ISSN 2348-1196 (print)
International Journal of Computer Science and Information Technology Research ISSN 2348-120X (online)
Vol. 2, Issue 2, pp: (243-247), Month: April-June 2014, Available at: www.researchpublish.com

Application of Data Mining Techniques in Health Care Industry

S.Dharani Shree[1], A.Kanimozhi[2] and Dr.G.N.K.Suresh Babu[3]

[1,2]Research Scholars, Dept of Computer Applications, GKM College of Engineering and Technology, Chennai-63
[3]Professor and Head, Dept of Computer Applications, GKM College of Engineering and Technology, Chennai-63

Abstract: **Data mining is a relatively new field of research whose major objective is to acquire knowledge from large amounts of data. In medical and health care areas, due to regulations and due to the availability of computers, a large amount of data is becoming available. On the one hand, practitioners are expected to use all this data in their work but, at the same time, such a large amount of data cannot be processed by humans in a short time to make diagnosis, prognosis and treatment schedules. The main objective of this paper is to analyze how the data mining can be used in health sector and discusses the importance of resident assessment instrument.**

Keywords: **Data Mining, Health, Forecast, Prediction.**

1. INTRODUCTION

Data mining has grown so vast that they can be used in many applications examples include predicting costs of corporate expense claims, in risk management, in financial analysis, in insurance, in process control in manufacturing, in healthcare, and in other fields. The Healthcare industry is among the most information intensive industries. Medical information, knowledge and data keep growing on a daily basis. It has been estimated that an acute care hospital may generate five terabytes of data a year. The ability to use these data to extract useful information for quality healthcare is crucial. Medical informatics plays a very important role in the use of clinical data. In such discoveries pattern recognition is important for the diagnosis of new diseases and the study of different patterns found when classification of data takes place The number of people feeling sick and getting admitted into clinics and hospitals are increasing proportionally. The growing number of patients indirectly increases amount of data that are required to be stored. If a small number of patients, visit a doctor during a given redundant, the doctor will be able to work efficiently and provide proper care of the patient. Now consider the case when there is a large number of patients' coming to meet this doctor in the same period. We will find the quality of care of the doctor will decrease. If the doctor has another colleague at his side he can at times ask him for a second opinion before making decisions about the patient.

2. DATA MINING

Data mining is one among the most important steps in the knowledge discovery process. It can be considered the heart of the KDD process. This is the area, which deals with the application of intelligent algorithms to get useful patterns from the data. Some of the different methods of learning used in data mining and as follows :

Classification learning:- The learning algorithms take a set of classified examples (training set) and use it for training the algorithms. With the trained algorithms, classification of the test data takes place based on the patterns and rules extracted from the training set. Classification can also be termed as predicting a distinct class.

· *Numeric predication:*- This is a variant of classification learning with the exception that instead of predicting the discrete class the outcome is a numeric value.

· *Association learning:*- The association and patterns between the various attributes are extracted are from these rules are created. The rules and patterns are used predicting the categories or classification of the test data.
· *Clustering:* - The grouping of similar instances in to clusters takes place. The challenges or drawbacks considering this type of machine learning is that we have to first identify clusters and assign new instances to these clusters.

3. KNOWLEDGE DISCOVERY IN DATABASES [KDD] AND DATA MINING

Traditional methods (Methods used before computers where introduced into healthcare) use manual analysis to find patterns or extract knowledge from the database. For example in the case of health care, the health organizations analyze the trends in diseases and the occurrence rates. This helps health organizations take precautions in future in decision making and planning of health care management. The traditional method is used to analyze data manually for patterns for the extraction of knowledge. Take any field like banking, mechanic, healthcare, and marketing; there will always be a data analyst to work with the data and analyzing the final results. The analyst acts like an interface between the data and knowledge. We can, using machine intelligence assist the analyst to produce similar results or knowledge from the data. When we encounter patterns within a database we state the findings (patterns or rules) as data mining, information retrieval or knowledge extraction and so on. The term data mining is used mostly by statisticians, data analysts and the management information systems (MIS). The difference between data mining and knowledge discovery is that the latter is the application of different intelligent algorithms to extract patterns from the data whereas knowledge discovery is the overall process that is involved in discovering knowledge from data. There are other steps such as data preprocessing, data selection, data cleaning, and data visualization, which are also a part of the KDD process.

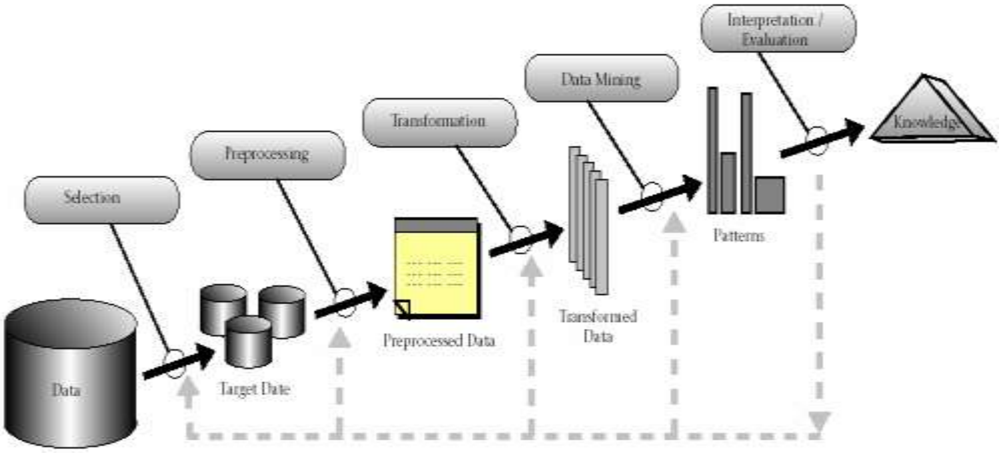

Figure 1. Overview of the steps involved in the KDD process

4. MACHINE LEARNING

Machine Learning is the study of computer algorithms that improve automatically through experience. Applications of machine learning range from data mining programs that discover general rules in large data sets, to information filtering systems that automatically learn users' interests. Machine learning can be used to develop systems resulting in increased efficiency and effectiveness of the system. Machine learning is also called concept learning. That is, computers can learn concepts and patterns within the data. Machine learning is considered successful when it can correctly find all the instances that consist of the right patterns and concepts. Although at times a machine cannot categorize correctly all the instances due to high variations in attributes present in the data.

5. DATA MINING PERSPECTIVE

Data mining lies at the interface of statistics, database technology, pattern recognition, machine learning, data visualization, and expert systems. A database is a collection of data that is organized so that its contents can easily be accessed, managed, and updated. Databases contain aggregations of data records or files, and a database manager provides users the capabilities of controlling read and write access, specifying report generation, and analyzing use. Databases and database managers are prevalent in large mainframe systems, but are also present in smaller systems and on personal computers. Databases usually include a query facility, and the database community has a tendency to view data mining methods as more complicated types of database queries. For example, standard query tools can answer questions such as, "How many surgeries resulted in hospital stays longer than 10 days?" Data mining is valuable for more complicated queries such as, "What are the important preoperative predictors of excessive length of stay?" Data mining techniques can be implemented retrospectively on massive data in an automated matter, whereas traditional statistical methods used in epidemiology require custom work by experts. Traditional methods generally require a certain number of predefined variables, whereas data mining can include new variables and accommodate a greater number of variables.

6. POLICY ISSUES IN HEALTHCARE REFORMS

The goal of any reforms should be to achieve a more efficient health delivery system and a more robust infrastructure. The changes should be more fundamental, strategic, sustainable and forward looking. Any partial or unplanned changes may not lead to long-term solutions. Reform efforts should be focused mainly on creating, changing or modifying certain 'central knobs'. These 'central knobs' include financing, payment, organization, regulation and consumer behavior. Any such changes in the very core of the health system have to be well-planned, politically acceptable and socially and economically feasible. After an overview of all such reform initiatives around the globe, there are two major strategies that may suit the Indian scenario. These two strategies include, Private healthcare- Public healthcare co-operation, and decentralization using a 'principal- agent approach'.

7. FORECASTING IN HEALTH SECTOR

In general predictions about future health - of individuals and populations - can be notoriously uncertain. However all projections of health care in India must in the end rest on the overall changes in its political economy - on progress made in poverty mitigation in reduction of inequalities, in generation of employment /income streams in public information and development communication and in personal life style changes. Of course it will also depend on progress in reducing mortality and the likely disease load, efficient and fair delivery and financing systems in private and public sectors and attention to vulnerable sections- family planning and nutritional services and women's empowerment and the confirmed interest of me ensure just health care to the Largest extent possible. To list them is to recall that Indian planning had at its best attempted to capture this synergistic approach within a democratic structure. It is another matter that it is now remembered only for its mixed success.

8. HEALTH INFORMATICS

Healthcare is a very research intensive field and the largest consumer of public funds. With the emergence of computers and new algorithms, health care has seen an increase of computer tools and could no longer ignore these emerging tools. This resulted in uniting of healthcare and computing to form health informatics. This is expected to create more efficiency and effectiveness in the health care system, while at the same time, improve the quality of health care and lower cost. Health informatics is an emerging field. It is especially important as it deals with collection, organization, storage of health related data. With the growing number of patient and health care requirements, having an automated system will be better in organizing, retrieving and classifying of medical data. Physicians can input the patient data through electronic health forms and can run a decision support system on the data input to have an opinion about the patient's health and the care required. An example in the advances in health informatics can be the diagnosis of a patient is health by a doctor practicing in another part of the world. Thus healthcare organizations can share information regarding a patient which will cut costs for communication and at the same time be more efficient in providing care to the patient.

There are other issues like data security and privacy, which is equally important when considering health related data. Thus Health informatics "deals with biomedical information, data, and knowledge--their storage, retrieval, and optimal use for problem solving and decision making". This is a highly interdisciplinary subject where fields in medicine, engineering, statistics, computer science and many more come together to form a single field. With the help of smart algorithms and machine intelligence we can provide the quality of healthcare by having, problem solving and decision-making systems. Information systems can help in supporting clinical care in addition to helping administrative tasks. Thus the physicians will have more time to spend with the patients rather than filling up manual forms. First the paper forms that are filled by the physicians are converted into electronic forms. Programs can be built around these forms to help in input validations. Some of the validation steps can be in the form of cautions provided when fields are inputted with invalid values; another type of validation can be to make sure attributes of high priority are not left empty by the user. The informatics part of health care can take care of the structuring; searching, organizing and decision making with the emergence in health informatics came many important research ideas and fields of study. One among them is the Resident Assessment Instrument (RAI).

9. INTER-RESIDENT ASSESSMENT INSTRUMENT (INTER-RAI)

The Inter-RAI is a comprehensive standardized instrument for evaluating the needs, strengths and preferences of psychiatric patients in institutional settings. Inter-RAI aims at patients with acute care and long term needs. Inter-RAI consists of a collection of patient assessment instruments, which are used to gather information, such as patient's strengths and needs, and are also used to develop individual care plans for different patients. These assessments can be updated according to the patients' health which should improve the care that is provided to the patient. The Inter-RAI is basically a structured idea of how to produce a well-defined approach to identify the problem with respect to treating a patient who requires long-term care. There are more than eight different types of Inter-RAI assessment instruments. These set of assessments are customized according to the patients requirements, thus not all the patients will have the same assessment form, which means a patient with acute care needs with regard to old age facilities will have different assessment forms as compared to one who requires acute care in mental health. The forms have all the information or questions that are related for a particular assessment. In Inter-RAI there are a number of forms that are required for diagnosis corresponding to certain health care issues such as with some acute care or diagnosis of patients with mental health. The Inter-RAI collection of instruments is also a kind of minimum data set instruments. This can be considered as the minimum number of questions that are required to make a proper diagnosis of a patient with respect to a certain acute problem. All well-defined problem identification process follows similar steps as mentioned below where RAP is the resident assessment protocol.

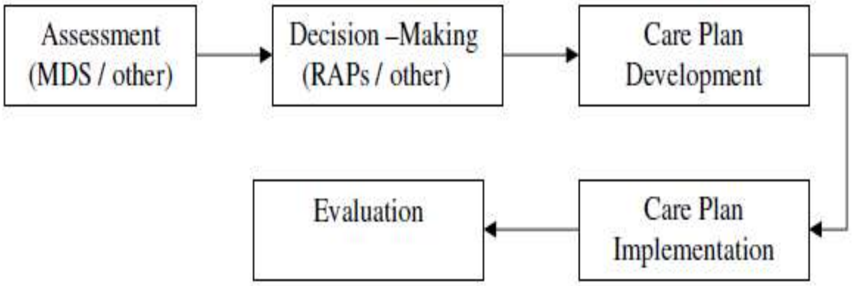

Figure 2. Assessment format for the Inter-RAI system

The end result of implementing these forms is, improved resident care and better quality of life due to the thorough diagnosis of the patient with the help of the Inter-RAI forms. Increasing attention provided to each resident should result in the patient responding better to treatment. Clinical staff will have a clearer picture having all the documentations of the patient in hand

and thus producing effective communication between staff members and individual residents. The documentation of the Inter-RAI is clear and there will be only one answer to each question. With proper documentation there should be fewer clerical errors and, at the same time, educating new staff members will be easier.

10. CONCLUSION AND FUTURE WORK

There have been significant advances in the healthcare system in India over last few decades. Despite these recent strides the health system remains ineffective in providing basic minimum care as promised in the Indian Constitution. The fiscal constraints on the government makes it obligatory for the private healthcare providers to take over part of the responsibility. New ways for establishing, strengthening and sustaining the private- public co-operation are essential for rejuvenating the system. At the same time decentralization exercises can make the health system more efficient and improve the quality of healthcare delivery. All these changes will need to be based on a strong political will and should be accompanied by economic and social reforms. Machine intelligence algorithms are improving as the number of data mining tools and algorithms increase. Healthcare data is a good test bed for data mining. A great deal of data in health care is still being gathered and organized using pen and paper.

Mobile computing plays a very important role in today’s information retrieval system. Some of the new handheld devices, cellular phones, PDAs, the Blackberry and others can be connected to the Internet and information can be received and sent from servers. There are a number of different data mining algorithms that produce rules that can be stored in mobile devices and used for data classification. A possibility for future work could be to implement a local interface for the device where user can input data directly into their mobile devices, and based on the rule set, can deliver the answer back, i.e. classification is done using rules stored in the database of the PDA. This can be a handy tool for medical practitioners. Automated surveillance systems offer obvious advantages over manual ones. When analytical technologies are embedded in automated hospital infection surveillance systems, it is not clear whether data mining outperforms traditional statistical methods.

REFERENCES

[1] Brosette SE, Spragre AP, Jones WT, Moser SA. A data mining system for infection control surveillance. Methods Inf Med 2000;39:303-310.

[2] Hirdes JP, Marhaba M, Smith, TF et al. 2001 Development of the Resident Assessment Instrument - Mental Health (RAI-MH), Hospital Quarterly, 4(2), 44-51

[3] Health Privacy Project (2003). Medical Privacy Stories, http://www.healthprivacy.org

[4] Kim, H. and Loh, W.-Y. 2001, Classification trees with unbiased multiway splits, Journal of the American Statistical Association, vol. 96, pp. 589-604.

[5] Matkovsky IP, Nauta KR. Overview of data mining techniques. Presented at the Federal Database Colloquium and Exposition; September 9-11, 1998; San Diego, CA.

[6] Ridinger M. American Healthways uses SAS to improve patient care. DM Review 2002;12:139.

[7] Shortliffe, EH.,Perrault, LE., (Eds.). Medical informatics: Computer applications in health care and biomedicine (2nd Edition). New York: Springer, 2000

[8] Thamar Solorio and Olac Fuentes, “Improving Classifier Accuracy using Unlabeled Data”. Proceedings of the IASTED International Conference on Artificial Intelligence and Applications (AIA2001), Marbella, Spain, Sept. 2001.

[9] Wang, Y. and Wong, A.K.C.;From association to classification: inference-using weight of evidence, IEEE Transactions on Knowledge and data engineering , Volume: 15 , Issue: 3 , May-June 2003 Pages:764 – 767

[10] Witten, T.H and Frank, E. 2000 Data mining: Practical machine learning tools and techniques with Java implementations. Morgan Kaufmann, San Francisco.