Available online at http://www.journalcra.com

International Journal of Current Research
Vol. 8, Issue, 01, pp.25174-25178, January, 2016

INTERNATIONAL JOURNAL OF CURRENT RESEARCH

ISSN: 0975-833X

RESEARCHARTICLE

AGENT-BASED SOFTWARE ARCHITECTURE FOR DECISION MAKING IN THE DATA WAREHOUSE

[1,*]Kalisa Wilson, [2]Ndatinya Eustache and [1]GakizaCanisius

[1]School of Information Science and Engineering, Xiamen University, Xiamen China
[2]Hunan University Collage of Foreign studies, Hunan China

ARTICLE INFO

Article History:
Received 11th October, 2015
Received in revised form
26th November, 2015
Accepted 20th December, 2015
Published online 31st January, 2016

Key words:

Agent-based architecture,
Multi-Agent, DSS, Data mining.

ABSTRACT

Using Agents to extract data from existing sources of information is a key development area to unlock previously unknown relationships between heterogeneous data sources. It becomes an issue when large volumes of data, such as in the case of data from different sources like SQL, ORACLE,IBM-DB2,MySQL,FLAT FILES and others need to be analyzed. The main focus of this paper is on how data warehousing and mining techniques can be applied in knowledge discovery. We therefore proposed architecture for heterogeneous data integration based on Multi-Agent Driven Rule-based Decision Support System in data warehousing. With our predictive model, we achieved high accuracies of over 83% compared to (65%-72%) and (66%-72%) achieved by Clark, P & Niblett, T and Michalski, R.Set. al respectively on the Breast Cancer Wisconsin dataset.

Citation: **Kalisa Wilson, Ndatinya Eustache and Gakiza Canisius, 2016.** "Agent-based software architecture for decision making in the data warehouse", *International Journal of Current Research,* 8, (01), 25174-25178.

1. INTRODUCTION

The issue of conciliating data coming from multiple, heterogeneous information sources such as SQL, DB2, Oracle, and MySQL, etc is an old acquaintance (Bent and Graham, 2008). There has always been this necessity and the appearance of new, larger and richer information sources only have made things worse. It is no longer just a matter of "digesting" flat files, paper documents and databases. The implantation and wide acceptance of the Information Systems within organizations impelled the generation and collection of huge, diverse volumes of data, as well as, the need for convenient analysis and consequently, their adequate processing (Naumann, 2001) (Guess 2000) (Lourenço *et al.,* 2004).

There are many kinds of abnormal situations, both instance and schema-related. From a single-source perspective, issues such as misspellings, duplicates, contradictory values, and the lack of integrity constraints are to be attended to. However, the problem rises when trying to integrate multiple and heterogeneous sources. Data warehousing scenarios (Goldring, 2002) are the perfect example as they are inherently heterogeneous. When the data is extracted from the sources, it is inspected and processed in order to obtain a single, homogeneous, consistent data volume, ready to be integrated into the Data warehouse. Agent technology has proven to be particularly useful in several applications especially in data processing. The DSS in Data warehouse approach requires some steps to prepare and integrate data for the final user decision making (Mazón, 2008). The goal of this paper is to present our new agent-based architecture and data mining techniques to improve the decision making process using agent paradigm. The paper is structured as follows, section 2 explains an agent is and the language used to communicate between them. Section 3 explains the DSS and gives the general structure of the DSS while Section 4 shows the experimental results and finally section 5 concludes and outlines the future work.

****Corresponding author: Kalisa Wilson,***
School of Information Science and Engineering, Xiamen University, Xiamen China.

Related works

We can recall that some of the authors defined the decision support system (DSS) before as an interactive computer-based information system that is designed to support solutions on decision problems (Marakas and George, 2003). Some of the related ideas in DSS can be traced back to preceding work in two main research streams: theoretical study of organizational decision making undertaken by Simon *et al.* (Eckermann*et al.,* 2007). at the Carnegie Institute of Technology during late 1950s and early 1960s, and technical work on interactive computer systems carried out by Gerrity *et al.* (1981) at the MIT in 1960s.

2. Agent Based Approach

Agents can be defined to be autonomous, problem-solving computational entities capable of effective operation in dynamic and open environments (Luck, 2008). Agents are often deployed in environments in which they interact, and maybe cooperate, with other agents (including both people and software) that have possibly conflicting aims. Such environments are known as multi-agent systems (Bellifemine, Fabio, Agostino Poggi, 2001). Agents can be distinguished from objects (in the sense of object oriented software) in that they are autonomous entities capable of exercising choice over their actions and interactions. Agents cannot, therefore, be directly invoked like objects. However, they may be constructed using object technology. These notions find application in relation to several distinct aspects, considered in turn below.

2.1. Agent communication languages

Agents need information about the system in order to work efficiently. Information can be gained by communication between agents. Agents need to understand something about the agent communication language to communicate with each other.Agent communication languages are used for agent system interaction between communicative agents. Agent communication language is an important part in building the distributed agent-oriented information systems. When agents use defined agent communication language, their communication is uncomplicated. FIPA ACL (The Foundation for Intelligent Physical Agents Agent Communication Language) (Singh and Munindar, 2000) and KQML (Knowledge Query and Manipulation Language) (Bollacker, 2008) are languages that allow agents to communicate.

3. Decision Support System (DSS)

Decision support systems (DSS) are a key to gaining competitive advantage. Many corporations have built or are building unified decision-support databases called data warehouses on which decision makers can carry out their analysis. Computer-based system that aids the process of decision making, data, documents, knowledge and/or models to identify and solve problems and make decisions

3.1. Predictive model

Depending upon the type of data, the objective is to infer from a collection of those data or dataset to facilitate decision-making processes (Forns, 2002). Black-box which can predict the future based on information from past and present.In our work we present the dataset as Breast cancer historical data. Example, we take Age, Menopause, tumor-size and others to be historical data and then we build a model which predicts that either the patient is having a breast cancer or Not. Actually, the outcomes are given a YES or NO answer.

The choice has been clear for our task to focus on: classification, ourDecision tree classifiers offers a highly practical method for generalizing classification rules.Decision tree classifier offers a highly practical method for generalizing classification rules.

3.3 Rules

IF <Antecedent>
THEN <Consequent>
IF <<DEG_MALIG = 2) AND (TUMOR_SIZE = 40-59>>
THEN <no>

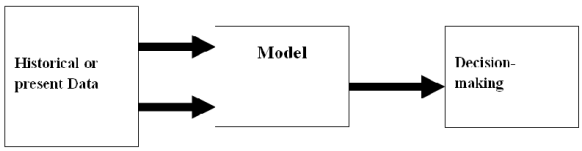

Fig. 3.1. General Structure of a model

Decision Tree

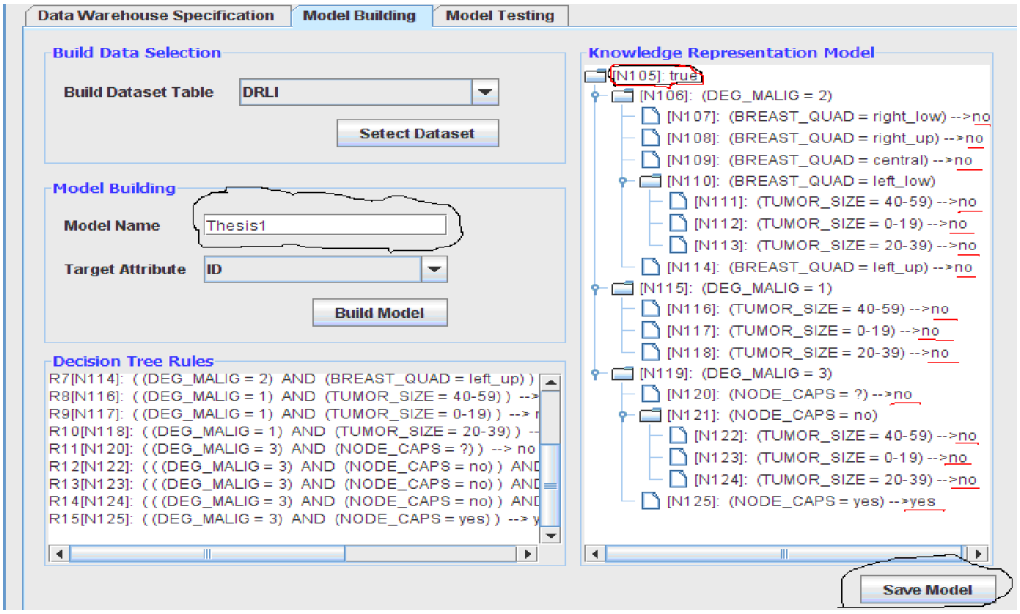

Fig. 3.2 Converting decision tree into Classification rules

IF <<<DEG_MALIG = 2>AND<TUMOR_SIZE = 0-19>> AND <MENOPAUSE = premeno>>> THEN <no>
IF << (DEG_MALIG = 2>AND <TUMOR_SIZE = 0-19>> AND <MENOPAUSE = ge40>>THEN <no>
IF <<TUMOR_SIZE = 20-39>AND <IRRADIAT = yes>> THEN < yes>
IF <<<DEG_MALIG = 3>AND<NODE_CAPS = yes>>AND <BREAST_QUAD = right_low>>> THEN <yes>

3.4 Model Testing

On way to test the prediction independently of the decision is to consider the four cases:

- **True positive** (*tp*) or *f*++, which corresponds to the number of positive instances correctly predicted by the classification model.
- **False negative** (*fn*) or *f*+–, which corresponds to the number of positive instances wrongly predicted as negative by the classification model.
- **False positive** (*fp*) or *f*–+, which corresponds to the number of negative instances wrongly predicted as positive by the classification model.
- **True negative** (*tn*) or *f*—, which corresponds to the number of negative instances correctly predicted by the classification model.

3.5. Rule-based DSS Implementation

3.5.1. System Implementation Diagram

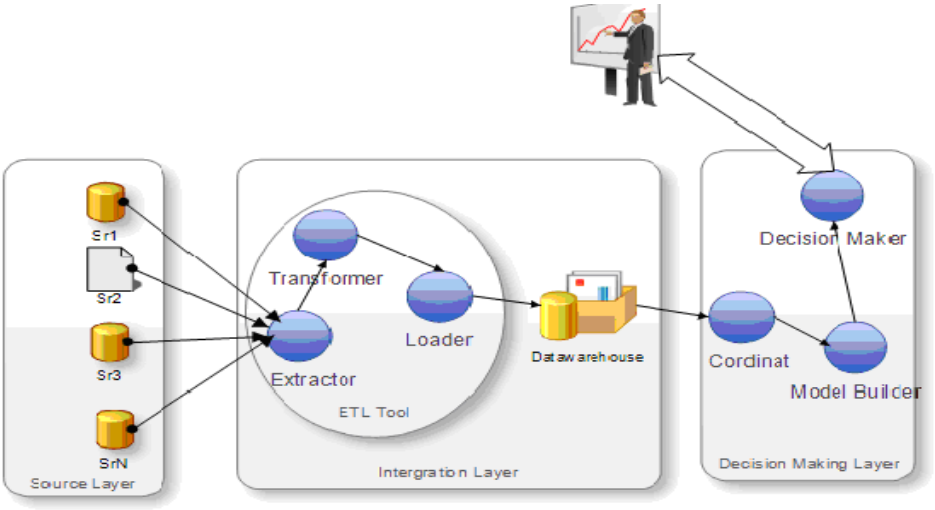

Fig. 3.5.1. System architecture in 3-layers layout

3.5.2. System Implementation

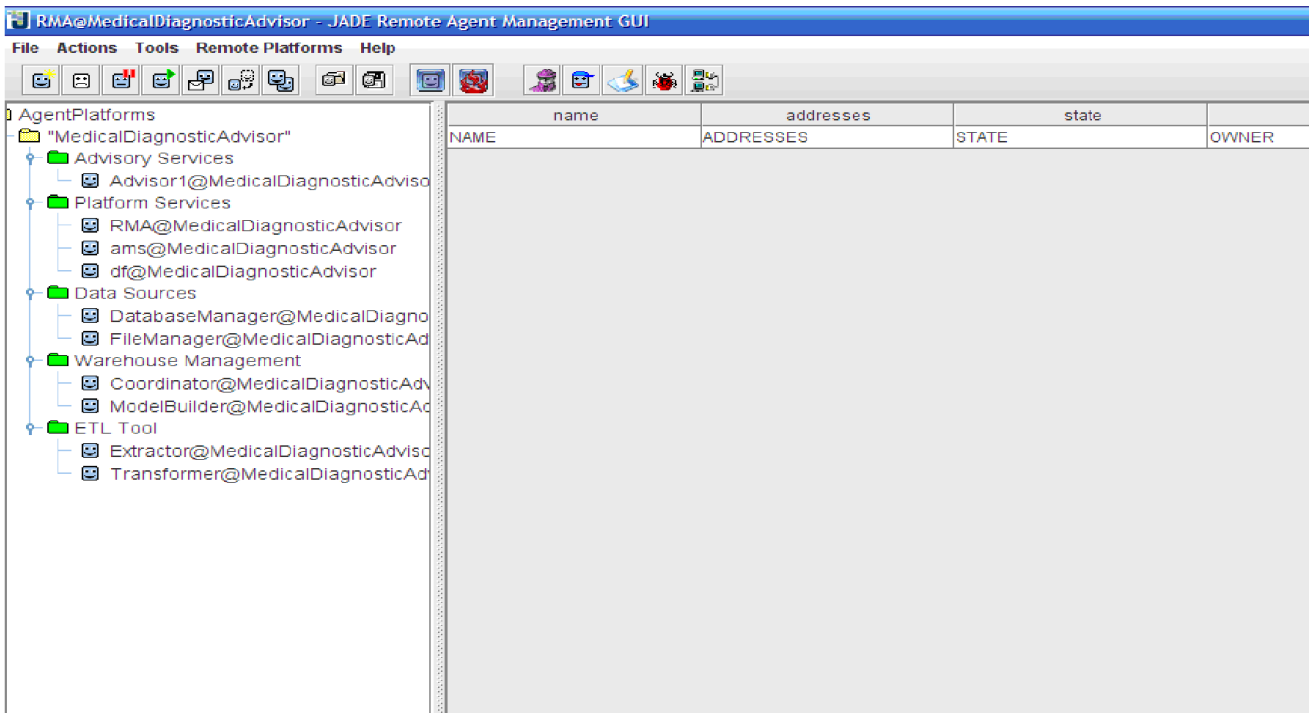

Fig.3.5.2. The Jade Remote Monitoring Agent showing the implementation Breast Cancer Advisory system layout architecture

4. Experiment Results

In our results we used the comparison of sensitivity and specificity in a table and an RCO curve. The term sensitivity is how good the test is at picking out the rule in correct way. It is simply the True Positive Fraction. In other words, sensitivity gives us the proportion of cases picked out by the test, relative to all cases that actually have the disease. Specificity is the ability of the test to pick out patients who do NOT have the cancer. In other words it is synonymous with the True Negative Fraction. The following figure shows our model testing which produce a high accuracy and hence reliable to predict the Breast cancer.

The accuracy is given by following formula:

True Positive (TP) is true cancer cases predicted correctly true. False Positive (FP) is false cancer cases predicted wrongly to be true. False Negative (FN) is the false cancer cases predicted wrongly to be negative. If we recall from the above table we can see that, TP=11, FP=2, TP=68, FN=14.

(TP+TN)/ (TP+FP+FN+TN) *100%

(11+68)/ (11+2+14+68)*100%=83%

Another way of proving right our model is the user of Receiver Operating Characteristics (ROC).

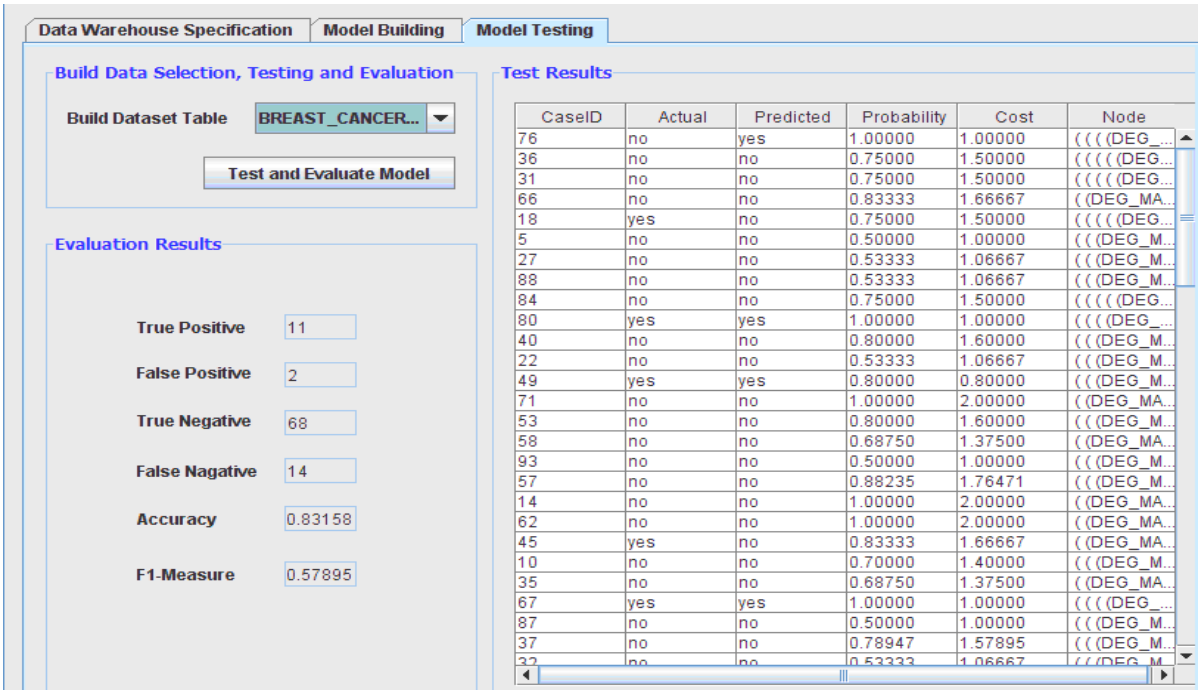

Fig.4.1 The accuracy of the model built

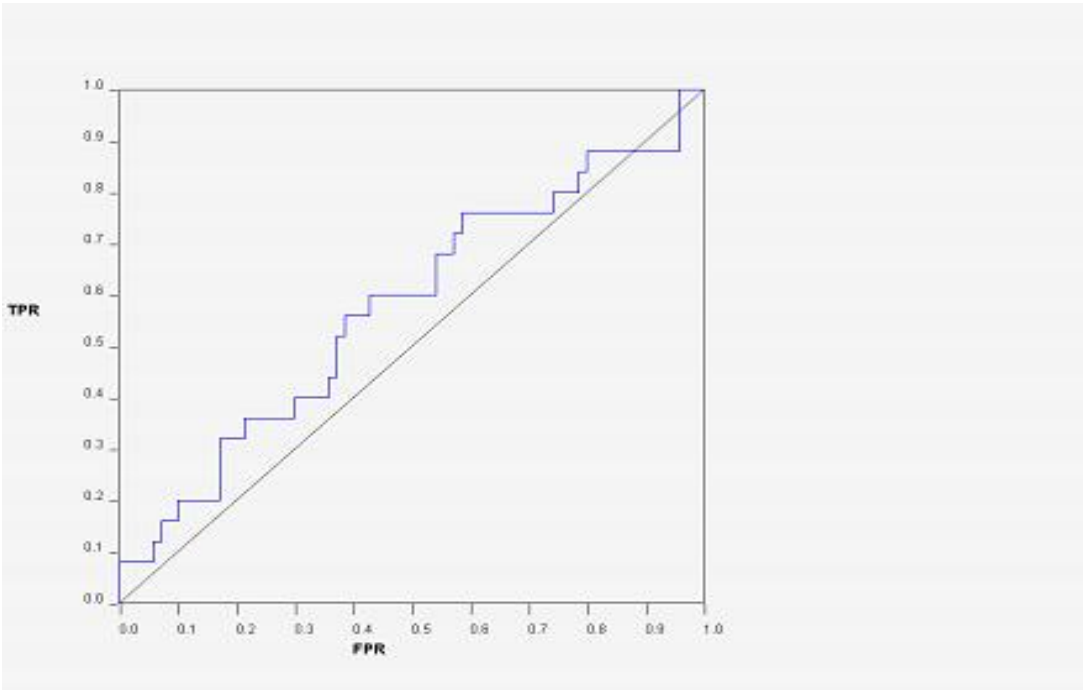

Fig.4.2 The ROC curve generated to test the accuracy of Breast cancer model

By definition the ROC receiver operating characteristic curve is a plot of the sensitivity against one minus the specificity (1-specificity) for different values of the threshold.

5. Conclusion and Future Works

The growth of research and applications of agent and data mining technologies in discovering knowledge from database has motivated an urgent need for techniques of integrating the two technologies in order to foster the weakness of one technology with the strength of the other. It is in that regards that, in this paper, we poured our efforts on integrating agent, data warehouse and data mining technology blocks, we used a three-layered architecture that can foster a synergy of complementation between agent technology and data mining technology to fully provide decision making to the end user.

As we know Breast cancer prediction is an important step in the complex decision-making process of deciding the type of treatment to be applied to patients after surgery. Some non-linear models, like neural networks, have been successfully applied to this task but they suffer from the problem of extracting the underlying rules, and knowing how the methods operate can help to a better understanding of the cancer relapse problem and give a high degree of performance and confidence. In this work we build a very effective and useful model which can be used to early cancer detection and hence reduce deaths to these needy women around the world.It is a prototype architecture which will be amended and ameliorated in the near future.From the above paragraph, we see that this paper has had some limitations. First and foremost the limited application scope of Breast Cancer concept in this paper was due to limited time and the accessibility of the clinical data especially those data for breast cancer. We believe that in the near future work we can try to accomplish this project by getting more data from different Breast Cancer data sources and develop the whole system

REFERENCES

Bellifemine, Fabio, AgostinoPoggi, and Giovanni Rimassa, 2001. "Developing multi-agent systems with JADE." *Intelligent Agents VII Agent Theories Architectures and Languages*.Springer Berlin Heidelberg, 2001.89-103.

Bent, Graham, *et al*. "A dynamic distributed federated database." *Proc. 2nd Ann. Conf. International Technology Alliance*. 2008.

Bollacker, Kurt, *et al*. 2008. "Freebase: a collaboratively created graph database for structuring human knowledge." Proceedings of the 2008 ACM SIGMOD international conference on Management of data. ACM.

Eckermann, Simon, and Andrew, R. Willan. 2007. "Expected value of information and decision making in HTA." *Health economics* 16.2: 195-209.

Forns, Xavier, *et al*. 2002. "Identification of chronic hepatitis C patients without hepatic fibrosis by a simple predictive model." *Hepatology,* 36.4 : 986-992.

Gerrity, Ross, G. 1981. "The role of the monocyte in atherogenesis: I. Transition of blood-borne monocytes into foam cells in fatty lesions." *The American Journal of Pathology,* 103.2: 181.

Goldring, Robert David. "Data replication in data warehousing scenarios." U.S. Patent No. 6,438,538. 20 Aug. 2002.

Lourenço, Anália, and Orlando Belo. "Abnormal data formats identification and resolution on data warehousing populating process."

Luck, Michael, and Peter McBurney. 2008. "Computing as interaction: agent and agreement technologies: 1-6.

Marakas, George M. 2003. *Decision support systems in the 21st century*. Vol. 134. Upper Saddle River, NJ: Prentice Hall.

Mazón, Jose-Norberto, and Juan Trujillo. 2008. "An MDA approach for the development of data warehouses." Decision Support Systems 45.1. 41-58.

Singh, Munindar, P. 2000. "A social semantics for agent communication languages." *Issuesin agent communication. Springer Berlin Heidelberg*, 31-45.

****