Volume 4, Issue 1, January 2014 ISSN: 2277 128X
International Journal of Advanced Research in Computer Science and Software Engineering
Research Paper
Available online at: www.ijarcsse.com

Evaluating data Reliability in a Clinical oriented data Warehouse using Multiagent Systems

T.P. Latchoumi[1] S. Anupriya[2], S.Jayashree[3], G. Saranya[4]
Department of Computer Science and Engineering
Christ College of Engineering and Technology, Pondicherry, India

Abstract: There are many available methods to combine information source reliability in an uncertainty representation, but there are only a few works focussing on the problem of evaluating this reliability. However, data reliability and confidence are essential components of a data warehousing system, as they influence subsequent retrieval and analysis. The reliability of data depends experimental protocol etc. Data warehousing can be defined as the subject oriented, time variant, integrated, non-volatile collection of data in support of management's decision making process. MAS may be viewed as a collection of distributed autonomous artefacts capable of accomplishing complex tasks through interaction, co-ordination, collective intelligence and emergence of patterns of behaviour. It is composed of multiple interacting intelligent agents. Data reliability means the trust we have in data values. We consider an integrated patient monitoring system, combining electronic patient records with high rate acquisition of patient psychological data. There remain many challenges in increasing the "Reliability" to a level at which they are clinically useful. In this paper, we propose a generic method to access data reliability by using MAS. The chosen illustrative example comes from real world clinical oriented data warehouse.

Key words-- Data Warehouse, Multi Agent Systems, Data Reliability, Clinical oriented data warehouse, generic method.

I. INTRODUCTION

Estimating data reliability is a major issue for many scientists, as these data are used in further inferences. During collection, data reliability is mostly ensured by measurement device calibration, by adapted experimental design and by statistical repetition. However, full traceability is no longer ensured when data are reused at a later time by other scientists. If a validated physical model exists and data values fall within the range of the model validated domain, then data reliability can be assessed by comparing data to the model predictions. However, such models are not always available and data reliability must then be estimated by other means. This estimation is especially important in areas where data are scarce and difficult to obtain (e.g., for economical or technical reasons), as it is the case, for example, in Life Sciences. The growth of the web and the emergence of dedicated data warehouses offer great opportunities to collect additional data, be it to build models or to make decisions. The reliability of these data depends on many different aspects and Meta information: data source, experimental protocol.

The term data warehouse (DW) is commonly used in industry and it denotes a kind of heterogeneous information system. We have to disclose firstly that a data warehouse is an environment, not a product. The need for building a data warehouse is that corporate data is often scattered in different databases and possibly in different formats. In order to view a complete picture of information, it is necessary to access these heterogeneous databases.

Developing generic tools to evaluate this reliability represents a true challenge for the proper use of distributed data. The rapid pace of development in "e-health" technologies within integrated healthcare systems (such as electronic patient records) has far outpaced their uptake in clinical practice. As evaluating data reliability is subject to some uncertainties, we propose to model information by the means of evidence theory, for its capacity to model uncertainty and for its richness in fusion operators.In this paper, a source (e.g., expert, sensor . . .) is considered as reliable if its information can be used safely (i.e., be trusted), while the information of an unreliable source has to be used with caution (note that information coming from an unreliable source may be true, but nothing guarantees it).

II. BACKGROUND DETAILS

Data warehouse can be explained as subject oriented, integrated, and non-renewable and data changing over time collection and it is used to support the processing in management decision. It consolidates data coming from different data sources. A real time data warehouse is used same purpose as data warehouse, in addition to these, data streams into real time data warehouse on time. The traditional data warehouse (DW) can only analyze historical data and data extraction cycle is so long that it greatly reduces the enterprise's adaptability to change this situation.

The concept of data warehousing dates back to the late 1980s when IBM researchers Barry Devlin and Paul Murphy developed the "business data warehouse". In essence, the data warehousing concept was intended to provide an architectural model for the flow of data from operational systems to decision support environments.

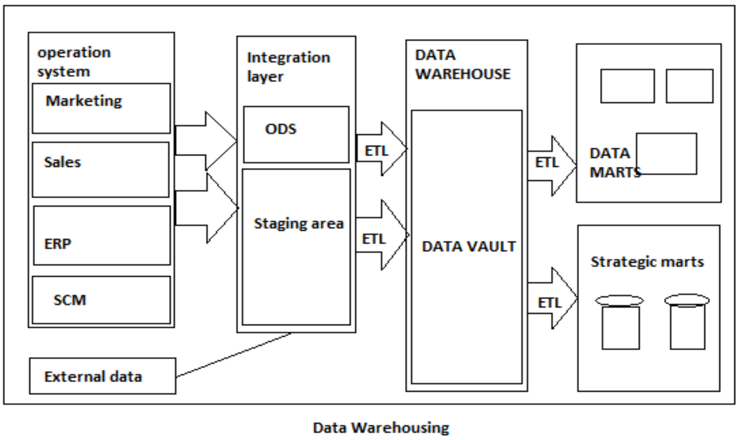

Fig 2.1 Data warehousing

The concept attempted to address the various problems associated with this flow, mainly the high costs associated with it. In the absence of a data warehousing architecture, an enormous amount of redundancy was required to support multiple decision support environments. In larger corporations it was typical for multiple decision support environments to operate independently. Though each environment served different users, they often required much of the same stored data. The process of gathering, cleaning and integrating data from various sources, usually from long-term existing operational systems (usually referred to as legacy systems), was typically in part replicated for each environment. Moreover, the operational systems were frequently reexamined as new decision support requirements emerged. Often new requirements necessitated gathering, cleaning and integrating new data from "data marts" that were tailored for ready access by users.

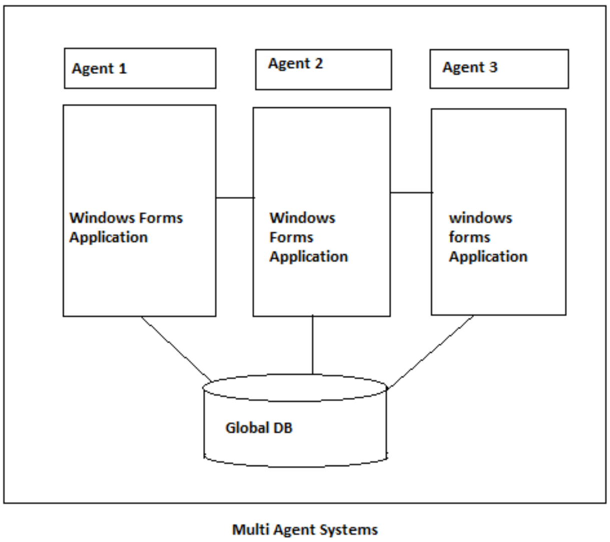

Fig 3.1 Multi Agent Systems

It is the collection of many numbers of intelligent agents. Agent is a calculation entity with autonomy, alternation, active and social characteristics in certain circumstances, it has some level of the learning ability and can complete particular task according to your own will.
It can also induct and adapt to the change of environment, and adjust its internal state in time. And the agents can communicate and cooperate with each other under the network environment, and complete a task together. The characteristics lay the foundation for the concept of agent system.
A **multi-agent system** (**M.A.S.**) is a computerized system composed of multiple interacting intelligent agents within an environment. Multi-agent systems can be used to solve problems that are difficult or impossible for an individual agent or a monolithic system to solve.
Intelligence may include some methodic, functional, procedural or algorithmic search, find and processing approach. Although there is considerable overlap, a multi-agent system is not always the same as an agent-based model (ABM). The goal of an ABM is to search for explanatory insight into the collective behavior of agents (which don't necessarily need to be "intelligent") obeying simple rules, typically in natural systems, rather than in solving specific practical or engineering problems. The terminology of ABM tends to be used more often in the sciences, and MAS in engineering and technology. Topics where multi-agent systems research may deliver an appropriate approach include online trading disaster response, and modelling social structures.

III. LITERATURE SURVEY

A. Data warehousing

An integrated patient monitoring system, combining electronic patient records with high-rate acquisition of patient physiological data[1]. There remain many challenges in increasing the robustness of “e-health” applications to a level at which they are clinically useful, particularly in the use of automated algorithms used to detect and cope with artefact in data contained within the electronic patient record, and in analyzing and communicating the resultant data for reporting to clinicians. There is a consequential “plague of pilots,” in which engineering prototype systems do not enter into clinical use. This paper describes an approach in which, for the first time, the Emergency Department (ED) of a major research hospital has adopted such systems for use during a large clinical trial.We describe the disadvantages of existing evaluation metrics when applied to such large trials, and propose a solution suitable for large-scale validation. We demonstrate that machine learning technologies embedded within healthcare information systems can provide clinical benefit, with the potential to improve patient outcomes in the busy environment of a major ED and other high-dependence areas of patient care. The rapid pace of development in “e-health” technologies within integrated healthcare systems (such as electronic patient records) has far outpaced their uptake in clinical practice. There is a perceived “plague of pilots” [1], in which prototype systems do not penetrate into clinical use, and there is a consequence lack of evidence required for adoption at scale. We address this problem by describing a large clinical trial in which the care of 10 000 patients in the Emergency Department of a major research hospital1 switches over to the use of integratedhealthcare systems that we have designed around.

B. Multiagent systems

Adopting new healthcare systems at scale in a clinical environment is a time-consuming and resource-intensive process, particularly in building the large bodies of evidence required to support adoption. This paper describes the trial needed to provide this evidence, in which algorithms for detecting physiological deterioration are embedded within an integrated healthcare system and are compared with the existing standard of hospital care (where the latter is introduced).

The infrastructure of the system is described. investigates the shortcomings of existing methods of using that infrastructure for patient care, and proposes techniques to overcome these shortcomings. We describe how methodologies for evaluating the success of electronic systems in clinical studies are inadequate:

In open and dynamic multi agent systems (MASs), agents often need to rely on resources or services provided by other agents to accomplish their goals. During this process, agents are exposed to the risk of being exploited by others. These risks, if not mitigated, can cause serious breakdowns in the operation of MASs and threaten their long-term wellbeing. To protect agents from the uncertainty in the behaviour of their interaction partners, the age-old mechanism of trust between human beings is re-context-ed into MAS s. The basic idea is to let agents self-police the MAS by rating each other on the basis of their observed behaviour and basing future interaction decisions on such information.

Over the past decade, a large number of trust management models were proposed. However, there is a lack of research effort in several key areas, which are critical to the success of trust management in MASs where human beings and agents coexist. The purpose of this paper is to give an overview of existing research in trust management in MAS s. We analyze existing trust models from a game theoretic perspective to highlight the special implications of including human beings in MAS, and propose a possible research agenda to advance the state of the art in this field.

Many systems that are commonplace in our lives, such as e-commerce platforms, crowdsourcing systems, online virtual worlds, and P2P _le sharing systems, can be modelled as open dynamic multi-agent systems (MASs). The agents in these systems can represent software entities or human beings. They are considered *open* because agents can come from any background with eterogeneous abilities, organizational af_liations, credentials, etc. They are considered *dynamic* ecause the decision-making processes of the agents are independent from each other and agents an join or leave the system at will.

A multi-agent system is a computerized system composed of multiple interacting intelligent agents within an environment. Multi-agent systems can be used to solve problems that are difficult or impossible for an individual agent or a monolithic system to solve.

Intelligence may include some methodic, functional, procedural or algorithmic search, find and processing approach. Although there is considerable overlap, a multi-agent system is not always the same as an agent-based model (ABM). The goal of an ABM is to search for explanatory insight into the collective behavior of agents (which don't necessarily need to be "intelligent") obeying simple rules, typically in natural systems, rather than in solving specific practical or engineering problems. The terminology of ABM tends to be used more often in the sciences, and MAS in engineering and technology. Topics where multi-agent systems research may deliver an appropriate approach include online trading disaster response, and modelling social structures.

IV. PROPOSED SYSTEM

The concept of our project is about to manage the evaluation of data reliability by using the Multi Agent Systems. The multi-agent system and the Data ware house can be combined at the 4-tier data warehouse.

This is the concept of combining both the data warehouse and the multiagent systems. It is a method to evaluate data reliability from Meta information using MAS. Several criteria are used, each one providing a piece of information about data reliability. These pieces are then aggregated into a global assessment that is sent back, after proper post-treatment, to the end user. In our opinion, such a method should

- Deal with conflicting information, as different criteria may provide conflicting information about the reliability. For example, data may come from a reputed journal, but have been collected with rather unreliable instruments;
- be traceable, as it is important to be able to detect conflict and to provide insights about its origins, or in the absence of such conflicts, to know why such data have been declared poorly (or highly) reliable;
- Be readable, both in its different input parameters and results, as the method and the system it is implemented in will be used mainly by non-computer scientists.

  To overcome the disadvantages with the existing systems, we are proposing the system. They are about Applying and managing generic method by using multiple experts and getting feedback from the user and inserting in DW.

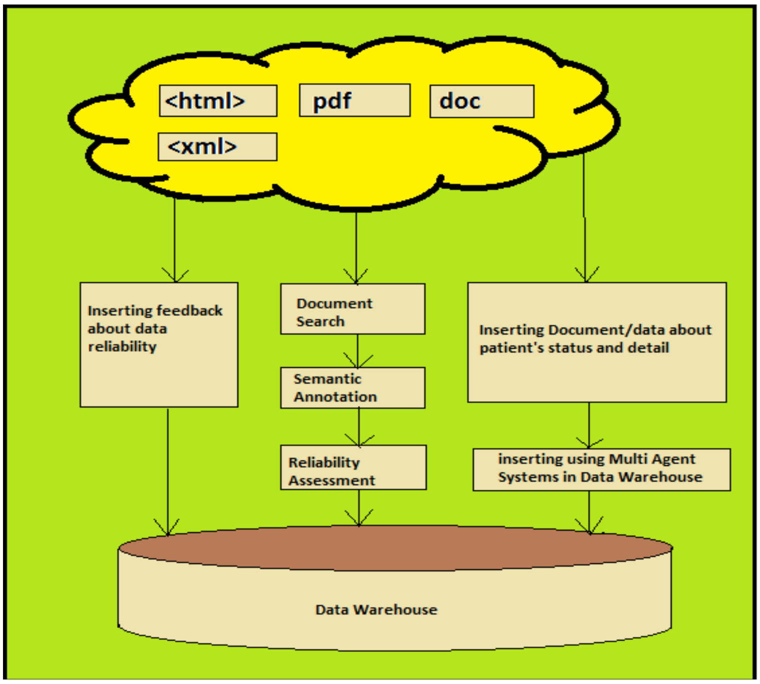

Fig 5.1 Evaluating data reliability in a clinical oriented data warehouse using multi agent systems

So far the method used for the data reliability is the generic method where the Maximal Coherent Systems (MCS) and merging have been used. Merging is the method of Combine or cause to combine to form a single entity, esp. a commercial organization. MCS consists in applying a conjunctive operator within each nonconflicting (maximal) subset of sources, and then using a disjunctive operator between the partial results. With such a method, as much precision as possible is gained while not neglecting any source, an attractive feature in information fusion.

Algorithm 1. Maximal Coherent Subset detection
INPUT: k intervals $I_{a,b}$
OUTPUT: List of maximal coherent subsets K_j
List = ∅ ; j=1; k=∅;
Order by increasing value
$\{a_i,a=1,.......,k\}U\{b_i,i=1,.......,k\}$(in case of ties, a_i is before b_i);
Rename them $\{c_i$, i=1,.........2k$\}$ with type(i)=a if
$C_i=a_m$ and type(i)=b if $c_i=b_m$;
for i=1,........ , 2k-1 do
if type(i) = a then
Add source m to K s.t. $c_i=a_m$;
If type(i+1)=b then
K_j=K;
Add K_j to list;
J=j+1;
Else
Remove source m from K s.t. $c_i=b_m$;

V. CONCLUSION

For the lack of automated reliability Assessment, it comes up with evaluating data reliability based on Multi-Agent Systems. if user searches a document in DW, then the document will be assessed and the feedback will be saved in the Data warehouse. For this purpose, we used Maximal Coherence subset (MCS) using MAS which evaluates Data reliability automatically and the feedback will be inserted in DW using MAS.

REFERENCE

[1] "Evaluating Data Reliability: An Evidential Answer with Application to a Web-Enabled Data Warehouse" Sebastien Destercke, Patrice Buche, and Brigitte Charnomordic Ieee Transactions On Knowledge And Data Engineering, Vol. 25, No. 1, January 2013

[2] "A Survey of Multi-Agent Trust Management Systems HAN YU1, ZHIQI SHEN1, CYRIL LEUNG2, CHUNYAN MIAO1, AND VICTOR R. LESSER3 Received March 27, 2013, accepted April 10, 2013, published May 10, 2013.

[3] "A Large-Scale Clinical Validation of an Integrated Monitoring System in the Emergency Department" David A. Clifton, David Wong, Lei Clifton, Sarah Wilson, Rob Way, Richard Pullinger, and Lionel Tarassenko, IEEE JOURNAL OF BIOMEDICAL AND HEALTH INFORMATICS, VOL. 17, NO. 4, JULY 2013

[4] "Agent Based Architecture in Distributed Data Warehousing" Bindia, Jaspreet Kaur Sahiwal, International Journal of Scientific and Research Publications, Volume 2, Issue 5, May 2012 1 ISSN 2250-3153

[5] "Multi-Agent System for Decision Support in Enterprises" Dejan Lavbi, Rok Rupnik, JIOS, VOL. 33, NO. 2 (2009)

