

Research on Data Mining and Knowledge Discovery of Computer Aided Medical Diagnostic System

Yang Hong¹, Ye Donghai², Lv Feng^{2*}

¹College of Information, Yunnan University of Chinese Medicine, Kunming, Yunnan, China

²Library, Yunnan University of Chinese Medicine, Kunming, Yunnan, China

*Corresponding Author

Keywords: Computer Aided, Medical Diagnostic System, Data Mining, Knowledge Discovery

Abstract: the Rapid Development of Medical Testing Technology Enables Patients to Generate Large Amounts of Data during Their Treatment. How to Analyze These Data in Depth and Find That Valuable Medical Rules Have Become a Research Hotspot in the Medical Field. Based on the Actual Situation of the Current Medical Diagnosis System, This Paper Analyzes the Specific Application Forms of Data Mining and Knowledge Discovery Technology in the Medical Field, and Points out That Data Mining and Knowledge Discovery Technology Have Certain Application Prospects in the Aspects of Association Analysis, Cluster Analysis, Prediction Trend and Control Deviation. on This Basis, a Computer-Aided Medical System Based on Data Mining and Knowledge Discovery Technology is Further Established, in Order to Provide a Useful Reference for the Hospital to Improve the Diagnostic Level and Optimize the Diagnostic Services.

1. Introduction

1.1 Literature Review

With the rapid development of information technology, computer technology is deeply integrated with various disciplines. In the medical field, many scholars have carried out research on the application of computer technology in diagnostic systems. Qiang Yonggan, Guo Youmin and others constructed a computer-aided medical diagnosis system based on Web text level to study the diagnosis of solitary pulmonary nodules (SPN). The system is further subdivided into a case upload subsystem, a maintenance management subsystem, and an auxiliary diagnostic subsystem. The diagnostic results will appear in the form of a report, along with a CT image and a diagnostic opinion. Diagnostic opinions are divided into malignant, benign, potentially malignant, potentially benign, and indeterminate levels with credibility (Qiang et al, 2011). Wang Weisheng, Luo Jiawei and others pointed out that one of the main factors restricting computer-aided diagnosis is the lack of a unified data storage model. In response to this, several scholars have built a data platform mainly for medical computer-aided diagnosis. For the impact data and process data in the diagnosis process, detailed analysis is carried out, and a unified data model is proposed and implemented by Oracle database (Wang et al, 2013). In order to solve the problem of target designation in automatic detection of medical images, Tao Pan and Fu Zhongliang proposed a deep learning-based algorithm, which is mainly used to detect target position and predict target posture. On this basis, the magnetic resonance imaging and transesophageal echocardiography were selected as experimental subjects to verify the accuracy of the algorithm. The results show that the algorithm is accurate and effective (Tao et al, 2019). Xiao Huanhui and Yuan Chenglang et al. focused on the classification of benign and malignant breast lesions and pulmonary nodules based on computer deep learning theory (Xiao et al, 2019). Li Yeshun and Bi Kai introduced data mining and knowledge discovery technology into the expert diagnosis system, proposed a new system framework and conducted experiments. The results show that the new system can help to improve the knowledge acquisition difficulty of the diagnostic expert system (Li et al, 2018).

1.2 Purpose of the Study

In the 21st century, the widespread use of computer and communication technologies has caused information to explode. While information is expanding, it is difficult for people to directly discover the hidden knowledge, which is the contradiction between rich data and knowledge discovery. Therefore, data mining and knowledge discovery have become one of the key research directions in the current computer field. With the hospital's computing plan and intelligent progress, many hospitals have begun to use the PACS system to collect medical influences and related medical parameters of a large number of patients, how to make full use of these confirmed cases, assist doctors in medical diagnosis, and help them quickly, Accurate judgment of the patient's condition is an urgent need for breakthroughs in the party-building computer-aided medical diagnosis system. In view of this, this paper deeply studies the data mining and knowledge discovery technology in computer-aided medical diagnosis system, in order to improve the modern medical diagnosis level.

2. Theoretical Overview

The so-called computer-aided diagnosis (CAD) specifically refers to the use of imaging, medical image processing technology and other physiological and biochemical methods to detect lesions in time and improve the accuracy of diagnosis. To some extent, computer-aided diagnosis can be seen as an extension of the auxiliary test, which can be turned over to support the diagnosis (Wang and Yang, 2019). For this reason, computer-aided diagnosis technology is also called the doctor's third eye, and its wide application is more helpful to improve the speed and accuracy of diagnosis. Currently, computer-aided diagnosis is mainly applied to medical imaging. The steps can be divided into three steps. One is image pre-processing, which extracts the lesion from the normal structure. The second step is the process of extracting and quantifying image signs, and quantifying the features of the lesions, such as size, density, and morphological features. The third step is the data processing process, the quantitative data is transmitted to the corresponding algorithm, and the CAD diagnosis system is used for classification processing. In this process, the neural network, rule extraction, decision tree and other methods are mainly used (Zhang and Bu, 2019).).

In the process of computer-aided diagnosis, there are two technologies that cannot be separated, namely data mining and knowledge discovery. Data mining refers to the process of filtering out the valid information hidden in it from a large amount of data through a specific algorithm. Data mining is a specific process of decision support. By using artificial intelligence, statistics, database, visualization technology, etc., it is highly automated to analyze data and make inductive sorting. The purpose of data mining is to promote knowledge discovery. The data mining process can be roughly divided into the following seven steps.

They are to define problems, build mining libraries, analyze data, prepare data, build models, evaluate models, and implement them. Knowledge Discovery (KDD) refers to the process of discovering knowledge from various information and according to its own needs. To a certain extent, the purpose of data mining is to discover knowledge, and the process of knowledge discovery must include data mining technology. KDD technology can be briefly summarized as a process for determining that data is valid, available, and understandable.

3. Application of Data Mining and Knowledge Discovery Technology in Medical Field

The variety of medical information includes text (patient symptom description, personal information, etc.), images (B ultrasound, EF and other medical imaging equipment), signals (electrocardiogram, EEG signals, etc.), pure data (test results, physical parameters) Etc.), animation, voice, video information. This is also a major feature of the medical field that distinguishes it from other fields. However, the variety of medical information also increases the difficulty of medical data mining. Due to the limitations of medical data, mining medical data needs to be transformed into a transactional data format containing inclusions. First, the name, address, etc., which have no

practical significance for diagnosis, are processed, and then the average value is used to fill the gap value, and finally the data is discretized into continuous attributes.

Applying data mining and knowledge discovery technology can help doctors discover hidden knowledge, predict future trends, and make scientific decisions. Specifically, there are mainly the following four aspects.

The first is to conduct correlation analysis. One of the core technologies in data mining technology is mostly data association technology. There are three ways of association, among which the correlation analysis is mainly used for detection, and it is also one of the main applications of the current hospital. For example, most technologies use this technique when researching diabetes data. This is due to the many complications of diabetes. At present, medical knowledge cannot quantitatively detect the relationship between blood glucose concentration and complications. With the aging of the population, the number of people with diabetes is increasing. Combine various physiological indicators of diabetic patients, and establish a related database to analyze the relationship between various parameters.

The second is to automatically predict trends. One of the goals of medical data mining is to predict disease and prevent it in advance. Therefore, there are many reports on such medical data, mainly using linear regression, nonlinear regression and generalized regression models. In recent years, fuzzy control technology and neural network technology have also been used in medical data mining, and achieved great success.

The third is cluster analysis. Due to the wide variety of medical data, an important step in data mining and knowledge discovery is to select classifications and then classify them into various combinations according to attributes, also known as clustering. Through clustering, medical data can be expressed more objectively and accurately, and it is also helpful to describe medical concepts. The completion of the clustering process requires genetic algorithm and neural network technology in data mining and knowledge discovery technology.

The fourth is to test the control deviation. In the medical database, occasionally some abnormal data will appear, and it is extremely important to find the data. Once these data are found, further testing is required, compared to the reference results, and the cause is analyzed. There are two main application scenarios. The first one is artificial auxiliary instruments such as ventilators and pacemakers. The second is the specific application effect of various drugs in the disease, so as to get more accurate feedback results.

4. Computer-Aided Medical System Based on Data Mining and Knowledge Discovery Technology

At this stage, many doctors still stay in the traditional experience stage for the diagnosis of the disease, and the level of diagnosis is closely related to the doctor's business ability. In particular, some intractable diseases often require one or several expert consultations to determine. In view of this, this paper attempts to develop auxiliary diagnostic systems using high-speed computer capabilities. On the basis of a large number of practices, a diagnosis and treatment system that can make full use of data mining and knowledge discovery technology has been developed to improve the doctor's diagnosis level. The specific system structure is shown in Figure 1.

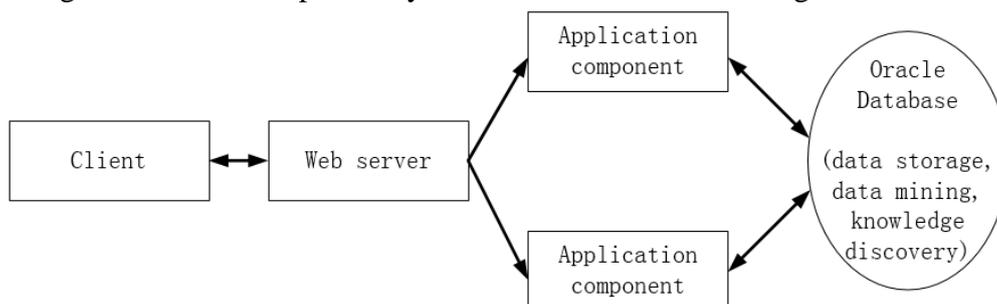


Fig.1 Computer-Aided Medical System Based on Data Mining and Knowledge Discovery Technology

As can be seen from Figure 1, the software system framework is used for the designed system for ease of use. The overall functional framework of the system uses the more mature J2EE architecture technology, and then establishes the corresponding information management platform, inspection rule base, data mining and analysis platform and case feature database. Among them, the web server is mainly used to receive HTTP page requests. The application server mainly uses J2EE to complete the application request of the web server.

The system also has the function of receiving the inspection data of the LIS system. The data is input into the server through the Internet, and the specific rules of the medical knowledge base and the large protective gear mining in the database are specifically compared, and corresponding opinions are given. In addition, with the help of the Internet, medical institutions and experts at all levels can conduct online consultations to get rid of the dependence of time and space. The database adopts the form of Oracle database, which not only needs to be responsible for data storage, but also needs to actively explore according to the aggregated data, thereby promoting knowledge discovery and improving medical level.

References

- [1] Qiang Y.G., Guo Y.M., Wang Q.P., et al.(2013) Computer-aided Diagnosis of Solitary Pulmonary Nodule Imaging Based on Web platform. *Journal of Xi'an Jiaotong University (Medical Sciences)*, 32(3): 367-370.
- [2] Wang W.S., Luo J.W., Lin H.L., et al.(2013) Research on Computer Aided Diagnostic Data Platform for Medical Image. *Chinese Journal of Biomedical Engineering*, 32(1), 105-108.
- [3] Tao P., Fu Z.L., Zhu W., et al.(2018). Research on medical computer aided detection based on deep learning. *Biomedical Engineering Journal*, 35(03),48-55.
- [4] Xiao H.H., Yuan C.L., Feng S.T., et al.(2019). Progress in computer-aided classification diagnosis of cancer based on deep learning. *International Journal of Medical Radiology*, 42(01): 25-28+61.
- [5] Li Y.S., Bi K., Zhao S.L.(2018). Research on Diagnostic Expert System Based on Knowledge Discovery and Data Mining Technology. *Electronic Technology and Software Engineering*, 7(1), 141-142.
- [6] Wang W., Yang Z.H.(2019). Discussion on Related Problems of Computer Aided Medical Diagnosis. *Computer Knowledge and Technology*, 15(11), 272-273.
- [7] Zhang W.J., Bu H.F.(2019). New Promotion of Subject Service Based on Data Mining Knowledge Discovery System--Taking Taizhou Vocational and Technical College as an Example. *Journal of Taizhou Vocational and Technical College*, 19(2), 46-18.