

A Study on Knowledge Discovery based on Big Data Background of Visceral Manifestation Theory in *Huangdi's Internal Classic of Medicine*

Yanda She*

School of Computer

Tonghua Normal University

Tonghua, China

E-mail: *dbcy9999@163.com

Abstract—This paper, under the background of big data, combines both the theory of "Huangdi undefined Internal Classic" and the pharmaceutical annotation theory of *Huangdi's Internal Classic of Medicine* with database, devises applied program, and probes into knowledge discovery, thus obtaining the theme distribution of the theory of liver storage image, and providing a methodological reference for the study of traditional Chinese medicine classical theory.

Keywords—*Big Data; Database; Visceral Manifestation Theory; Knowledge discovery; Huangdi's Internal Classic of Medicine*

I. INTRODUCTION

The theory of Visceral manifestation theory is an important part of the theory system of *Huangdi's Internal Classic of Medicine* and the theoretical basis of clinical dialectical treatment. The contents of Visceral manifestation theory theory are very large, which makes it very difficult to systematically understand the contents of Visceral manifestation theory theory in *Huangdi's Internal Classic of Medicine*. Since the publication of the *Huangdi's Internal Classic of Medicine*, the physicians of the past dynasties have carried out a great deal of sorting out and annotating it, forming a number of representative notes, therefore, refer to the annotated book, choose a good one, and follow through on it." It is an essential and important way to study the Internal Classic.^[1] "however, there are more than a dozen annotators from the Sui and Tang dynasties to the Ming and Qing dynasties. These annotations are sometimes consistent, sometimes they differ greatly, and the amount of data is huge, which makes it more difficult for the later scholars to learn.

In todayundefineds big data era, the data get the world. Big data has never been lacking in the history of Chinese medicine, and a typical feature of Chinese medicine is the large amount of data. In the study of Visceral manifestation theory theory of traditional Chinese medicine, how to systematically arrange and study the Visceral manifestation theory theory of *Huangdi's Internal Classic of Medicine* under the background of big data and referring to the notes from various countries is the problem discussed in this paper. On the basis of systematically collecting and sorting out the

Visceral manifestation theory theory of *Huangdi's Internal Classic of Medicine*, and referring to 14 annotated books of *Huangdi's Internal Classic of Medicine*, this paper constructs big data analysis system of *Huangdi's Internal Classic of Medicine*, and introduces the method of knowledge discovery (KDD). In order to find the potential and unknown knowledge, this paper studies the contents and annotations of the liver and Visceral manifestation theorys in *Huangdi's Internal Classic of Medicine*, and provides a methodological reference for the theoretical study of the Visceral manifestation theorys in the *Huangdi's Internal Classic of Medicine*.

II. THE ESTABLISHMENT OF DATABASE SYSTEM UNDER THE BACKGROUND OF BIG DATA

A. The design database structure

According to the Tibetan theory of Huangdi Classic and annotatorundefineds annotation, the relational model database is used in this database, that is, the two dimensional table data is used to express the relationship between the data. This model can not only reflect the one-to-one relation between attributes, but also reflect the one-to-many relation between attributes and the many-to-many relationship between attributes [3]. The conceptual structure of the database is designed in the form of a conceptual structure diagram as follows:

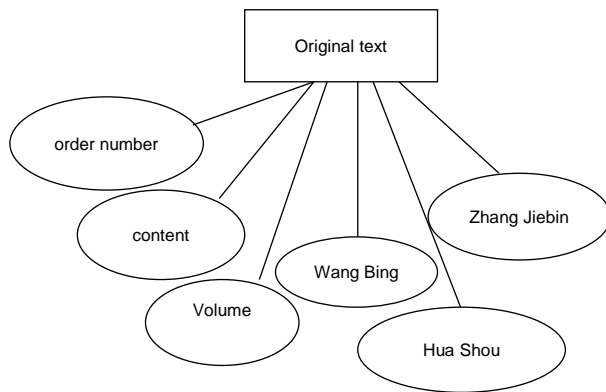


Fig. 1 Conceptual structure chart of the database of the visceral manifestation

In figure 1, the original liver visceral manifestation is an entity with corresponding attributes with serial number, content, and notes of Wang Bing, Zhang Jiabin and other annotators. According to the concept structure diagram, we can get the logical structure of the special topic database of "the liver visceral manifestation annotator", and on this basis, select the appropriate database management system, we can establish the database table.

According to the above concept structure chart, we can get the logical structure of the annotator database of the Visceral manifestation theory's Internal Classic of Medicine. According to the conceptual structure and logical structure, using SQL-Server as the database management system, each database table in the 14 annotated database of liver hide image theory can be established.

B. Database development tools

Under the background of big data, to study the knowledge discovery of the traditional Chinese medicine, we should first choose the suitable database management tool. SQL Server. SQL Server is a relational database management system developed by Microsoft, it can provide secure and reliable processing, storage, retrieval, analysis and other functions for relational and structured data.

Using Microsoft Visual C# is an object oriented programming language developed by Microsoft. It is a new program language tailored by Microsoft for .Net platform [4].

The SQL Servers of this database system is used to manage the database, and Microsoft Visual C # is used to develop the Windows window application program which is easy to operate, which makes the complex database transparent and ensures the integrity of the data. Usability and recoverability, and easy to operate, the database can be manipulated by the application program, and scientific inquiry and statistical analysis can be carried out on the theory of liver storage and annotated notes of the Inner Classic. To provide reliable technical support for further discussion of relevant theories.

C. The principles and methods of Entry

"liver" is used as the key word, and besides "liver" as the key word, "soul", "anger", "tendons", "claw", "eye", "Jueyin", "wood", "acid", "application and harmony" are used. "A and B", "Qing", "East", "Wind" and so on related to liver hiding image related to the search for articles related to liver hide image. After literature analysis, the theoretical data table of liver storage image and the data table of liver storage image theory were inputted into "element question" and "Ling Pivot" respectively.

III. THE ESTABLISHMENT OF DATABASE SYSTEM UNDER THE BACKGROUND OF BIG DATA

A. The brief introduction to knowledge discovery method

As an interdisciplinary subject, knowledge discovery has attracted much attention with the development of computer technology and information science and technology. The study of knowledge discovery begins with the concept of finding useful schemas from databases, that is, KDD (KnowledgeDiscovery in Database). KDD is the recognition of valid and novel ones that exist in databases. A nontrivial process with potentially useful and ultimately understandable patterns. [5] data mining (DM) is generally considered to be a step in the process of knowledge discovery (KDD), or to equate knowledge discovery with data mining.

B. The application of knowledge discovery

The task of knowledge discovery is to find and extract hidden, unknown, potentially useful and understandable rules and patterns from a large number of data. Therefore, knowledge discovery is widely used to solve data classification and data clustering. Decline and prediction, correlation and correlation, sequence discovery, description and discrimination, time series analysis and so on [6]. Knowledge discovery is also widely used in the field of traditional Chinese medicine. If literature [7] collects relevant information such as formula with prolactin effect and traditional Chinese medicine and so on, the data will be standardized and the database will be set up. The frequency analysis and association rule method were used to excavate the formula with prolactin effect and the compatibility of traditional Chinese medicine, which provided a reference for clinical treatment of postpartum lack of milk. In reference [8], data frequency statistics and co-occurrence matrix were generated by using SATI, a tool for statistical analysis of bibliographic information, and a visualized co-occurrence network was studied by using social network analysis software (ucinet & underdraw). This paper draws and analyzes the development law of the research on the five elements music of TCM, the main author groups and the hot spots, which lays a foundation for the theoretical research and practical application of the music of the five elements of TCM. Literature [9] based on the database of pulmonary diseases of traditional Chinese medicine, discusses the use of various data mining methods to study the etiology, pathogenesis, syndromes and characteristics of lung diseases, and the application of prescription drugs. It was found that the pathogeny and pathogenesis, syndromes, syndromes distribution, disease-prescription, symptom-prescription, syndrome-prescription and compatibility of lung diseases could be preliminarily extracted by different mining

methods. Literature [10] collected the medical records of the first national medical master and the first batch of national famous Chinese medicine practitioners, and made data mining.

Through the retrieval of the literature on the application of knowledge discovery in the field of traditional Chinese medicine under the background of big data, it is found that the application of knowledge discovery is mainly focused on the study of the law of compatibility of prescriptions and medicines, and the application of the theory of Tibetan image in the Classic of Internal Medicine has not been reported in the literature at present. Therefore, this paper applies the method of knowledge discovery to 14 annotated databases of liver hidden image theory, in order to find some valuable knowledge.

C. The distribution of the liver visceral manifestation theory in Huangdi's Internal Classic of Medicine

According to the frequency of cooccurrence of words and phrases, the theme distribution of liver hide image theory was analyzed. The total choice was: liver soul, liver- blood, liver-fire, liver-blood, liver-gallbladder, liver and qi, liver-fire. Liver-oriental and other 17 pairs of word pairs related to liver hide images were recorded into the knowledge discovery module for word frequency statistics. The results are shown in figure 2:

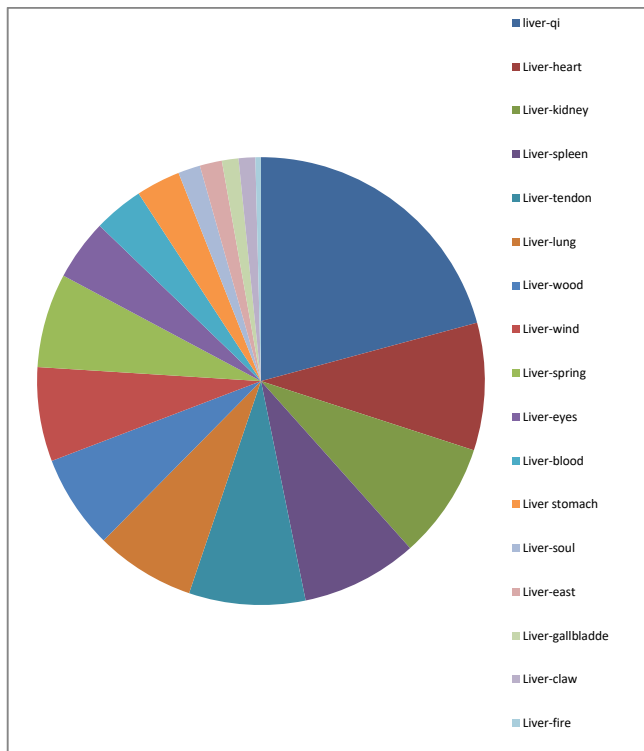


Fig. 2 Statistical results of word Frequency of liver visceral manifestation

From the above analysis, we can see that in the 17 pairs of words, the highest frequency is "liver-qi". For example, in the database: "liver gas failure, general atmosphere in the liver, liver qi changes, in the liver, the liver Qi is a dream anger, liver Qi in the heart, liver Qi heat," There are as many as 52 items, such as the origin of liver-qi, the aggravation of liver-qi and the "qi of the fascia of the liver". It can be seen that in the time of the "Internal Classic", when discussing the physiology and

pathology of the liver, there are more discussions from the "Qi" level. This has something to do with the philosophical thinking of the time. The Nei Jing holds that people and all things come from nature and are born in heaven and earth, so the basic material that constitutes the human body is also equal to all things [11]. But the fire here, which is not a modern description of pathological phenomenon, so called "liver-fire", can be seen that the theory of liver-fire played a role in the process of describing the pathological manifestations of liver.

IV. CONCLUSIONS AND DISCUSSIONS

The method of knowledge discovery belongs to modern information science, and its mode of thinking is mainly conceptual thinking and logical thinking from the West. In the context of big data, for the applicability of knowledge discovery methods in the field of traditional Chinese medicine, the general idea is to apply specific algorithms (such as association analysis, cluster analysis, etc.) to find texts that appear frequently and have a high degree of association from the database.

REFERENCES

- [1] Xiong Jibai. Refer to the Note "Inner Classic" [N]. Chinese Medicine Journal, 2014-04-21 (004): 1.
- [2] Liu Denyi. Big data should bring great value to the development of traditional Chinese medicine [N]. Health Journal, 2017-03-29 (005).
- [3] introduction to Database system [M]. Wang Shan, sa Shixuan. Beijing: higher Education Press, 2014.
- [4] Sun Sancui, Xu Yanyin. Proficient in C # and ASP.net programming. Beijing: China Railway Press, 2003.
- [5] Ni Zhiwei, Cai Qingsheng. Neural network expert system based on knowledge discovery technology. Journal of Xiamen University (Natural Science Edition), 2003: 288-292
- [6] Sun Jihong, Jiao Yuying. Research on knowledge Discovery and its Development trend. Information theory and practice. 2006 (5): 528-530
- [7] Kang Chen, Song Xingxing, Tian Huiling, et al. Mining the law of prolactin prescription and compatibility of traditional Chinese medicine based on knowledge discovery. Chinese Journal of Chinese Medicine Library and Information, 2015, 39 (1): 12-15
- [8] Hu Songjie, Ma Yanping. Research on Visualization of traditional Chinese Medicine five elements Music Therapy based on knowledge Map. Journal of Changchun University of traditional Chinese Medicine, 2015N 31 (3): 534-537
- [9] Chen Liping, Li Jian-sheng, Cai Yongmin, et al. Train of thought and method of constructing and excavating TCM pulmonary disease database. Chinese Journal of traditional Chinese Medicine, 2017 (12): 5530-5534
- [10] Liu Jiahui, Wei Zhihui, Lu Dongyong, et al. Study on the Drug use of Old Chinese Medicine for Primary Hepatocellular carcinoma based on data Mining. Chinese Journal of traditional Chinese Medicine, 2016N 31 (1): 58-61
- [11] Zhao Mingshan, Ju Baozhao. An interpretation of the Culture of Huangdiundefineds Internal Classic [M]. Shenyang: Liaoning Science and Technology Press, 2014: 87.