ensemble of bagged networks as a filter and identifying the regions in the data space that have been consistently misclassified or have high impact on the system performance. The RIDC-ANNE technique treats the ensembles as a black box or data filtering tool in order to provide a body of new training examples to feed standard rule induction or decision tree systems. This strategy considers the diversity and expertise of the component networks in the rule generation process. However, unlike other rule generation techniques, it does not focus on the identifying ensemble members that are relevant in explaining the prediction (output) associated with a particular case. Instead RIDCANNE explains the output of the ensemble based on a cluster of cases that consistently generate agreement across the classifiers with similar expertise. The use of RIDCANNE not only produces rule sets that are substantially easier to understand but it also reduces the computational cost, since it creates a cluster of new training examples that allow faster rule generation process.

The RIDC-ANNE approach [5] uses the Bagging algorithm to train neural networks and extracts the input patterns (examples) that were included across the ANN series in the final results. In this strategy for each of the patterns or examples in the validation set, the votes were recorded first. Then the system searched for patterns or examples that caused at least 50% of the networks to agree on one category (fail or success). Then, it searched for patterns that caused 51% agreement among the networks. This procedure was repeated until the model reached the 100% agreement among the networks for one category (fail or success). It should be noted that as the model approaches the 100% agreement condition, fewer patterns were able to satisfy the condition and be included in the report. Finally, for each condition (i.e. network agreement), the prediction accuracy and the number of examples that satisfy the condition were reported by the system.

This investigation revealed that by using 19% of data points (84 cases), the validation set can reach a 70.2% accuracy rate (with 87% agreement among the networks). Furthermore, a close study of the rule set generated in this scenario revealed that the donor state and age were considered two important factors in determining the success of kidney transplants [5]. The investigation also revealed great success in the application of two well-known medical datasets, Pima Indian Diabetes and Wisconsin Cancer [3].

Furthermore, Figure 2 demonstrate how multi-agent, data mining, and RIDC-ANNE approaches can be combined to generate knowledge models for agents. The Data Miner (DM) is the core component of this prototype and can be used to provide the user with a number of DM algorithms [2]. The DM algorithms are then validated before they can be store as a set of facts in knowledge based models of a rule engine. For Example, as described in above the generation of the rules can be done through RIDC-ANNE.

We are currently expanding the above ideas for using on clinical pathology databases for recognizing patterns that will identify small groups in the population that are at higher relative risk of infection by the respiratory pathogens Chlamydia pneumoniae (Cp) and/or Mycoplasma pneumoniae (Myco). Once rules have been discovered, they

can be used to develop a system (e.g. a multiagent) of electronic alerts that will identify patterns of pathology results that identify a patient at a high relative risk of a true infection through enhancing the predictive capacity of combined laboratory test results [9].

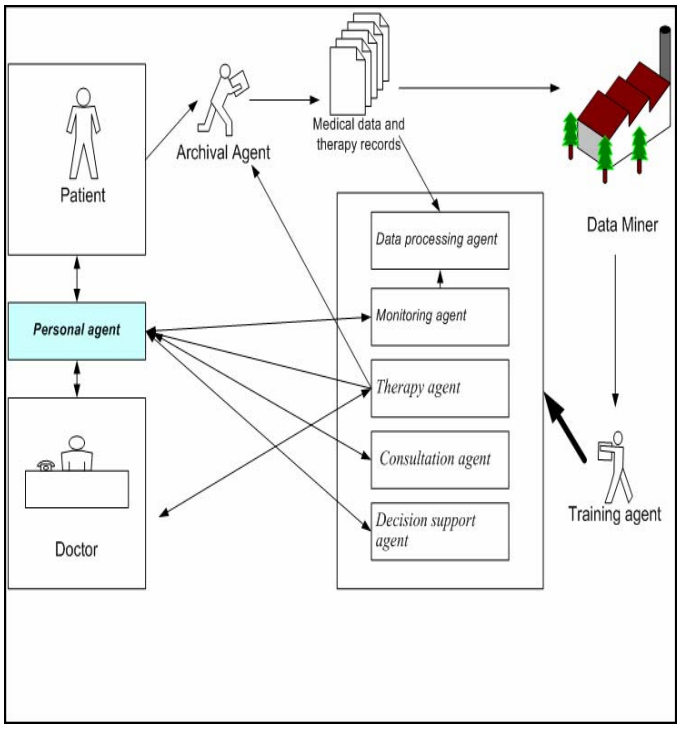

Figure 2. Combining the RIDC-ANNE and Multi Agent Technology [2]

III. CONCLUSION

This paper reviewed the potential impact of advanced information systems and knowledge-based techniques on clinical medicine, by referring to a few real world success stories. Procedures such as patient diagnosis, treatment and the storage, retrieval and analysis of clinical information can all be assisted or improved by such technologies.

Novel data mining and machine learning technologies can be used to assist medical services. However, it may take many years and thousands of such trials to fully demonstrate the potential of integrating advanced data mining modules into computerized patient records and clinical findings for improving the quality of care.

REFERENCES

[1] Ira. S. R. 2004, INTELLIGENT AGENTS. *Communications of the Association for Information Systems Volume14, 2004*pp. 275-290.

[2] Sharma D. and F. Shadabi, An Intelligent Multi Agent Design in Healthcare Management System- Lecture Notes in Computer Science, Volume 4953/2008, pp. 674-682, 2008.

[3] IDI. International diabetes institute - diabetes research, education and care. *2007(10/30/2007),*

[4] Mitchell, T. (1997), *Machine Learning.* (McGraw-Hill, New York).

[5] Shadabi, F. (2007), "Medical Outcome Prediction: A Hybrid Artificial Neural Networks Approach" School of Information Sciences and Engineering, University of Canberra, Australia.

[6] Cheung, N. (2001), "Machine Learning Techniques for Medical Analysis" Honours Thesis, School of Information Technology and Electrical Engineering, University of Queensland,

[7] Duch, W., Adamczak, R., and Grabczewski, K. (2001), "A new methodology of extraction, optimization and application of crisp and fuzzy logical rules", *IEEE Transactions on Neural Networks* **12:** 277-306.

[8] Cuong T., D. Sharma and F. Shadabi, A Multi-Agents Approach to Knowledge Discovery, International conference on Intelligent Agent Technology(IAT'08), 9-12 December, Sydney, Australia

[9] Richardson, A., Hawkins, S. ,Shadabi, F., Sharma, D., Fulcher, J and Lidbury, B.A, "Enhanced Laboratory Diagnosis of Human Chlamydia pneumoniae Infection through Pattern Recognition Derived from Pathology Database Analysis" Third IAPR International Conference on Pattern Recognition in Bioinformatics (PRIB 2008) ,October 15 – 17, Melbourne , Australia