Research on the Application of Data Mining in the Medical Field

Yishan Liang[1]

[1] The School of Computer Science and Technology at Chongqing University of Posts and Telecommunications, Chongqing 400065, China
2021212083@stu.cqupt.edu.cn

Abstract. Currently, there is a growing demand for the healthcare industry in society, generating vast amounts of complex data that are difficult for humans to process. Therefore, relying on data mining technology to analyze data can help the development of the medical field. This paper discusses the application of data mining in the medical field. Regarding data mining techniques, this paper references several articles to identify commonly used data mining algorithms. In the healthcare domain, concerning medical management and resource allocation, this paper refers to hospital management cases and analyzes the specific applications of data mining technology in hospital business, cost-effectiveness, and optimization of hospital resource allocation. For clinical medical diagnosis, this paper references some highly completed medical data mining models and analyzes their specific applications in predicting diseases, assisting clinical diagnosis, and treatment, thereby improving the accuracy and efficiency of clinical treatment. Finally, based on existing research, the paper proposes suggestions for the development of data mining in the medical field application.

Keywords: Data mining, Clinical medicine, Machine learning, Medical management.

1 Introduction

Currently, with the rapid development of technology, massive amounts of data are generated in the operations of various industries. The ability to correctly analyze these data plays a crucial role in the development of various fields, while relying on traditional methods involving manual data manipulation proves to be highly inefficient.

Taking medical data as an example, medical data exhibits characteristics of disease diversity, heterogeneity of treatments and outcomes, as well as the complexity of data collection, processing, and interpretation [1]. The quantity and complexity of these data are extremely high, and relying on manual processing would increase time costs and inevitably lead to errors. In such a scenario, the accuracy and efficiency of medical decision-making would be greatly reduced. Data mining technology, however, can analyze these data to discover patterns and trends. Data mining is the

Y. Wang (ed.), *Proceedings of the 2024 International Conference on Artificial Intelligence and Communication (ICAIC 2024)*, Advances in Intelligent Systems Research 185,
https://doi.org/10.2991/978-94-6463-512-6_30

process of discovering interesting patterns and knowledge from large amount of data [2]. It involves detecting and extracting hidden information, patterns and specific data connections of the prediction idea [3]. By utilizing data mining technology, the efficiency of data analysis can be improved, and substantial analysis results can be obtained.

Regarding diseases, data mining can analyze the patterns and characteristics of disease occurrence, transmission, and treatment. It helps people better understand the correlations between different diseases, predict the development trends of diseases, and improve the efficiency of disease prevention and treatment. Nowadays, billions of people worldwide rely on medical diagnoses to maintain their health, indicating the continuous need for the development of the medical field to meet people's needs. For many countries' medical institutions, advancing the research and application of medical big data has become a key factor driving modern medical research [1]. It's not just about the technology to treat diseases; this also applies to maintaining and improving the operations of various institutions.

This study explores how data mining is applied in the medical field to address the challenges and problems faced by the existing medical system, thereby better meeting the global population's health needs. The study first discusses the application of data mining technology in solving the management and resource allocation issues of hospitals as public service institutions. This includes ensuring operational efficiency, optimizing business structures, timely adjusting operational strategies, and improving operational effectiveness. Based on patient's medical needs and the supply of medical resources, resource optimization allocation is conducted to ensure the rational distribution of medical resources. Subsequently, the study discusses the application of data mining in clinical medical diagnosis. By referencing past highly completed disease prediction models, it analyzes the contributions of data mining technology to predicting diseases, assisting clinical diagnosis, and treatment.

2 Overview of Data Mining Techniques

2.1 Classification

Classification is a data analysis technique where items are assigned to target categories within a set. This technique groups items with similar feature sets into a category [3]. It is the process of identifying a collection of models to distinguish between concepts and data types [4].

Classification with Machine learning can be efficiently used in difficult measuring applications in medicine [5]. Common classification methods include decision trees, Bayesian networks, artificial neural networks, support vector machines, etc.

2.2 Clustering

Clustering is the process of grouping items based on certain attributes, aiming to maximize the differences between clusters while maximizing the similarities within clusters [6].

Unlike classification techniques, clustering techniques can determine the categories and objects within each cluster. Classification techniques, on the other hand, involve assigning objects to predefined categories [3].

Common clustering methods include neural network methods and the K-means method. Compared to other clustering methods, the k-means algorithm and its variants are efficient in clustering large data sets, thus very suitable for data mining [7].

2.3 Decision Tree

The decision tree algorithm is a commonly used method for summarizing logic, based on conditional probabilities. It is one of the methods for approximating discrete-valued expected functions [3]. This algorithm analyzes data under classification concepts and constructs a tree, where each node of the tree predicts based on the values of different columns in the training dataset [8]. Decision trees present category partition patterns through tree structures, with nodes within the tree representing continuous classification of data [8].

2.4 Artificial Neural Networks

Artificial Neural Network (ANN) is an algorithm based on biological neural networks, used to estimate or approximate functions that rely on a large number of typically unknown inputs [3]. This algorithm mimics biological neural networks, employing artificial neurons as computational units and forming a network through connections between neurons [9].

Due to the characteristics of neural network algorithms, they can intelligently handle nonlinear, incomplete, uncertain, and inaccurate data. Medical data is often incomplete, uncertain, and inaccurate [9]. Therefore, neural networks exhibit good fault tolerance, robustness, and high accuracy when dealing with medical data, meeting the accuracy requirements of medical data mining models [9].

3 Analysis of the Application of Data Mining in the Medical Field

3.1 Application in Medical Management and Resource Allocation.

Hospital Operations and Cost-effectiveness. Analyzing the cost-effectiveness status of hospitals allows for a comprehensive understanding of the hospital's operational status and provides insights into its financial situation and fund trends. For example,

analyzing dynamic trends in the inventory levels of different medications enables the adoption of strategies to reduce medication inventory levels, thus accelerating capital turnover.

In the field of medical business, the core part of data mining and analysis focuses on further exploration and analysis of hospital financial data [10]. Through financial analysis, hospitals can grasp their revenue, cost structures, and understand the contribution of different business sectors, thereby evaluating the hospital's profitability. For instance, utilizing data mining techniques such as ratio analysis and trend analysis to analyze hospital financial growth, and fitting curves to estimate future cash demand, providing reference for hospital decision-making.

Not only can financial analysis serve as the basis for selecting key hospital investments, but data mining can also utilize revenue data as a reference in terms of time dimension, distinguishing between peak and off-peak seasons for hospitals, further rationalizing hospital fund planning to ensure normal hospital operations.

Data mining techniques can equally be applied to expanding the scope of hospital business. Through data mining and analysis, patient characteristics can be summarized, thereby tailoring different treatment or prevention plans for different populations. This can enhance the hospital's proactive ability to grasp and intervene in specific population health risks, ultimately expanding the hospital's business scope [10].

Taking the teacher population as an example, due to the prolonged standing work environment, teachers are more prone to lumbar and cervical spine diseases. In response to this characteristic, hospitals can utilize data mining and analysis to provide health counseling and preventive healthcare knowledge services to the teacher population, effectively expanding the hospital's business scope.

By mining hospital business data, understanding the cost-effectiveness of hospitals, hospitals can grasp their characteristics. Based on this, they can develop advantages in business handling, optimize hospital business structures, and improve hospital economic benefits.

Optimizing Hospital Resource Allocation. Hospital resources mainly include medications, medical equipment, and hospital infrastructure [10]. By utilizing data mining to optimize hospital resource allocation, it is possible to alleviate the pressure on resource utilization, improve resource efficiency, and reduce waste during the treatment process, thus accelerating the efficiency of patient reception and treatment.

Currently, bed resources are the most scarce in hospital resource allocation. Data mining can be used to obtain and analyze patient admission information, such as admission and discharge times, to understand the utilization of hospital beds at different times and derive predictive models for hospital bed usage. Based on the predictive results, coordination among various departments can be fully achieved to efficiently and reasonably allocate hospital bed resources, further increasing the utilization rate of hospital beds. This reduces the time cost for patients waiting, ultimately aiming to improve the efficiency of patient admission and treatment in the hospital.

3.2 Application in Clinical Medical Diagnosis

Predicting Diseases. Data mining techniques have tremendous potential in clinical medical diagnosis. With the abundance of complex data generated in the healthcare industry, including patient information, disease diagnoses, electronic medical records, etc., applying data mining techniques to correctly predict and determine patients' diseases can improve the efficiency of patient recovery and enhance hospital treatment effectiveness [3].

In 2022, researchers Zhu Xiaotong etc. introduced a cardiovascular disease prediction model based on one-dimensional convolutional neural networks[11]. This model preprocesses 13 relevant parameters related to cardiovascular diseases and then performs convolution operations using two-dimensional matrices, followed by model optimization. After experiments using two datasets from the UCI database, the accuracy obtained was 93.36% and 94.48%.

Building upon medical imaging-based cardiovascular disease prediction models, this model utilizes convolutional neural algorithms to classify and recognize echocardiography images, achieving an accuracy rate of 98% in disease identification. The high accuracy of this technology ensures the reliability of the cardiovascular disease prediction model proposed by the researchers.

In addition to electrocardiogram information, this model predicts cardiovascular diseases based on 13 clinical parameters such as blood pressure, blood sugar, and blood lipids. Compared to models based solely on physiological parameters, the model proposed by Zhu Xiaotong etc. achieves higher accuracy. By incorporating more clinical parameters on top of medical imaging-based prediction models, this model is easily accessible and ensures result accuracy.

This predictive model serves as a case study of data mining techniques in clinical medicine, enabling disease prediction in medical diagnosis. Its excellent ability to identify cardiovascular diseases can assist doctors in disease diagnosis and improve patient treatment efficiency.

Assisting Clinical Diagnosis and Treatment. With the global population growth, more people are facing health issues, posing challenges to limited medical resources to meet the rapidly growing diagnostic data. Simultaneously, the substantial increase in diagnostic data provides a basis for medical research, treatment assistance, and disease diagnosis. However, the capacity of healthcare professionals to analyze and process such vast and complex data is limited. Therefore, leveraging data mining techniques to analyze and summarize diagnostic data can provide references for healthcare professionals in patient diagnosis and treatment, thereby accelerating the efficiency of patient treatment.

In 2019, researchers Mao Ye etc. provided a case of optimizing orthopedic rehabilitation information using data mining techniques [10]. The test results indicated that orthopedic rehabilitation information data mining could be used for precise treatment of orthopedic diseases, assisting clinical diagnosis by doctors.

The system evaluates doctors' diagnoses to assist in clinical diagnosis and treatment. By recording doctors' treatment data and applying data mining techniques to analyze diagnostic data, the system provides corresponding reference results. The system uses clustering methods to categorize diseases into three types: high misdiagnosis rate, medium misdiagnosis rate, and low misdiagnosis rate [10], serving as auxiliary factors in evaluating diagnostic results.

The system can continuously learn through manual judgment to ensure more accurate future judgments [10]. Based on a large amount of orthopedic diagnostic information, the system establishes a diagnostic information analysis model to identify patterns. When the system encounters data that is difficult to discern or has a low accuracy rate, it can be diagnosed by experienced experts. This approach enables the system to continuously learn and reduce reliance on manual judgment in the future.

4 Conclusion

In conclusion, the application of data mining technology is beneficial for the development of healthcare businesses. For hospital operations, data mining techniques can analyze the cost-effectiveness and business operations of hospitals and be used to optimize resource allocation.

Cost analysis of hospitals can provide insights into the hospital's financial situation and predict future demands, enabling rational planning for hospital development and providing references for hospital decision-making. Hospitals can also expand their business rationally by analyzing patient characteristics using data mining techniques.

For hospital resource allocation, data mining techniques can analyze past visit data, identify and predict the demand for various resources in different departments at different times, and adjust more reasonable resource allocation plans based on predictions.

In terms of clinical medical diagnosis, data mining technology can help predict diseases and assist in clinical diagnosis and treatment.

Currently, numerous researchers have utilized data mining techniques to design different algorithms based on large clinical data sets, launching predictive models for various diseases. These models, mostly based on classification and clustering techniques, process and analyze data, some achieving high accuracy rates.

By applying high-completeness disease prediction models, doctors can use the model results combined with patient conditions and their own experience to decide treatment plans. Predictive models can also help doctors intervene early in patients at risk of illness, change their treatment strategies based on identified risks, thereby avoiding disease deterioration and improving treatment effectiveness.

Data mining technology can assist doctors in faster and more accurate clinical diagnosis. By analyzing and summarizing large amounts of data obtained clinically, data mining applications can discover potential disease patterns in patients, assisting in determining disease types. Moreover, data mining can be applied to evaluate doctors' diagnoses, helping identify biases or errors in diagnosis. This increases

diagnostic accuracy and provides a basis for quality management and improvement of medical services for healthcare institutions.

However, the application of data mining in the healthcare field also faces challenges. Medical knowledge concepts are complex, and there are numerous details. Researchers of data mining technology usually do not have as thorough a grasp of medical knowledge as experienced medical scholars. This may lead to limitations in the application of data mining technology, as researchers may not fully understand the complexity of medical data and the special requirements of the medical field, affecting their precise application of data mining technology. Therefore, how to master the details of medical knowledge, correctly handle complex medical data, and obtain more accurate and appropriate data mining technology solutions is a question that researchers in this field need to discuss in the future.

Training hybrid talents who simultaneously master medical knowledge and data mining technology can serve as a solution to this issue, while also presenting new challenges to the education industry. Additionally, establishing more research institutions to recruit medical scholars and algorithm researchers to jointly conduct in-depth research on the application of data mining in the healthcare field can promote the integration of medical science and technological innovation.

References

1. Yang, J., Li, Y., Liu, Q., et al.: Brief introduction of medical database and data mining technology in big data era. Evidence-Based Medicine 13(1), 57-69 (2020).
2. Khedr, M. A., Aghbari, A. Z., Ali, A. A., Eljamil, M.: An Efficient Association Rule Mining from Distributed Medical Databases for Predicting Heart Diseases. IEEE Access 9, 15320-15333 (2021).
3. Ghorbani, R., Ghousi, R.: Predictive data mining approaches in medical diagnosis: A review of some diseases prediction. International Journal of Data & Network Science 3(2), 47-70 (2019).
4. Kasperczuk, A., Dardzińska, A.: Comparative Evaluation of the Different Data Mining Techniques Used for the Medical Database. Acta Mechanica et Automatica 10(3), 233-238 (2016).
5. Das, S., Saini, G., Rai, K., et al.: A Review on Healthcare Sector using Data Mining and Classification Techniques. International Journal of Innovative Research in Technology and Management 6(6), 2581-3404 (2022).
6. Wang, J., Zhong, X., Yang, F.: The Application of Data Mining in Medical Clinical Pathway. Modern Hospital 11(3), 1-3 (2011).
7. Xie, B.: The Application of Data Mining and Analysis in Medical Business. Electronic Technology & Software Engineering (15), 181-182 (2021).
8. Saxena, P., Lehri, S.: Analysis of various clustering algorithms of data mining on health informatics. Technology 7(2), 1 (2016).
9. Zhang, C.: Application Research of Data Mining Algorithm in Medical Management. Software 44(6), 157-159 (2023).
10. Ye, M., Zhang, H., Li, L.: Research on Data Mining Application of Orthopedic Rehabilitation Information for Smart Medical. IEEE Access 7, 177137-177147 (2019).

11. Zhu, X., Pang, C., Zhu, H.: Deep learning-based cardiovascular disease prediction model. Computer application 41(S2), 346-350 (2021).