Research on key technology of data mining based on hospital information system

Rui Wang [1,a], Jinguo Wang [2,b*], Na Wang [3,c]

[1]Department of Information Engineering,Jilin Business and Technology College China

[2]Department of Urology, the First Hospital of Jilin University,China

[3]Department of Anaesthesiology, the First Hospital of Jilin University, China

[a]Xiaoben6666@126.com,[b]wangjinguolily@163.com,[c]lilyly12345@163.com

*corresponding author

Key word: data mining. hospital information system. HIS.

Abstract: At present, the processing of medical information is mostly based on the level of the operation of database technology, it is the specific application service. Hospital information system is a branch of medical informatics, which has accumulated a lot of management information and clinical information resources in hospital information system. How to excavate the deep level, implicit and valuable knowledge in the large amount of data resources is a difficult problem that we face. Data mining has the ability to solve this problem. Using data mining technology to carry out scientific research, improve the level of medical technology and hospital management is very necessary.

1. Introduction

Hospital Information system(HIS) is a comprehensive management of the hospital and the Department of the hospital and its affiliated departments of the hospital and its affiliated departments[1]. The data collected, stored, processed, transmitted, collected, processed and generated by the various departments of medical activities. Hospital information system is an indispensable infrastructure and supporting environment in the construction of a modern hospital. With the rapid development of computer technology today, there is no computer in the hospital wide application; there is no modern hospital management, modern hospital construction and modernization of medical standards. The application of computer and hospital information system has become a new breakthrough point for the hospital to strengthen scientific management, the hospital service quality competition new starting point, the hospital technology construction affects the new commanding heights of various disciplines and links, the hospital in-depth development of economic management and improve the efficiency of the new growth point[2].

The computer information technology application in the medical field has accumulated a large amount of data resources, covering the medical activities of the text, images, sound and other multimedia data. How to use data mining technology to obtain knowledge is becoming more and more attention. With the support of computer and database technology, the data has become an important resource for the implementation of scientific management and research in medical field. The database scale is growing and the complexity of the database is growing[3]. It is becoming more and more important to get the policy information from the large amount of data. It is based on this new demand. The appearance of data mining technology provides a technical tool for medical

administrators, scientific research workers to analyze and use these data resources for scientific management, decision making and medical research. Exploring data mining technology in medical information research has more important practical value and broad prospects for development[4][5].

2. Data mining

Data mining technology is the result of the gradual evolution of information technology[6]. It is the result of the research and development of the database technology in the long term. At first, all kinds of business data is stored in the database, and then gradually to the database of business data query, and then to develop the real-time data. Data mining makes the application of database technology into a more advanced stage; it can not only query past data, and can find the intrinsic relationship between the data, so as to increase the depth of the application of information. With the development of data collection, powerful multi-processor computer and data mining algorithm based on the three basic technologies, data mining technology has been widely used in commercial applications[7].

The definition of data mining is a lot of different ways of expression. From a technical point of view, data mining is a process of information and knowledge from a large amount of data, which is not completely, with noise, fuzzy and random. The main features of the commercial database is to extract, transform, analyze and model the business data, and extract the key knowledge from the auxiliary business decision, namely, from a database to find relevant business model: data mining is stored in the database, data warehouse or other information repository in the process of mining interesting knowledge.

3. Commonly used data mining technology

Statistical analysis. The data mining model used in statistical analysis is linear analysis, nonlinear analysis, regression analysis, logistic regression analysis, univariate analysis, multivariate analysis, time series analysis, nearest neighbor algorithm and clustering analysis. Using these techniques, we can check the abnormal data, and then use a variety of statistical models and mathematical models to explain the data, explain the market rules and business opportunities hidden behind these data. For example, you can use statistical analysis tools to find the best business opportunities, increase market share and profit; the use of quality management procedures, improve the quality of products or services, so that customers are more satisfied; through the adjustment of product manufacturing or business restructuring, increase profits. In all data mining techniques, the statistical data mining technology is one of the most mature, and has been widely used in data mining.

Knowledge discovery class. Knowledge discovery data mining technology is a kind of data mining technology, which is completely different from the statistical data mining. It can filter information from a large amount of data warehouse, looking for the new operating mode that may appear in the market, and find out the fact that people do not know. Knowledge discovery data mining techniques include artificial neural network, decision tree, genetic algorithm, rough set, rule discovery and association order, etc..

4. Key techniques of medical data mining

Data preprocessing. data preprocessing is an important step in data mining process, especially when the database contains noise, incomplete, and even inconsistent data, it needs the data preprocessing.

In a complete data mining, data preprocessing usually takes about 60% of the time, while the excavation work is only about 10% of the whole process. Data preprocessing includes data cleaning, data integration, data conversion and data reduction.

Anonymity and identity conversion. Because medical information is related to patient privacy information, medical data need to be processed in special data processing, which is to make a special data processing.

Anonymous refers to the identification of patients who are removed from the record, such as the name, address, hospital records, etc., or use the wrong logo instead of the correct identification. After anonymity, researchers cannot know any information about the patient by observing the record. Logo conversion and anonymity have some minor differences. After the transformation, the identity of the patients may still be implied, but the implied true information can only be obtained by the authorized researchers, and the unauthorized persons cannot obtain any real information of the patients through the implied clues. Anonymous refers to the identification of patients who are removed from the record, such as the name, address, hospital records, etc., or use the wrong logo instead of the correct identification. After anonymity, researchers cannot know any information about the patient by observing the record. Logo conversion and anonymity have some minor differences. After the transformation, the identity of the patients may still be implied, but the implied true information can only be obtained by the authorized researchers, and the unauthorized persons cannot obtain any real information of the patients through the implied clues.

Medical text data mining. medical text information, medical experts on the image, signal or other clinical data interpretation is not standardized, it is difficult to directly carry out data mining, and therefore need to be standardized for text data processing. Currently, it has been a certain effect on the standardization of medical text data by computer. Machine conversion mainly consists of three steps: the analysis of the source statement, the conversion. One difficulty of the conversion is that the source statement is not unique. So we need to collect all kinds of forms of the source statement. The current machine translation can only handle statements that are less than 10 words. XML is a structured language, which provides another way of standardization of text data.

XML not only can create text that contains the structured data, but also can share and process data. XML is the key technology of data mining and knowledge discovery, which is mainly reflected in the communication between various data mining tools and data mining tools and data warehouse, the data mining tools are implemented based on different software platforms.

Image data mining. the current medical imaging data are mainly from some imaging instruments, they have been more and more medical experts as a reliable diagnostic tool. Therefore, the development of effective image data mining tools has become one of the key technologies in medical data mining, which is not only different from the data mining methods, but also difficult to achieve. Medical image data mining mainly includes: remove or reduce the impact of image noise, improve the target image quality or edge extraction of the target tissue, the concept description of the target organization, and summarize the relevant characteristics of the object, so as to obtain or verify the parameters of the dynamic range; medical impact data management and retrieval. At present, it has made a breakthrough in data mining of SPECT image.

References

[1] Muhaimenul Adnan,Reda Alhajj. DRFP-tree: disk-resident frequent pattern tree[J]. Applied Intelligence . 2009.

[2] Shi Zhong. Semi-supervised model-based document clustering: A comparative study[J]. Machine Learning . 2006.

[3] Rakesh Agrawal,Johannes Gehrke,Dimitrios Gunopulos,Prabhakar Raghavan. Automatic Subspace Clustering of High Dimensional Data[J]. Data Mining and Knowledge Discovery . 2005.

[4] Tomasz Imieliński,Aashu Virmani. MSQL: A Query Language for Database Mining[J]. Data Mining and Knowledge Discovery . 1999.

[5] Xin-Dong Wu,Xing-Quan Zhu,Qi-Jun Chen,Fei-Yue Wang. Ubiquitous Mining with Interactive Data Mining Agents[J]. Journal of Computer Science and Technology . 2009.

[6] Elisa Bertino,Giovanna Guerrini,Marco Mesiti. Measuring the structural similarity among XML documents and DTDs[J]. Journal of Intelligent Information Systems . 2008.

[7] Jiawei Han,Hong Cheng,Dong Xin,Xifeng Yan. Frequent pattern mining: current status and future directions[J]. Data Mining and Knowledge Discovery . 2007.