

August 2023

Data Analysis of Traditional Chinese Medicine Disease Diagnosis from the Perspective of Computational Sociology.

Zhou Haodong

Biquan Academy, Xiangtan University, Xiangtan 411105, China, 512837914@qq.com

Follow this and additional works at: <https://www.interscience.in/ijcct>



Part of the [Digital Communications and Networking Commons](#)

Recommended Citation

Haodong, Zhou (2023) "Data Analysis of Traditional Chinese Medicine Disease Diagnosis from the Perspective of Computational Sociology.," *International Journal of Computer and Communication Technology*: Vol. 9: Iss. 2, Article 2.

DOI: 10.47893/IJCCT.2023.1451

Available at: <https://www.interscience.in/ijcct/vol9/iss2/2>

This Article is brought to you for free and open access by the Interscience Journals at Interscience Research Network. It has been accepted for inclusion in International Journal of Computer and Communication Technology by an authorized editor of Interscience Research Network. For more information, please contact sritampatnaik@gmail.com.

Data Analysis of Traditional Chinese Medicine Disease Diagnosis from the Perspective of Computational Sociology

Zhou Haodong

Biquan Academy, Xiangtan University, Xiangtan 411105, China

Correspondence: 512837914@qq.com

Abstract

In Traditional Chinese Medicine (TCM), the diagnosis and treatment of diseases typically involve viewing the patient as a system and considering both the intrinsic natural mechanisms of the disease and the external sociological factors. However, a comprehensive and scientific standard for understanding the external sociological factors in TCM diagnosis and treatment has not yet been established. The main reason for this is the complexity of computing these sociological factors due to their openness, multidimensionality, and heterogeneity. Drawing insights from computational sociology, this study explores the latent sociological factors in TCM disease diagnosis and treatment. It aims to obtain sociological factor data related to diseases from various online sources, such as internet-based medical consultation platforms and social networks. Through data analysis, it seeks to reveal the correlations between diseases and sociological latent factors. The ultimate goal is to establish a pre-diagnosis sociological factor database for TCM diseases. This endeavor serves as a foundation for developing a more scientific online TCM disease consultation system, providing references for TCM disease diagnosis and treatment, and offering evidence-based health behavioral

recommendations for disease prevention.

Keywords: Computational Sociology; Data Analysis; Big Data Technology; Disease Diagnosis

Introduction

Social sociology was coined by Western scholar Auguste Comte, and the first book on sociology was written by the British philosopher Herbert Spencer. Almost all the foundational theories of sociology were established by Western scholars¹. The existing social sociology theories originate from the West, and achieving a breakthrough and localizing sociology, a field that has been leading in the West for hundreds of years, is a current concern for Chinese sociologists². Sociology is a discipline based on human self-reflection, and people's understanding of society varies based on different cultural backgrounds. The Western development of sociology for hundreds of years has produced some theories that may not be applicable to China's national conditions and our traditional ideological system³. Sociology itself is used to analyze and solve social problems, and different countries' social problems may manifest in different ways under their respective systems. The development of Traditional Chinese medicine (TCM) and sociology in China share many similarities. TCM has a

long history, but it lacks a data-driven system, and part of TCM data relies on individual doctors' experiences without being recorded and stored, leading to the incomplete transmission of many valuable TCM principles. Presently, the diagnostic methods of many TCM hospitals even rely on Western medicine and Western standards⁴. Western medical standards are based on the physical characteristics, habits, and diets of Western people, which are significantly different from ours. Many indicators used for diagnosing diseases in Chinese patients may not be suitable. Moreover, TCM's diagnostic ideology differs from Western medicine. Western medicine focuses on the symptoms and treatment of diseases, while TCM considers the patient as a system, taking into account the intrinsic natural mechanisms of the disease and external sociological factors. Similar to the localization development of sociology in China, TCM also needs to establish diagnostic and treatment standards that suit its own characteristics⁵.

Pre-diagnosis Data of TCM Diseases from the Perspective of Computational Sociology

The emergence of the concept of computational sociology in 2007 and the proposition of "new computational sociology" in 2014 have brought new perspectives to the development of sociology. Traditional sociological research involves developing basic theories, identifying problems, proposing hypotheses, and then confirming these hypotheses. Researchers need to have a deep understanding of sociological theories before conducting their studies. However, with the advent of computational sociology, researchers can

break free from the constraints of Western foundational theories and directly study sociological phenomena and problems. Computational sociology is an empirical science based on sociology that utilizes large-scale data analysis to better explore latent factors influencing sociological phenomena. It can also aid in the establishment of sociological theories that suit the national conditions⁶. Since some traditional diagnostic and treatment methods of TCM have been lost, the establishment of TCM diagnostic and treatment standards requires a reverse process. The development of big data and the rise of computational sociology offer new ideas for establishing TCM standards. During the TCM diagnostic process, a pre-diagnosis sociological factor database for TCM diseases can be established. By analyzing the data in the database, the connections among different components can be explored, aligning with the fundamental ideas of TCM diagnosis and treatment⁷.

In the field of computational science, problems can be categorized into two types: numerical computation problems and non-numerical computation problems. For numerical computation problems, the mathematical solutions for practical problems are obtained by converting the collected information into data for computation. The input is limited and has a clear deduction process and computation result. Society itself is a complex and diverse system. In the study of sociological problems, due to the numerous latent factors influencing the development of things, a sociological model that includes as many influencing factors as possible needs to be constructed. In the research process, researchers need to continuously discover latent factors

influencing the development of things to make sociological studies and predictions more accurate. Therefore, sociological research is more of a non-numerical computation problem. For non-numerical computation research, non-numerical problems need to be quantified as much as possible. Computational sociology is precisely a computational science that continuously explores latent factors. By analyzing existing data, more correlations between data can be discovered, expanding the database through machine learning, and enabling deeper data mining. The study of big data can help us find patterns in the development of things and make more accurate predictions. Under the general rules of the development of things, there are also differences between different individuals, and traditional research methods have advantages in studying specific individuals. For example, in poverty alleviation, the country needs a clear understanding of all impoverished groups and uses big data to develop national strategies for poverty alleviation. In the implementation stage, different poverty-stricken groups require different poverty alleviation methods. Traditional methods and the study of big data in computational sociology are not contradictory; they complement each other. Traditional research methods improve big data databases in one dimension, and the research results of big data in one dimension develop traditional research methods in sociology. All things can coexist without harming each other, and the path can go side by side without contradicting each other. For computational sociology, computational data can be big data or traditional "small data."⁸ As Professor Chen Yunsong of Nanjing University said, the integration of

disciplines requires discarding inherent concepts from different disciplines. The integration of computational sociology and health big data can help doctors better diagnose patients' diseases. The process of TCM diagnosis usually involves asking patients where they feel uncomfortable, their recent conditions, and then making preliminary judgments based on the doctor's experience. Combining physical examination indicators and health problems with social big data can reveal many latent factors that affect health, some of which may not be medical indicators but sociological indicators closely related to people's work and life. Through the analysis of these latent factors, more non-traditional, relevant latent factors leading to diseases can be found. The establishment of a social health big data system can not only help medical workers diagnose more accurately but also save diagnosis time and establish a health big data system that belongs to us. At the same time, after the digitalization of health big data, cloud diagnosis platforms have more uses and can gradually alleviate the problem of unbalanced regional medical allocation. It can also provide assistance in establishing TCM standards⁹.

From the perspective of computational sociology, the social attributes of pre-diagnosis data in TCM can be analyzed to discover the social correlations within the pre-diagnosis data. TCM disease data can be divided into two parts. The first part consists of the questions that patients are asked by doctors when they visit the hospital. The other part includes the questions patients ask about their potential illnesses on online platforms. The goal is to collect the questions raised on these online platforms. These individuals who raise questions on

the internet have different social attributes, such as gender, age, region, education level, occupation, etc. Analyzing this data can reveal the differences in questions raised by individuals with different social attributes. Simultaneously, training these data samples can discover more correlations between TCM and sociology. In the digital social system, more connections between diseases and sociological factors can be discovered and data-driven to form a pre-diagnosis sociological factor database for TCM diseases¹⁰, providing a data foundation for researching, analyzing, and revealing the sociological factors in TCM diagnosis and treatment.

Pre-diagnosis data of TCM diseases refer to sociological data before diseases have fully formed and before formal medical consultations. These data are sourced from online platforms such as

internet-based medical consultation platforms and social networks.

In TCM, the diagnosis and treatment of diseases usually consider the patient as a system, taking into account both the intrinsic natural mechanisms of the disease and external sociological factors. Sociological factors in the pre-diagnosis of TCM diseases commonly include attributes such as gender, age, diet habits, and bad habits, which are directly related to the disease. There are also implicit sociological factors that are not directly related to the disease, such as economic conditions and social relationships. This article summarizes the sociological factors considered in various TCM disease diagnoses and classifies them. It designs the main sociological factors in pre-diagnosis TCM diseases, as shown in Table 1.

Primary Classification of Sociological Factors	Secondary Classification of Sociological Factors	Quantitative Calculation of Sociological Factors
Basic Information	Sex	M/F
	Age	Under 18 years old/19-30 years old/ 31-45 years old/46-55 years old/ 56-65 years old/66 years old and above
	Residential area	Eastern region/Western region/ Central region/Southern region/ Northern region
	Educational level	High school and below/ Associate's degree/Bachelor's degree/ Master's degree/Doctoral degree
Work situation	Job nature	Private enterprise/ State-owned enterprise/ Public institution/ Government department/Other
	Annual income	Below 50,000/50,000 to 100,000/ 100,000 to 200,000/Above 200,000
Primary social	Marital status	Married/Unmarried

relationships	Parents	Alive/Not alive
	Children	Yes/No
	Siblings	Yes/No
Living habits	Sleep	Sufficient/Average/Less
	Diet	Mild/Spicy/Moderate
	Leisure interests	None/Reading/Traveling/Shopping/Other
	Outdoor sports	Less/Normal/More
	Unhealthy habits	Yes/No/Occasionally
Financial condition	Liabilities	No
		Yes:below 500,000/ 500,000 to 1 million/above 1 million
	deposit	No
		Yes:below 100,000/ 100000 to 500,000 /above 500000
Housing situation	self-owned property	Per capita: below 30m ² / 30-50m ² /Above 50m ²
	no property	renting a place

Table 1. Sociological Factors in Pre-diagnosis of Traditional Chinese Medicine Diseases

By obtaining, processing, and analyzing the data from various internet platforms related to TCM disease consultations and social behavior data recorded in mobile social networks, the correlations between diseases and sociological latent factors are revealed. This process leads to the establishment of a pre-diagnosis sociological factor database for TCM diseases, laying the foundation for the development of a more scientific online TCM disease consultation system, providing references for TCM disease diagnosis and treatment, and offering evidence-based health behavioral recommendations for disease prevention.

Data Analysis of Pre-diagnosis TCM Diseases Based on Computational Sociology

Internet-based TCM disease consultation platforms, mobile social networks (such as blogs, WeChat mini-programs, TikTok, etc.), and other

online platforms contain a massive amount of pre-diagnosis sociological factor data for various TCM disease treatments, which have significant reference value. However, these data possess characteristics of openness, multidimensionality, and heterogeneity, resulting in computational complexity. Traditional sociological research methods have difficulty obtaining such valuable vast amounts of data. With the emergence of big data technology and the development of new computational sociology theories and methods, guidance in the analysis and computation of pre-diagnosis sociological factor data for TCM diseases has been provided.

Based on the fundamental theories and methods of new computational sociology and big data analysis technology, combined with the pre-diagnosis sociological factors designed and analyzed in this paper, the data analysis of pre-diagnosis TCM diseases is realized.

The processing steps mainly include data acquisition, data cleaning, data storage, data analysis, and visualization. Python and Java provide mature methods and techniques for obtaining, processing, and analyzing big data. Python, in particular, is an extremely efficient scripting language

for data analysis. Based on Python, the architecture of data analysis for pre-diagnosis TCM diseases is designed, as shown in Figure 1.

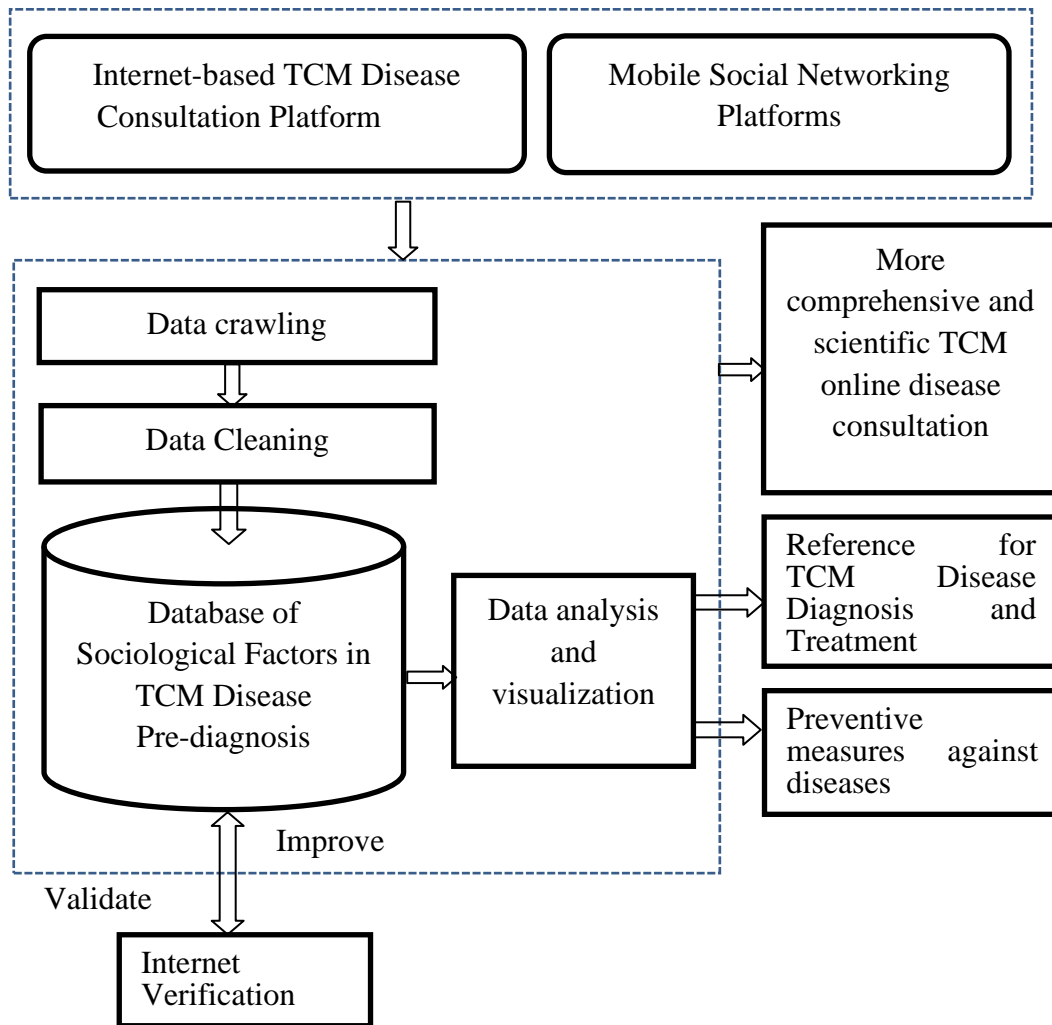


Figure 1. System Architecture of Data Analysis for TCM Disease Pre-diagnosis

Data crawling

Python's web crawling technology can be used to obtain pre-diagnosis sociological factor data for TCM diseases from internet-based TCM disease consultation platforms, mobile social networks (such as blogs, WeChat mini-programs, TikTok, etc.), and other online platforms. Python provides web

crawling methods and techniques such as the scrapy library, BeautifulSoup library, and Requests library. In Python's data acquisition process, the URL manager effectively manages and schedules the URLs of resources to be crawled and those already crawled through URL lists. The web page downloader downloads the web page information determined by the URLs and converts it into text data for local

storage. The web page parser extracts new URLs associated with the web page and stores them in the URL manager. Simultaneously, it extracts valuable data from the downloaded and locally stored text data. In the crawling process of

Data Cleaning and Storage

In the obtained data through web crawling, there might be duplicate, erroneous, or incomplete data. Data cleaning is performed to remove duplicate and erroneous data and to complete the missing data. Through data cleaning and organizing, the subsequent structured storage and utilization of data are optimized, improving data analysis and computational performance. Python provides technologies such as numpy and pandas for data cleaning.

The data that has been crawled, cleaned, and stored in text format is then transferred to a database using structured storage. This process establishes the pre-diagnosis sociological factor database for TCM diseases, making it more convenient for subsequent data analysis and visualization applications. The database is primarily established using MySQL, a relational database that supports data computation and applications based on Structured Query Language (SQL).

Data Analysis and Visualization

As a commonly used data analysis tool in Python, Pandas can directly perform calculations on data in a relational database. It has strong capabilities for two-dimensional or three-dimensional array computations and functions for data restructuring, slicing, aggregation, and subset extraction, making it suitable for data statistical analysis. Python's Matplotlib tool can achieve 2D

pre-diagnosis TCM disease data, the sociological factor data will be crawled according to the classification of TCM diseases and based on the predetermined themes.

visualization of data analysis results.

By analyzing and visualizing the crawled pre-diagnosis TCM disease data, the sociological latent factors in TCM disease diagnosis and treatment are revealed. The quantitative analysis of the correlations between diseases and various sociological factors is conducted. This provides valuable references for TCM disease diagnosis and treatment, effective health management recommendations for disease prevention, and foundational support for the construction of a more scientific online TCM disease consultation system. Utilizing Python's ability in big data analysis, the analysis of pre-diagnosis sociological factor data for TCM diseases can be conducted from multiple dimensions and perspectives. Python's two-dimensional and three-dimensional computational capabilities allow the construction of multidimensional correlation models between diseases and sociological factors, enabling multi-dimensional relationship calculations. For example, calculating the degree of correlation between pre-diagnosis sociological latent factors in TCM disease treatment for different time periods, regions, and economic conditions.

Internet Verification and Correction of Pre-diagnosis TCM Disease Data

Internet social experimentation and research are important components of the new computational sociology research method [10]. Applying internet-based methods to verify pre-diagnosis

sociological factor data for TCM diseases can involve online surveys, building internet-based verification experimental systems, etc. Various forms of verification experiments are designed based on the verification objectives. Through internet experiments, the correctness and completeness of the pre-diagnosis sociological factor database for TCM diseases are validated. The analysis of verification experimental data further contributes to correcting, expanding, and improving the pre-diagnosis sociological factor database for TCM diseases. The verification process has a cyclic and iterative nature.

Conclusion

In TCM disease diagnosis and treatment, consideration of both the intrinsic natural mechanisms of diseases and numerous external sociological explicit and implicit factors is essential. The study of external sociological factors in TCM disease diagnosis and treatment is a sociological research topic, requiring guidance from sociological theories and adherence to sociological research methods and paradigms. External factors in TCM disease diagnosis and treatment represent a non-numerical computational problem that cannot be deduced and verified through "small data" in a closed system. Instead, it requires big data methods and techniques for modeling, analysis, and validation. Therefore, the analysis of pre-diagnosis TCM disease data is a branch of research in the field of new computational sociology. By starting from the perspective of new computational sociology to study sociological factors in TCM diagnosis and treatment, this research provides valuable references for TCM disease diagnosis and treatment and

has significant guiding implications for disease prevention. In this paper, we have considered and preliminarily discussed the methods of analyzing pre-diagnosis sociological factor data for TCM diseases from the perspective of new computational sociology. However, this consideration and discussion do not delve into mathematical modeling and specific implementations, which would be one of the issues that require further exploration in future research.

References

1. He, Yijin. The Shift of the Geographical Center of Sociology: A Perspective from the History of the Discipline. *Chinese Social Sciences Review*, 2018(2), 4-18.
2. Hu, Yipeng. The Inner Path of Wu Wenzao's Thought on the Localization of Sociology. *Sociological Review*, 2021(2), 31-48.
3. Jing, Tiankui; Wang, Junbai. Cultural Awareness and Research on Chinese Sociology. *Journal of Jiangnan University (Humanities & Social Sciences)*, 2016, 15(1), 5-13.
4. Yi, Jianliang. Reflections on Chinese Culture and the Development of Traditional Chinese Medicine. *Journal of Nanjing University of Traditional Chinese Medicine (Social Sciences Edition)*, 2006.09, 7(3), 138-140.
5. Bei, Runpu. Defending Traditional Chinese Medicine and Reflecting on Its Internationalization by Jiang Chunhua. *Chinese Medicine Culture*, 2009(3), 17-20.
6. Chen, Yunsong. Development of Chinese Computational Sociology: Characteristics, Advantages, and Prospects. *Journal of Hunan Normal U*

- niversity (Social Sciences), 2020(5), 1-10.
7. Chen, Yao; Lan, Xu; Zhao, Junnan; Xu, Fengqin. Discussion on the Application of Artificial Intelligence in the Academic Inheritance of Famous Traditional Chinese Medicine Experts. *World Science and Technology-Modernization of Traditional Chinese Medicine*, 2021, 23(1), 165-169.
 8. Fan, Xiaoguang. Paradigm Revolution of Computational Sociology Explores New Space. *China Social Sciences Newspaper*, 2021-08-21(003).
 9. Chen, Yunsong. Understanding the Value of Algorithms to Assist Social Governance. *People's Daily*, 2021(16), 36-37.
 10. Luo, Wei; Luo, Jiaojiang. New Computational Sociology: Sociology Research in the Big Data Era. *Sociological Research*, 2015(3), 222-241+246.
 11. Wang, Haoru; Ning, Jianwen; Ding, Chenyan. Research Progress on the Correlation between Irritable Bowel Syndrome and Psychosocial Factors. *Proceedings of the 2014 Zhejiang Medical Association Academic Conference on Psychosomatic Medicine and Forensic Psychiatry*, 2014, 83.