IJISET - International Journal of Innovative Science, Engineering & Technology, Vol. 1 Issue 3, May 2014.
www.ijiset.com
ISSN 2348 - 7968

Therapeutic Decisions Making System Using Data Mining Techniques: A Review

Shareen S. Anthony[1], **Prof. Amit A. Sahu**[2]

[1] *Dept of CSE GH Raisoni, Amravati,*

[2] *Dept of CSE GH Raisoni, Amravati*

Abstract

Medical prognosis has played an increasing role in health, namely in the critical care medicine. These factors have induced the medical community to take a more active interest in developing models for mortality prediction based on Artificial Intelligence (AI) techniques [1]. that make possible the doctors pro-active action. In this context, the existence of large Databases (DB) containing Intensive Care Units (ICU) clinical information, motivate and enable the application of Data Mining (DM) techniques, in a Knowledge Discovery Database process (KDD). to induce prediction models of organ failure in a much more efficient way than other approaches (e.g.. Logistic Regr es sion) [2].

Clinical guidelines carry medical evidence to the point of practice. As evidence is not always available, many guidelines do not provide recommendations for all clinical situations encountered in practice. We propose a decisions making system for exploring physicians' therapeutic decisions with data mining techniques to fill knowledge gaps. The various data mining techniques are explored in this paper for the proposed system

Keywords: *Decision support system, knowledge discovery, intensive care, data mining, Artificial Neural Networks*

1. Introduction

Modern organizations use several types of decision support systems to facilitate decision support. For the purposes of analysis and decision support in the business area in many cases OLAP based decision support systems are used [1]. Performing analysis through OLAP follows a deductive approach of analyzing data [8]. The disadvantage of such an approach is that it depends on coincidence or even luck of choosing the right dimensions at drilling-down to acquire the most valuable information, trends and patterns. We could say that OLAP systems provide analytical tools enabling user-led analysis of the data, where the user has to start the right query in order to get the appropriate answer [9]. Such an approach enables mostly the answers to the questions like: “What is overall revenue for the first quarter grouped by customers?” What about the answers to the questions like: “What are characteristics of our best customers?” Those answers cannot be provided by OLAP systems, but by the using data mining. Performing analysis through data mining follows an inductive approach of analyzing data [8].

Data mining is a process of analyzing data in order to discover implicit, but potentially useful information and uncover previously unknown patterns and relationships hidden in data. The use of data mining to facilitate decision support can lead to an improved performance of decision making and can enable the tackling of new types of problems that have not been addressed before. The integration of data mining and decision support can significantly improve current approaches and create new approaches to problem

The idea of conceiving Decision Making Support Systems (DMSS) making use of the Knowledge Discovery from Databases (KDD) and Agent-Based Systems paradigms, as a way to solve complex and dynamic problems, is not new [11]. However, a great part of these concepts (and architectures) need to be corroborated by real-world applications, in order to obtain a valuable feedback of its effective use. The intensive care medicine, due to its high volume and complexity of data, is a rich field to test this thesis.

DMSSs are computer based systems that support one or several phases of the individual, team organizational or inter-organizational decision making process [12]. While DMSS constituted one of the most popular areas of research in information systems in the past, a closer look seems to indicate that the interest in DMSS is declining [13] or is not increasing at the rate that is expected [14]. This is due to the requirements for decision support tools posed by new dynamic and complex environments (like the one in an ICU), demanding for situated and active DMSSs. A situated DMSS is similar to an open system in Artificial Intelligence (AI), a ‘‘view’’ that has been gaining popularity since the late 1980s. In a situated system the focus is on the interaction with its environment, the behavior is the combined result of its purpose and its interaction with the environment.

Other concerns come from the extended Keen's agenda for DMSS research to the year 2007, which postulates that researchers and developers should be more prescriptive about effective decision making by using intelligent systems and methods. The incorporation of Artificial Intelligence techniques, opened room for a new and more successful kind of DMSS, called intelligent decision making support systems (i-DMSS).

This paper introduces the Decision Making System based on the KDD and the Agent-Based paradigms, to support intensive care medical activities. In particular, the system assists the physicians' decision making by: (i) detecting action demanding conditions by continuously scanning automatically acquired data and applying the relevant model to predict next day failure of six systems (liver, respiratory, cardiovascular, coagulation, central nervous and renal); (ii) maintaining an up-to-date in-hospital death probability value used in end-of-life decision making and (iii) evaluating scenarios for the evolution of the condition of the patient, allowing physicians to compare the consequences of different medical procedures. The INTCare system is being tested in a real environment, the Intensive Care Unit (ICU) of Hospital Geral de Santo António (HGSA), Oporto, north of Portugal.

2. Related Work

In the beginning of the 80s, several expert systems were developed for medicine, such as MYCIN , CASNET and CADUCEUS, just to name a few. A decade later, in the 90s, the common knowledge was that these expert systems, based on information retrieved from experts only, were not sufficient for the complex real-world problems. A shift was induced, where the main emphasis was the gathering of knowledge directly from the data, using intelligent data analysis. At the end of the 90s, this approach gained an increasing interest, especially in the medicine area, due to the high volume and complexity of the clinical data. In terms of intensive care medicine, the application of DM techniques (e.g. Artificial Neural Networks) is new, although possessing a huge potential. However, it should be stressed that the number of studies were these techniques are applied in a real environment is very limited.

Everyday, doctors have to make decisions that seriously affect their patients. First, they make a diagnosis and then conceive therapeutic plans (e.g. prescribe treatments) in order to improve the patients' condition. However, ICU doctors have to make decisions that are even more challenging, such as the ones related to life-support treatments. Resource availability limitations force them to make sure that intensive care is applied only to those who are likely to benefit from it. Critical decisions include interrupting life-support treatments and writing do-not-resuscitate orders when intensive care is considered futile. In addition, the condition of ICU patients is such that doctors are not able to completely assess it and will benefit from having extra quality information available.

Within the Medicine arena, huge databases, with large, complex and multi-source information (e.g. text, images or numerical data), are commonplace. However, human experts are limited and may overlook important details. Furthermore, the classical data analysis (e.g. logistic regression) breaks down when such vast amounts of data are present. Hence, an alternative is to use automated discovery tools to analyze the raw data and extract high level information for the decision-maker [10].

An important step of the KDD process is related to the choice of the DM function, which will effect the selection of the DM algorithm [11]. In ICUs, the most useful DM goals are:

- *classification* - labels a given set of attributes into one of several predefined classes (e.g. diagnosing a disease according to the patient's symptoms);

- *regression* - maps a data item into a real-value variable (e.g. estimation of the patient's heart rate); and

- *clustering* **-** searches for natural groupings of objects based on similarity measures (e.g. segmenting patients into clusters according to similar profiles).

The above goals may involve the application of Machine Learning algorithms such as Decision Trees (DT), Learning Classifier Systems (LCS) [16] and Artificial Neural Networks (ANN) [10]. More recently, there has been an emergent DM research area that involves the use of ensembles for supervised learning, where a set of classification/regression models are combined in some way to produce an answer.

Using the large databases of electronic patient records now available, it is possible to use data mining and knowledge discovery techniques to identify common therapeutic decisions made by physicians for a given clinical condition. There have been some limited attempts at using these techniques for generating practice guidelines from data. Mani et al. [9] presented a two-stage machine learning model as a data mining method to develop clinical practice guidelines, and showed its value in staging dementia. They modeled the methodology used by clinicians by deriving intermediate concepts in the first phase, and in the second phase they used the intermediate concepts for staging dementia. However, the dementia scoring scale that they learnt led to a less complex guideline than those usually implemented in other domains. It is also not clear whether their method can be generalized to different domains. Morik et al. [10] used a combination of prior knowledge from experts and learning from data to generate protocols automatically for decision support in intensive care. They used support vector machines (SVM) to learn the appropriate dose adjustment for each continuously administered drug, based on the response of the patient's vital signs to that change. They then verified the learnt dose adjustment against a medical knowledge base.

However, the resulting protocol was again much less complex than most guidelines. In another attempt, Mani et al. [11] used C4.5 and Ripper algorithms with a database of 369 patients and showed that data mining methods could be used for generating simple guidelines and checking compliance to guidelines. Nevertheless, the guidelines resulting from these efforts for creating entire guidelines from data only handle simple problems and lack detail and readability.

A synopsis of the current decision making process in the ICU of the HGSA is presented in Figure 1. This process can be seen as unfolding in four phases (Simon 1977) (Turban et al., 2004):

Intelligence - The medical and nursing staff collect physiological data regarding the patient’s condition. Then, based on this information, the physicians evaluate the illness state to determine the pertinence of immediate action;

Design - Doctors conceive the possible therapeutic scenarios. Two main categories of scenarios can be considered: immediate action assumption (where there is a pressing need for intervention) and delayed group decision (guidelines for discharge or long term treatments). In the former case, if needed, the physician can consult internal or external colleagues. Delayed group decision choices include: patient’s discharge, suspension of life-supporting treatment, writing of do-not resuscitate orders and applying a specific treatment;

Choice – After the formation of the alternatives and entwined with the previous stage, a final decision is taken (either individually or group based) and

Implementation – The corresponding therapeutic procedure is applied

I. USE OF EVIDENCE-BASED RULES AT THE POINT OF CARE

Due to the massive economical impact of the health system, great changes in medical treatments are notable. Apart of humanitarian and healing nature of medicine, this industry is becoming more and more business like. Various serious medical studies show, that the even patients with harmless health complaints, are often exposed to unnecessary but expensive therapies. In case of serious diseases, clinicians often recommend new and costly medical treatments instead of traditional therapies. Since they are assured to get a quick recovery, patients are usually willing to try newest therapy approaches.

It is obvious, that this does not mean that new methods are not efficient. It solely means, that tests and clinical studies are needed, in order to prove their effectiveness. Evidence-based medicine offers a collection of proven best practise guidelines to recommend drugs and medical treatments. With the help of a data warehouse, clinicians can easily navigate through the knowledge database and find appropriate therapy for the particular disease.

In this case study, we show how a data warehouse is used to determine the best suitable treatment for a particular patient, using evidence-based guidelines. The described healthcare institution is a clinic, in which the patients are treated ambulatory as well as hospitalised.

Clinical business management is striving to deliver best medical treatment to the patients by applying the most effective therapies. In order to cope with future technology challenges and newest medical achievements, as well as to reduce administrative and treatment costs, this clinic is humanitarian and healing nature of medicine, this industry is becoming more and more business Various serious medical studies show, that the even patients with harmless health complaints, are often exposed to unnecessary but expensive therapies. In case of serious diseases, clinicians often recommend new and costly medical treatments instead of traditional therapies. Since they are assured to get a quick recovery, patients are usually willing to try newest therapy approaches.

It is obvious, that this does not mean that new methods are operating a data warehouse based on evidence-based guidelines, to support decision making at the point of care.

Fig. 2 represents an extract of a clinical data model, holding personal data about the *patient*. A patient is uniquely identified by social security number and possesses additional attributes: name, date of birth, gender etc. Patient is characterised by the relations to another entities, which may be seen as the dimensions of the particular patient. Patient may have one or more drug *prescriptions*. The set of possible drugs is given in a *drug*-entity, giving information about drug characteristics, like description, size or pharmaceutical form. Each patient may have one or more *diagnoses*. These are specified by diagnoses codes and classifications. Since one patient can have many diagnoses, one diagnose is only valid in a certain period of time. For each patient and each diagnose made, there is a responsible clinician stored in the *clinician* entity. For each diagnose, there is one or more *therapies* assigned. Apart of therapy descriptive attributes, *charging* information is available as well. Information about the medical institutions (*clinic*), where the patients are treated and clinical facilities (*equipment*) are also available in this model.

3. Conclusions

In this paper the overall literature survey related to different various data mining techniques for different applications are mentioned. It is observed various data mining techniques are very helpful for the efficient working. Different applications used various methods and algorithms for implementing various data mining techniques and they have been proved as efficient in their domain of work.

References

[1]. S. M. Metev and V. P. Veiko, Laser Assisted Microtechnology, 2nd ed., R. M. Osgood, Jr., Ed. Berlin, Germany: Springer-Verlag, 1998.
[2]. J. Breckling, Ed., The Analysis of Directional Time Series: Applications to Wind Speed and Direction, ser. Lecture Notes in Statistics. Berlin, Germany: Springer, 1989, vol. 61.
[3]. Zhang, C. Zhu, J. K. O. Sin, and P. K. T. Mok, “A novel ultrathin elevated channel low-temperature poly-Si TFT,” IEEE Electron Device Lett., vol. 20, pp. 569–571, Nov. 1999.
[4]. M. Wegmuller, J. P. von der Weid, P. Oberson, and N. Gisin, “High resolution fiber distributed measurements with coherent OFDR,” in Proc. ECOC’00, 2000, paper 11.3.4, p. 109.
[5]. R. E. Sorace, V. S. Reinhardt, and S. A. Vaughn, “High-speed digital-to-RF converter,” U.S. Patent 5 668 842, Sept. 16, 1997.(2002) The IEEE website. [Online]. Available: http://www.ieee.org/
[6]. M. Shell. (2002) IEEEtran homepage on CTAN. [Online]. FLEXChip Signal Processor (MC68175/D), Motorola, 1996.
[7]. “PDCA12-70 data sheet,” Opto Speed SA, Mezzovico, Switzerland.
[8]. Karnik, “Performance of TCP congestion control with rate feedback: TCP/ABR and rate adaptive TCP/IP,” M. Eng. thesis, Indian Institute of Science, Bangalore, India, Jan. 1999.

[9]. J. Padhye, V. Firoiu, and D. Towsley, “A stochastic model of TCP Reno congestion avoidance and control,” Univ. of Massachusetts, Amherst, MA, CMPSCI Tech. Rep. 99-02, 1999.Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specification, IEEE Std. 802.11, 1997.

[10]. Hand D., Mannila H., Smyth P.(2001) Principles of Data Mining. MIT Press, Cambridge, MA

[11]. Fayyad U., Piatesky-Shapiro G., Smyth P. (1996) From Data Mining to Knowledge Discovery: An Overview. In Fayyad et al. (eds) Advances in Knowledge Discovery and Data Mining. AAAI Press / The MIT Press, Cambridge MA, pp 471-493

[12]. Forgionne G.A., Mora M., Cervantes F., Kohli R (2000) Development of integrated decision making support systems: a pratical approach. In Proceedings of the AIS Conference, Long Beach CA, pp 2132-2134.

[13]. Claver E., Gonzales R., Llopis J. (2000) An analysis of research in information systems (1981-1997), Information and Management 37 181– 195.

[14]. Vahidov R., Kersten G. (2004) Decision station: situating decision support systems. Decision Support Systems (38) 283-303.

[15]. Lanzi P., Solzman W., Wilson S. (2000) Learning Classifier Systems from Foundations to Applications. Lecture Notes in Artificial Intelligence 1813, Springer, Berlin Heidelberg New York.