RESEARCH ON DOCTOR-PATIENT BEHAVIOR AND MEDICAL RESOURCE ALLOCATION ON INTERNET MEDICAL PLATFORM

Name: 廖家瑞 (Jiarui Liao) Student ID: 1801212884 Class: Fintech

1 Introduction

1.1 The project background

With the rapid development of the global economy, people are enjoying a constantly improving quality of life. But at the same time, the increasing aging of society, environmental problems and increased awareness of health management make the original medical resources in short supply. Compared with developed countries, China has been suffering from serious shortage of medical resources and unreasonable allocation of medical resources for a long time. According to the statistics of our ministry of health, our country has a huge population, accounting for more than one fifth of the world's population, but we only have two percent of the world's medical resources. In addition to the shortage of medical resources, the distribution of medical resources in China is also seriously unbalanced. The quantity and quality of medical resources in developed eastern regions are far higher than that in backward western regions [1]. The big cities in the east of China have most of the high-quality medical resources, and most of the medical resources in the cities are concentrated in some famous big hospitals. In addition, the traditional medical service system, which is opaque and inefficient, has been unable to meet the growing and increasingly diversified demands of the national health service. All kinds of problems force the society to explore a more reasonable medical service system.

The above mentioned problems create good opportunities for the development of Internet medical treatment. With the cross-border integration of technologies such as big data and cloud computing and mobile Internet, emerging technologies and new business models are rapidly penetrating into various segments of medical care. For example, network hospitals, online consultation and telemedicine have realized the transtemporal and spatial allocation of limited medical resources by virtue of the Internet, improved the communication ability between patients and doctors, broke through the traditional on-site service mode, and alleviated the shortage of medical resources.

In the Internet medical model, most patients do not need to see a doctor on the spot, but only need to make early medical diagnosis, obtain medical information and even medical plans through the network. Therefore, Internet medical treatment will greatly change the traditional medical model, and have a certain impact on patients' medical decision. How to use information and network technology to solve the imbalance of supply and demand and distribution of medical resources has become a hot issue in the field of research and practice.

Online medical service refers to the access of medical information by users through the Internet,

mainly through instant messaging such as online message dialogue or telephone communication to professional medical personnel for medical consultation and access to medical services (online diagnosis, treatment advice, health consultation).At the same time, patients can also make appointments for consultations, communication with patients and so on. There are currently tens of thousands of health-related websites in the world [2].

Established in 2006, Haodf medical platform is the leading Internet medical platform in China. (https://www.haodf.com) By the end of 2016, Haodf had collected 480,000 doctors from 7,216 regular hospitals across the country online. Among them, 142,000 doctors registered their real names on the platform and directly provided medical consultation, disease management, science knowledge and other services to patients, reducing the cost of treatment for patients (especially those from other places).

Considering the availability of data in the big data project, this project is based on one data crawler project during my undergraduate period. The project has completed the work of data crawling, but has not carried out subsequent cleaning, analysis and visualization. I have collected all hospitals of Shanghai city department of the relevant data, including basic information on doctors and patients, the geographical position, participation, including the time of registration, login, and stay logged in time, number of posts, etc.), post (post, reply, questions), the content of the post (the length of the posts, posts, the theme of the emotion), evaluation data, etc. This project will examine the impact of the Internet on healthcare from the perspective of doctor-patient interaction (reflected in doctor-patient consultation and response).The project will focus on the improvement effect of the Internet on the imbalance of medical resources, and also plan to visualize the geographical distribution and spatial distribution of medical allocation on the Internet medical platform. The research conclusion of the project hopes to put forward corresponding Suggestions on the influence of Internet medical treatment on doctor-patient behavior and resource allocation.

1.2 Research Propose

This project is based on the Internet medical platform -- doctor-patient data of good doctors. **Through data mining and analysis, the purpose of the project is to study the impact of Internet medical platform on the allocation of medical resources and relevant evaluation, to obtain the regional distribution characteristics and temporal and spatial changes of doctor-patient communication in Internet medical treatment, and to analyze the main factors influencing the regional differences of Internet medical treatment in order to give optimization directions and suggestions.**

This project is planned to focus on the following parts:

1) Visualization of the geographical distribution of doctor-patient interaction

The geographical distribution of various indicators is the focus of this study. We take Shanghai as the radiation center, as the geographical location of doctors, is the representative of high-quality medical resources; The location of the patient is the radiation location, which is the geographical point to obtain high-quality medical resources through the Internet. Therefore, through data analysis and visualization, we can make the geographical distribution of doctor-patient interaction frequency in each department, and study the geographical distribution characteristics of doctor-patient communication on the Internet medical platform.

(2) Visualization of doctor-patient interaction time distribution

The characteristic that the frequency of doctor-patient interaction changes with time is also one of the objects of this study. I collected the data of doctor-patient communication and consultation on Haodf platform during the ten years from 2006 to 2016, and drew a monthly visual map to show the change of geographical distribution over time, from which I could study the development characteristics and status of the Internet medical platform and give corresponding Suggestions.

(3) Influencing factors of regional differences in doctor-patient interaction

We can use the data to further study the regional factors that influence indicators such as doctor-patient communication level. Medical quality level, geographical location (spherical distance from Shanghai) and economic development level (GDP) are used to describe the regional differences, establish appropriate mathematical models, and give corresponding development Suggestions for the Internet medical platform.

1.3 Research Methods

This study mainly focuses on the regional and temporal distribution of doctor-patient communication as well as the regional influencing factors of doctor-patient communication, aiming to analyze the main factors influencing the regional difference of Internet medical treatment taking good doctors as an example. The study provides reference and basis for users of online medical community and operators of online medical community. The research is mainly carried out from the following parts:

1) Theoretical modeling

The construction of the theoretical model needs to consult a large number of literatures. Through reading and understanding the literatures, we find the supporting theory of the project research and establish the research conceptual model.

2) Data analysis and visualization

The data used in the research mainly come from Haodf medical platform, where patients can interact with doctors accordingly. **The data included the location of the patient, times of consultation, consultation time, as well as the identity information of the doctor, the location of the hospital and the grade of the hospital.** Data analysis can adopt data cleaning, database establishment and connection and other methods. Data visualization can use a variety of visualization software for data visualization operations.

3) Mathematical model verification

According to the supporting theory and the research hypothesis, the mathematical model is established and the corresponding hypothesis is put forward. In this study, R language and python were used to analyze and model the established data set, and corresponding conclusions and Suggestions were given.

2 Data Description

2.1 Data overview

On the Haodf website, the project can be collected relevant data of all hospitals in Shanghai, including basic information on doctors and patients, the geographical position, participation,

(including the time of registration, login time, stay time, number of posts, etc), post (post, reply, ask questions), the content of the post (the theme of the posts, the length of the posts, the emotion), evaluation data, etc.

The data used in the project is divided into three parts.

The first part includes hospital information, such as location, grade, department name, etc.

Figure 1: Hospital information interface (Source: Haodf website)

The second part includes the information of doctors, such as location, title and service status.

Figure 2: Doctor information interface (Source: Haodf website)

The third part includes patient information, such as: location, consultation times, consultation time, etc.

Figure 3: Patients information interface (Source: Haodf website)

2.2 Intrinsic big data properties

The data of this project conforms to the inherent characteristics of big data.

1) **Variety:** both structured and unstructured. Structured data, such as the age of the patient, the title of the doctor, the number of consultants, etc.

2) **Volume:** the project has collected relevant data of all Shanghai hospitals on the website of Haodf for nearly ten years, with the total number of tens of millions, and the storage capacity is TB, which is consistent with the characteristics of large data Volume.

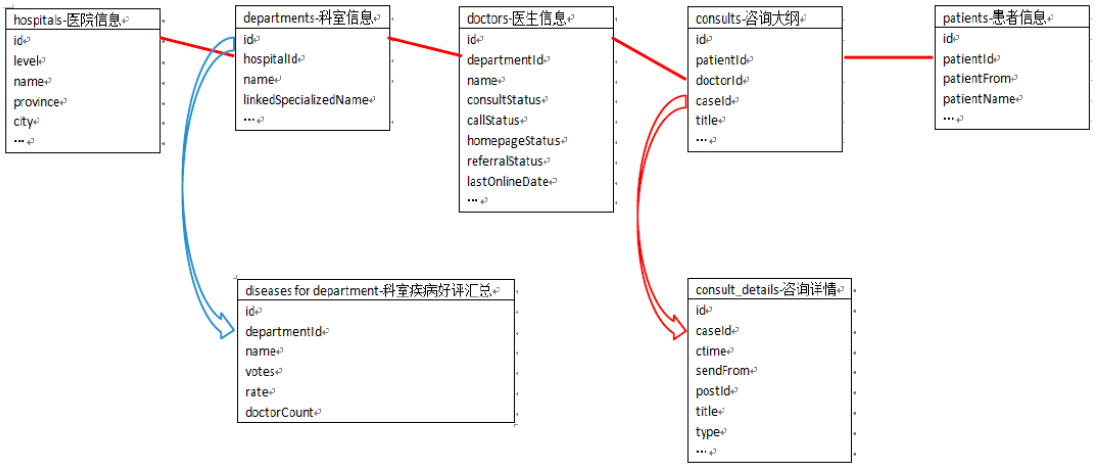
3) **Velocity:** project data includes doctor-patient communication data all the time. The data changes quickly and must be processed quickly.

4) **Value:** the data of the project itself has a low Value density, but if appropriate processing is carried out, valuable information can be mined, such as the improvement effect of Internet medical model on the allocation of medical resources.

2.3 Relational databases

This project plans to use MySQL relational database to store data. The MySQL database can connect the data tables used by the project to obtain the table of research objects, representing the relationship between the data. The connection relationship between the tables is shown in the figure below.

Figure 4: Data storage relationship description



3 Project workflow

The big data project is a data project for the Internet medical platform. **The process of the project includes data crawling, data cleaning, data storage, data analysis and visualization.** This project is very novel and explores the geographical distribution of patients on the Internet medical platform and the regional factors influencing their decision-making. This project collected ten years' doctor-patient consultation data in Shanghai, and planned to study and compare the differences in the geographical distribution of consultation volume of patients in different departments and at different times, and to make a visual expression. At the same time, the project plan analyzed the factors influencing the decision-making of patients in different regions, and speculated that it might be related to **the local medical level, economic development level and geographical distance.** The project is of great significance in understanding the role of Internet medical platform in medical resource allocation, laying a solid foundation for further research on the impact of Internet medical platform on medical resource allocation.

References

[1] 张文娟，郝艳华，吴群红等，我国医患关系紧张的原因及对策[J]，医学与社会，2014，27(4)：44-46.

[2] 中投顾问，2013-2017 年中国医疗信息化建设投资分析及前景预测报告[M]

Appendix

The appendix shows some data visualization charts:

Figure 5: Regional distribution of oncology patients

Figure 6: Heat map of patient area distribution

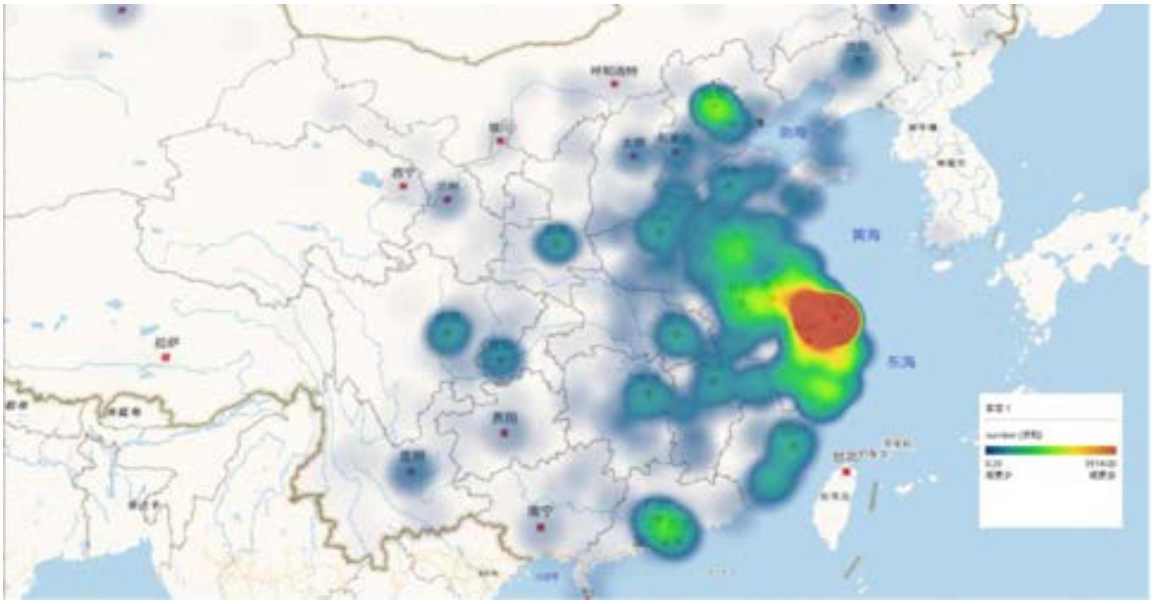


Figure 7: GDP per capita by region, number of patients and Internet users