

Case Study on Data Mining Application in Health Care Monitoring Systems

Prof. GVNKV Subbarao¹, Prof.MD Sameeruddin Khan², K.VIJAY KUMAR³
^{1,2,3} CSE Department, Sree Dattha Institute of Engineering & Science

Abstract: In this research paper, we have focused to compare a variety of techniques and various approaches and different types of tools and its effect on the healthcare sector. The objective of data mining application is to find out that data are nothing but facts, numbers, or text which can be processed by a computer into knowledge or information. The purpose of data mining application in healthcare monitoring systems is to develop an automated tool for identifying and disseminating relevant healthcare information. This paper aims to make a study report of different types of data mining applications in the healthcare monitoring and to reduce the complexity of the study of the healthcare data transactions. Also presents a detailed study of different data mining applications, technologies and different methods applied for previous knowledge from database healthcare industry. Finally, the existing data mining techniques concepts with data mining algorithms and its application tools which are more valuable for healthcare services are discussed in detail in this research paper in the fullest perspective.

Keywords: Data Mining, Knowledge Discovery Database, Data Mining classification, Data Mining Application, WEAK.

I. Introduction

The objective of data mining is to extract useful information from large databases. Data mining applications are useful for commercial and scientific sides. This paper mainly discusses the Data Mining applications in the scientific application. Scientific data mining distinguishes itself in the sense that the nature of the datasets is often unique from traditional market driven data mining applications. In this research paper, a detailed survey is carried out on data mining applications in the healthcare sector, different types of data used and various information extracted.

Data mining algorithms applied in healthcare sector play a major role in prediction and diagnosis of the diseases. There are different types of data mining applications found in the medical field such as Medical device industry, Pharmaceutical Industry and Hospital. To find the useful and hidden information from the database is the objective behind the application of data mining. Popularly data mining called information discovery from the data. The information discovery is an interactive process, consisting by developing an understanding of the application domain, selecting and forming different data set, preprocessing, data transformation.

Data Mining has been used in a various types of applications such as marketing, customer relationship management, engineering, and medicine analysis, expert prediction, web mining application , mobile and mobile computing.

Data mining:

Data mining is the non trivial process of verifying valid, novel, potentially useful, and ultimately understandable different patterns in data. Data mining is the use of automated data analysis techniques to uncover previously undetected relationships among data items. Data mining often include the analysis of data saved in a data warehouse.

Definition:

Data mining or knowledge discovery in database, as it is also known, is the non-trivial extraction of implicit, unknown and potentially very useful information from the data. This collect different number of technical approaches, like as clustering, data summarization, classification, verifying dependency networks, analyzing changes, and detecting anomalies.

Development of data mining:

The current evaluation of data mining functions and products is the results of influence from many disciplines, including various types of databases, information retrieval, statistics, algorithms, and machine learning (See Fig. 1).

Fig. 1. Historical perspective of data mining

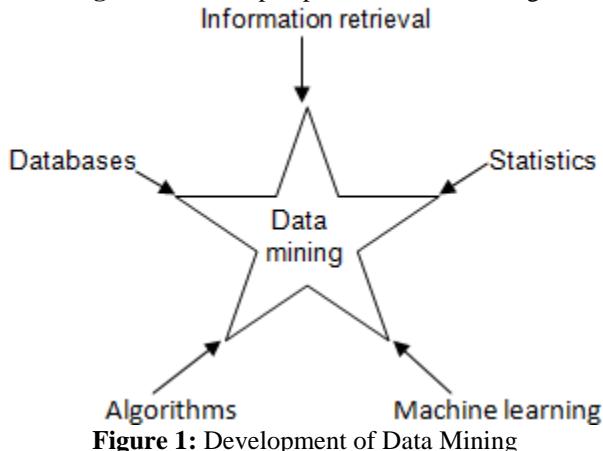


Figure 1: Development of Data Mining

KDD and Data Mining:

The motivation for handling data and performing computation is the finding of information. In this research paper, we store data about a certain process and analyzed later that information in order to useful it in a meaningful manner. The KDD process employs data mining methods to find out patterns at some measure of interestingness. The KDD is the method of turning the low-level data into high-level data. Hence, Frawley et al. referred the knowledge discovery in database as the nontrivial process of identifying valid, novel, potentially useful, and ultimately understanding patterns in data mining application. The terms Knowledge Discovery in Databases (KDD) and Data Mining are often used interchangeably. This dilemma is due to three different perspectives to look at the data mining but in real data mining is an important method in the KDD process. Han et al. defined the data mining as the process of discovering interesting and useful information from large amounts of data stored either in databases, data warehouses, or other various information repositories. Witten et al. defined data mining as the process of extracting implicit, previously unknown and potentially is very useful knowledge from databases. Hand et al. defined data mining as the analysis of observational data set to identify unsuspected relationships and to summarize the data in novel ways that are both understandable and useful to the data owner.

On studying the various definitions on data mining Zhou finally suggested that the database, machine learning and statistics perspectives of data mining put particular emphases on efficiency, effectiveness and validity respectively. Data mining refers to the nontrivial extraction of implicit, previously unknown and potentially is very useful information from data in databases. It uses various machine learning, statistical and visualization techniques to find out and present knowledge in a form which is easily comprehensible to us in proper manner. The richness and rapid evolution of the data mining discipline comes from its large variety of research field of interest. Data mining applications can use various kind of parameters to analyzed the data. They include an association, sequence or path analysis (patterns where one event leads to another event), classification (identification of new patterns with predefined targets) and clustering techniques.

II. Literature Review

A literature review is a text written by critical points of current knowledge including substantive find theoretical and methodological contributions to a particular topic. Literature reviews are secondary sources and do not report any new or original experimental work.

Hian Chye Koh and Gerald Tan mainly discusses data mining and its applications with major areas like Treatment effectiveness, Management of healthcare, Detection of fraud and abuse, Customer relationship management [1].

Jayanthi Ranjan presents how data mining discovers and extracts useful patterns of this large data to find observable patterns. This paper demonstrates the ability of Data mining in improving the quality of the decision making process in pharma industry. Issues in the pharma industry are adverse reactions to the drugs [2].

M. Durairaj, K. Meena illustrates a hybrid prediction system consists of Rough Set Theory (RST) and Artificial Neural Network (ANN) for dispensation medical data. The process of developing a new data mining technique and software to assist competent solutions for medical data analysis has been explained. Propose a hybrid tool that incorporates RST and ANN to make proficient data analysis and indicative predictions. The experiments on spermatological data set for predicting excellence of animal semen is carried out. The projected hybrid prediction system is applied for pre-processing of medical database and to train the ANN for production prediction. The prediction accuracy is observed by comparing observed and predicted cleavage rate[10].

Classification techniques on healthcare data:

Data mining has been applied to a different variety of healthcare sector or management to improve decision making and its application.

But one of the major challenges in healthcare sector is the extraction of comprehensible information. Here, the application of data mining classification techniques is studied which are applied on the different diagnostic datasets and application.

Orlando Anunciacao et al. explored the applicability of decision trees for detection of high risk breast cancer groups over the dataset produced by Department of Genetics of faculty of Medical field of Universidad Nova de Lisboa with 164 controls and 94 cases in WEKA machine learning tool and technology . To statistically validate the association found, different permutation tests were used. They found a high-risk breast cancer group composed of 13 cases and only 1 control, with a Fisher Exact Test value of 9.7×10^{-6} and a p-value of

0.017. These results conclude that it is possible to identify statistically significant associations with breast cancer by deriving a decision tree and choosing the best leaf. A. Soltani Sarvestani et al. provided a comparison among the capabilities of different types of neural networks such as Multilayer Perceptron (MLP), Self Organizing Map (SOM), Radial Basis Function (RBF) and Probabilistic Neural Network(PNN) which are used to classify WBC and NHBCD data. The presentation of these neural network structures was investigated for breast cancer diagnosis related problem. RBF and PNN were proved as the best classifiers in the training set.

But the PNN gave the best classification accuracy when the test set is considered. This paper shows that statistical neural networks can be effectively useful for breast cancer diagnosis as by applying several types of neural network structures a diagnostic system was constructed that performed quite well. Dr. Medhat Mohamed Ahmed Abdelaal et al. investigated the capability of the classification SVM with Tree Boost and Tree Forest in analyzing the DDSM dataset for the extraction of the mammographic mass features along with age that discriminates true and false cases. Here, SVM techniques show promising results for increasing diagnostic accuracy of classifying the cases witnessed by the largest area under the ROC curve comparable to values for tree boost and tree forest.

History of data base and data mining:

Data mining development and the history represented in the Fig. 2. The data mining system started from the year of 1960s and earlier. In this paper , the data mining is simply based on file processing. The next step its Database management Systems to be started year of 1970s early to 1980s. In this OLTP, Data modeling tools and Query processing are working. From database management system there are three broad categories to be worked.

First is Advanced Database Systems, this evaluated year of Mid- 1980s to present in this Data models and Application oriented process are worked. The Second is Data Warehousing and Data Mining worked since the year of the late 1980s to present. The third part is Web based Database Systems which is started from 1990s to present and in this Web mining application and XML based database systems are included. These are three broad categories are joined and form the new process that is called New generation of the Integrated Information system which is started in 2000.

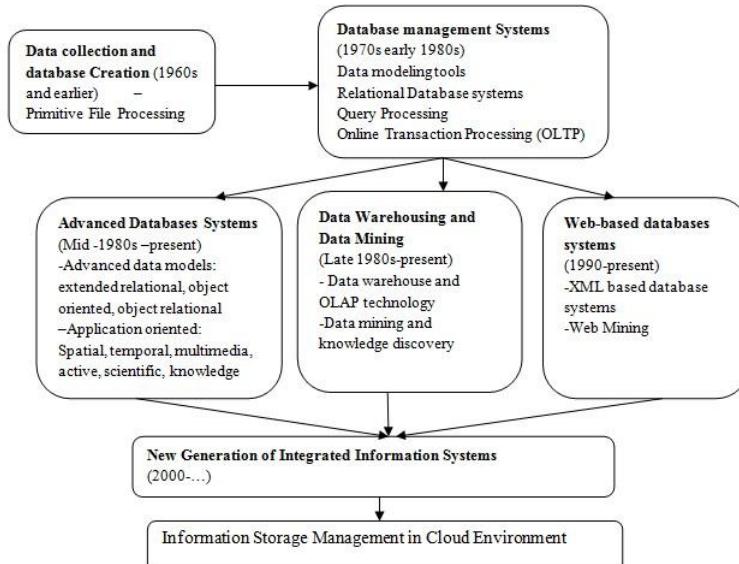


Figure 2: History of Database Systems and Data Mining

Data mining applications in healthcare sector:

Healthcare sector nowadays creates a large amounts of complex data about patients, hospital resources, disease diagnosis, electronic patient records and various types of medical devices . Larger amounts of data are a key resource to be processed and analyzed for knowledge or information extraction that enables support for cost-savings and decision making. Data mining applications in healthcare sector can be combined as the evaluation into broad categories.

Treatment effectiveness:

Data mining applications can develop to evaluate the effectiveness of medical treatments in hospitals. Data mining can deliver an analysis of which course of action proves effectiveness by comparing and contrasting causes, symptoms, and different courses of treatments.

Healthcare management:

Data mining applications can be developed to better find out and track chronic diseases states and high-risk patients, design appropriate interventions, and minimize the different number of hospital admissions and claims to aid healthcare management. Data mining is used to analyze massive volumes of data and statistics to search for patterns that might shows an attack by bio-terrorists.

Customer relationship management:

Customer relationship management is a core process to maintaining interactions between commercial organizations and industries such as banks and retailers-and their customers, it is more important in a healthcare management. Customer interactions may occur through call centers, physicians' offices, billing departments, inpatient settings, and healthcare settings.

Fraud and abuse:

Detect fraud and abuses establish norms and then find out abnormal patterns of claims by physicians, hospitals, or others attempt in various data mining applications. Data mining applications fraud and abuse applications can highlight prescriptions which are not appropriate or referrals and fraudulent insurance and medical claims.

III. Conclusion

This paper aimed to compare the different data mining application in the healthcare management for extracting useful information. The prediction of diseases using different data mining applications is a challenging task and it drastically minimizes the human effort and increases the diagnostic accuracy. Healthcare organization and agencies could look to found ideas on how to extract knowledge from their own database system.

References

- [1]. HianChyeKoh and Gerald Tan, —Data Mining Applications in Healthcare, journal of Healthcare Information Management – Vol 19, No 2.
- [2]. JayanthiRanjan, —Applications of data mining techniques in pharmaceutical industry, Journal of Theoretical and Applied Technology, (2007).
- [3]. RubanD.Canalas Jr., MSIT., MBA , — Data mining in Healthcare: Current applications and issues.
- [4]. K. Srinivas , B. Kavitha Rani and Dr. A. Goverdhan, —Applications of Data Mining Techniques in Healthcare and Prediction of Heart Attacks International Journal on Computer Science and Engineering (2010).
- [5]. ShwetaKharya, —Using Data Mining Techniques ForDiagnosis And Prognosis Of Cancer Diseases, International Journal of Computer Science, Engineering and Information Technology (IJCSEIT), Vol.2, No.2, April 2012.
- [6]. EliasLemuye, —Hiv Status Predictive Modeling Using Data Mining Technologyl.
- [7]. Arvind Sharma and P.C. Gupta —Predicting the Number of Blood Donors through their Age and Blood Group by using Data Mining Tool International Journal of Communication and Computer Technologies Volume 01 – No.6, Issue: 02 September 2012.
- [8]. Arun K Punjari, —Data Mining Techniques, Universities (India) Press Private Limited, 2006.
- [9]. Margaret H.Dunham, —Data Mining Introductory and Advanced Topics, Pearson Education (Singapore) Pte.Ltd., India. 2005.
- [10]. PrasannaDesikan, Kuo-Wei Hsu, JaideepSrivastava, —Data Mining For Healthcare Managementl, 2011SIAM International Conference on Data Mining, April, 2011.
- [11]. ShusakuTsumoto and Shoji Hirano, —Temporal Data Mining in Hospital Information Systems.