

Intelligent Analysis and Research on Clinical Data of Traditional Chinese Medicine Diagnosis and Treatment of Coronary Heart Disease Based on Data Mining

Kun Zhang^{1,2,3}, Xiaoyan Chen^{1*}, Haifeng Wang^{1,2}, Chong Shen^{2,3}

¹College of Ocean Information Engineering, Hainan Tropical Ocean University, Sanya, Hainan, 572022, China

²State Key Laboratory of Marine Resources Utilization in South China Sea, Hainan University, Haikou, Hainan, 570228, China

³College of Information Science and Technology, Hainan University, Haikou, Hainan, 570228, China

*E-mail: xiaoyan0205@126.com

*The corresponding author

Keywords: Data mining, Intelligent data analysis, Wisdom medical, Chinese medicine treatment rules

Abstract: Coronary heart disease is a serious harm to human health and common disease in many cases, in Hainan, is also one of the common multiple fatal diseases. This paper analyzes the association rules of clinical diagnosis and treatment of coronary heart disease in patients with coronary heart disease, analyzes the research situation of domestic and foreign research, expounds the advantages of using intelligent medical treatment and traditional Chinese medicine diagnosis and treatment, and puts forward the clinical practice of TCM based on data mining Application of Intelligent Analysis in Clinical Diagnosis and Treatment.

1. Introduction

Coronary heart disease is a serious harm to human health and common disease in many cases, in Hainan, is also one of the common multiple fatal diseases. It is of great theoretical and practical significance to study how to quickly and effectively extract the association rules from the clinical diagnosis and treatment data of traditional Chinese medicine (TCM) and to apply the relevant association rules to the clinical diagnosis and treatment of coronary heart disease.

According to the characteristics of Chinese medicine diagnosis and treatment of coronary heart disease and the need of mining, this project will introduce the vector method into the association rules mining, from the pulse, tongue, pathogenesis and other 18 aspects and 26 kinds of coronary heart disease TCM syndromes excavated between the TCM syndrome, Has received a series of rules of coronary heart disease syndrome, for the diagnosis and prevention of coronary heart disease provides an important basis for decision-making.

Coronary heart disease diagnosis and treatment data are often multi-valued attributes, multi-class standard data, the traditional association rules mining algorithm is usually not a good combination of domain knowledge, direct mining of the rules, the efficiency is low. To this end, this project combined with the characteristics of coronary heart disease TCM diagnosis and treatment data, decision-making attributes and non-decision attributes of the block coding, proposed a pre-and posterior constraints of the association rules mining algorithm, can effectively dig out the traditional Chinese medicine treatment of coronary heart disease medication, Greatly improving the efficiency of the algorithm. This algorithm has some theoretical and practical value for finding the data mining problem of association rules between decision attributes and non - decision attributes.

Traditional association rules often use a unified support threshold to mine frequent itemsets, resulting in low support items with low support levels that can not be tapped. If you reduce the support threshold will produce a large number of redundant short project set, and will lead to

greatly reduce the efficiency of the algorithm. To this end, this project proposes a kind of association rule mining method with front and back and length decreasing support constraints, which can effectively excavate more long patterns and reduce useless short patterns, making the generated rules more theoretical and practical value, Can effectively assist in the diagnosis and treatment of coronary heart disease.

2. Domestic and foreign technology development situation

At the 11th International Joint Artificial Intelligence Conference (IJCAI-89), held in August 1989, G.P.ShaPirO first proposed the concept of Knowledge Discovery in Database (KDD). Data mining (DM) is considered a specific step in the KDD process, which uses a specialized algorithm to extract patterns from the data. From 1991 to 1994, the workshop on KDD was held. In 1995, the conference was held for the first time in the International Conference on Data Mining and International Conference on Data Mining. The concept of data mining was put forward. Because data mining is the most critical step in the knowledge discovery process, data mining is often used to represent the entire knowledge discovery process. Over the past 20 years, data mining technology has been attracting a large number of artificial intelligence and machine learning researchers, creating a data mining research boom, become the current computer science industry a hot spot.

There are many universities in the world to study data mining. More well-known universities include Carnegie Mellon University (Center for Machine Building Data Mining, Multimedia Database Data Mining Research Center, Internet Data Mining Research Center), Massachusetts Institute of Technology, Stanford University. The well-known research institutions are ACM (ACM Special Interest Group on Knowledge Discovery in Data and Data Mining), NCDM (The National Center for DataMining (NCDM) at the University of Illinois at Chicago), KNet (the European Knowledge Discovery Network of Excellence), Coronary heart disease TCM clinical diagnosis and treatment data model structure shown in Figure 1.

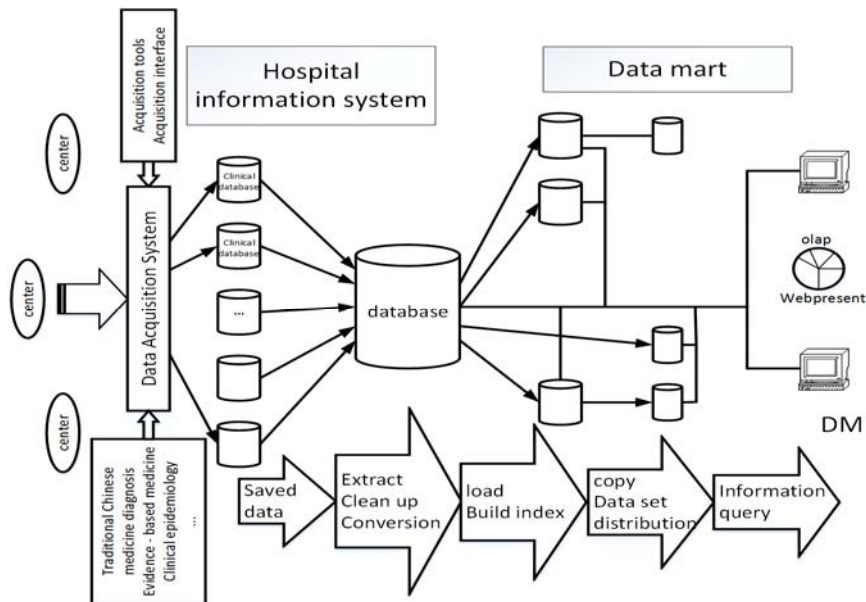


Figure1. Coronary heart disease TCM clinical data model structure

Tsinghua University, Peking University, Shanghai Jiaotong University, Xi'an Jiaotong University, Nanjing University, National University of Defense Science and Technology, Huazhong University of Science and Technology, Fudan University, Zhejiang University, China University of Science and Technology, Chinese Academy of Sciences Institute of Mathematics, Jilin University, Dalian Maritime University, Air Force Third Institute, Institute of Computing Technology, Chinese Academy of Sciences and many other institutions of higher learning and scientific research units competing to carry out knowledge discovery theory and its application research.

3. Application of Data Mining and the Advantages of Traditional Chinese Medicine

3.1. The development of data mining

The history of data mining is based on the development of related disciplines. With the development of database technology and its application, people accumulate more and more data. The surge of data is hidden behind a lot of important information, simple query and statistics have been unable to meet the needs of business, the need for a hidden data behind the hidden knowledge of the means.

At the same time, artificial intelligence has made significant progress since its birth. Experienced the game period, natural language understanding, knowledge engineering and other stages, the recent hot is the machine learning. Using the database management system to store data, using machine learning methods to analyze data, mining a lot of data behind the knowledge, the combination of the two contributed to the database of knowledge discovery (KDD: Knowledge discovery in databases) generation. Knowledge discovery in the database is a cross-cutting subject, involving a number of areas such as machine learning, pattern recognition, statistics, intelligent database, knowledge acquisition, data visualization, high performance computing, expert systems and so on. The knowledge found in the database can be used in information management, process control, scientific research, decision support, and so on. Some people also refer to KDD as data mining, but it is generally considered that data mining is a processing phase of KDD, which is the KDD process In a key step.

3.2. Common algorithms for data mining

3.2.1. Association rules

The association rule is the main model of the current data mining research, focusing on determining the relationship between the different domains in the data and finding the dependencies between multiple domains that satisfy the given support and confidence threshold. Mining association rules refers to the rules in the data warehouse dug with this form of rules: due to the occurrence of some events caused by the occurrence of other events. A typical example of association rule mining is the shopping basket analysis. The association rule study helps to find the link between the different commodities in the transaction database and find out the behavior of the customer buying behavior, such as the purchase of a commodity for the purchase of other goods. The analysis results can be applied to commodity shelf layouts, inventory arrangements, and classification of users based on purchase patterns.

3.2.2. Classification

Classification is one of the main contents of data mining, mainly through the analysis of training data samples, resulting in accurate description of the category. This category is usually composed of classification rules, can be used to predict the future of the data, has a wide range of applications. Classification and prediction are closely related to the two forms of analysis, the difference is that the classification is predicted classification (discrete) and the forecast is to establish a continuous value model.

Common algorithms for data classification include decision tree, Bayesian classification and Bayesian network, neural network and so on. Most of the algorithms are memory-resident algorithms, and it is often assumed that the amount of data is small. Recent database mining studies, based on these tasks, propose scalable classification and prediction techniques that can handle large amounts of data on resident disks.

3.2.3. Clustering

Unlike classification analysis, cluster analysis inputs a set of unclassified records, and these records should be divided into categories that are not known in advance. Clustering analysis is through the analysis of the database records, according to a certain classification rules, a reasonable

division of the record set to determine the type of each record. The classification rules it uses are determined by the clustering analysis tool.

Clustering is to divide the data points into several clusters or clusters so that the data points in each cluster are maximally similar, and the data points in different clusters are maximally different; thus, the data set is valid, Novel, useful data model that can be used. Clustering methods are common division, hierarchy, density, model and grid.

3.3. Application direction

Coronary heart disease TCM diagnosis and treatment data are often multi-valued attributes, multi-class standard data, the traditional association rules mining algorithm is often not a good combination of domain knowledge, direct mining of the rules, the efficiency is low. To this end, this project combined with the characteristics of coronary heart disease TCM diagnosis and treatment data, decision-making attributes and non-decision attributes of the block coding, proposed a pre-and posterior constraints of the association rules mining algorithm, can effectively dig Chinese medicine treatment of coronary heart disease medication.

Traditional association rule mining usually adopts unified support threshold, so it is difficult to effectively exploit long-mode rules with low support. To this end, this project combined with the characteristics of coronary heart disease TCM diagnosis and treatment data, proposed a back and forth and length reduction support degree of association rules mining method, can reduce the useless short mode rules, can dig out more effective adjuvant coronary heart disease Of the long pattern rules.

Aiming at the shortcomings of the association rules in the global correlation, this project proposes a method of segmented nonlinear regression and reverse verification, and verifies the relevance of the association rules. This method makes the analysis of the relevance of the association rules more accurate, reducing the number of rules, can dig out more meaningful rules. Experiments on TCM diagnosis and treatment data show that this method is more practical.

4. Summary

Based on data mining and intelligent analysis of clinical practice of traditional Chinese medicine diagnosis and treatment of coronary heart disease in Hainan Province, the development of clinical techniques for the diagnosis and treatment of coronary heart disease has a positive effect. The research of this project is strong and practical, and it is applied to medical field. Innovative development and promotion, in line with modern, science and technology of medical development trends.

5. Acknowledgments

This research was financially supported by the Key Project of Scientific Research of Hainan Province(No.Hnky2015ZD-14); the Scientific and Technological Cooperative Project for College and Region of Sanya(No.2015YD16).

References

- [1] Feng Yuan, Shouqiang Chen. Model Construction on Efficient mMining Association Rules in Clinical Data of Hypertension, *Computer Engineering and Applications*, 2011, Vol. 47, No. 36: 226-229.
- [2] Gang Yan. Application of Association Rule Algorithm in TCM Data Mining, *Silicon Valley*, 2009, No. 24: 65.
- [3] Jue Wang. Realization of Diagnosis and Treatment Data Dmining System Based on Asocciation Rules, *Modern Electronics Technique*, 2013, Vol. 36, No. 19: 124-126.

- [4] Guang Liu, Peng Liu. Prospects of Data Mining Technology in TCM Diagnosis and Treatment, *Medical Information*, 2011, Vol. 24, No. 09: 5616-5617.
- [5] Li Jin, ZongDian Wang, Hongxing Kan, Ru Han. Research on the Application of Data Mining in the Extraction of TCM Diagnosis and Treatment Rules, *Lishizhen Medicine and Materia Medica Research*, 2013, Vol. 24, No. 04: 1015-1017.
- [6] Guang Liu, Lei Wu. Study on the Mining of Medication and Prescription of Traditional Chinese Medicine Based on Association Rules, *Jiangsu Journal of Traditional Chinese Medicine*, 2011, Vol. 43, No.11: 72-74.
- [7] Fengxia Ou, Zongdian Wang. Application of Association Rules in Data Mining Technology to TCM Diagnosis, *Journal of Henan Institute of Engineering(Natural Science Edition)*, 2011, Vol. 23, No.02: 53-58.
- [8] Zhi Liu. Research of Association Rule Mining Algorithms and Their Applications of Coronary Heart Disease Diagnosis and Treatment with Traditional Chinese Medicine, Dalian: Dalian Maritime University, Master Dissertation, 2012.
- [9] Lisha Shi. Data Warehousing and Data Mining on the Clinical Database of Coronary Treatment with Chinese Traditional Medicine, Dalian: Dalian Maritime University, Master Dissertation, 2008.
- [10] Kun Zhang, HongXu Wang, HaiFeng Wang, Zhuang Li. New Exploration on Definition of Similarity Measures Between Vague Sets, *Journal of Natural Science of Heilongjiang University*, 2012, Vol.29, No.03, pp.412-415.
- [11] Kun Zhang, HongXu Wang, HaiFeng Wang, Zhuang Li. Vague Spatial Decision Method and Its an Application for the Location of Tailings Dam, *Computer Science*, 2014, Vol. 41, No. 04: 260-262.
- [12] Kun Zhang, Hong-Xu Wang, Hai-Feng Wang and Zhuang Li. The Time Series Prediction Algorithm of Enrollment based on the Hesitancy degree Correction Rule of Vague Sets, *ICIC Express Letters*, 2015, Vol. 9, No. 5:1311-1318.
- [13] Kun Zhang, Hongxu Wang, Haifeng Wang, Zhuang Li. Fuzzy Time Series Prediction Model and Application based on Fuzzy Inverse, *International Journal of Signal Processing, Image Processing and Pattern Recognition*, 2015, Vol. 8, No. 10:121-128.