Design of cancer classification and visualization platform based on Internet big data

Yipeng Xu¹, Zhixiao Zhang² and Min Liu³

¹School of Mathematics and Science, Tian Gong University, Tianjin, 300387
²School of Remote Sensing and Information Engineering, Wuhan University, Wuhan, Hubei, 430079
³School of Physical Science and Technology, Yangzhou University, Yangzhou, Jiangsu, 225000

Abstract. With the development of Internet technology, Internet technology is widely used in the medical industry. Based on the medical Internet big data, using visualization, data mining and other technologies, combined with distributed crawler, HTTP data transmission technology, vue2 front-end framework and the application of visualization technology, this paper completed the development and design of the analysis platform based on the internet medical big data. From the three aspects of presentation layer, business logic layer and data access layer, the architecture can show the number of hospitals, the proportion of doctors, the number of departments, the hospital's praise, poor evaluation and other information. Then, the cancer big data classification model based on the new bionic optimization algorithm is established. The experimental results show that the method has high classification accuracy and speed. In the era of big data, the integration of early diagnosis and treatment of cancer resources and information construction is very necessary and has far-reaching significance.

Keywords: Internet, big data, INBO, architecture, visualization.

1. Research background and significance

1.1. Background
In recent years, with the rapid development of China's economy, it not only improves people's living standards, but also drives the development of medical and health fields. Medical informatization continues to heat up, mobile medical technology and Internet of things technology continue to develop, and cloud computing big data technology is also emerging [1]. Medical and health information platform, business system and other rapid popularization [2]. From 2015 to 2018, more than 20 important documents were issued by important national departments to promote the combination of medical industry and Internet technology, which shows the investment and determination of the state to promote the development of medical industry. With the support of national policies, China's medical and health undertakings have made great progress in medical resource allocation, medical institution services, and online medical services. Meanwhile, hospitals are constantly trying to integrate Internet plus into their own businesses, and constantly upgrade the overall level of hospital business.
1.2. Significance
Under the background of big data era, with the development of medical and health undertakings and the promotion of intelligent medical treatment, the amount of medical information is also growing exponentially every day [3], which is closely related to people's work, learning and life. It not only brings convenience to people, but also brings great difficulty to data analysis and display. The medical big data accumulated year by year contains rich information and inestimable value. How to obtain specific analysis results efficiently, accurately and intuitively to support decision-making is a problem that people have to face while enjoying the convenience brought by big data. The so-called big data, that is, based on the "massive" and "huge" multi-type data, through its rapid collection, analysis and processing, to achieve data application and management value creation process. With the rapid development of cancer prevention and treatment research in China, the comprehensive informatization and digital networking are the important trends. It is necessary to strengthen the construction of standardization and sharing of the overall information.

2. Design of medical big data analysis platform

2.1. Overall architecture design of the platform
The analysis platform of medical big data based on the Internet is designed with B/S architecture. The main functional framework is based on vue2, supplemented by element UI style model. Java script is the overall programming language based on vue2 component module. The storage and processing of medical data use Java script combined with mongoose. Mongodb database stores most of the development data, a small amount of which directly uses JSON The text format is stored on the server. As a whole, the platform can be divided into three parts: presentation layer, business logic layer and data access layer. These three parts are related to each other, but the layers are weakly coupled. If you need to maintain a part, you only need to maintain the corresponding module, which is easy to expand and maintain.

(1) Presentation layer
The presentation layer defines the human-computer interaction interface. Its main task is to be responsible for displaying the data processed in the background and the corresponding functions of the browser visualization platform to the target users in the form of pictures, tables or other intuitive forms. This layer directly displays the data from the background or displays the background data by receiving user commands.

(2) Business logic layer
The business logic layer includes the implementation of the business logic of the whole analysis platform. This part of business logic includes but is not limited to a series of business rules, such as the pull of background data, the aggregate query of data, the rendering of acquired data, and the network communication protocol. Its main focus is to develop business rules, achieve business processes, business optimization and design business-related systems. From this, we can see that the business logic layer is the bridge between the presentation layer and the data access layer. The business logic layer knows nothing about the view layer, whereas the view layer does not contain the business logic but knows the business logic layer. The data access layer is equivalent to the white box to the business logic layer, and vice versa. In short, there is a downward dependency between the three layers, and the business logic layer bridges the view layer and the data access layer.

(3) Data access layer
The data access layer is mainly connected to database or local JSON format data. In this part, the data is pulled from the database according to the instructions from the users in the view layer, and then fed back to the corresponding display module through the data business logic layer. Database is the direct source of the vast majority of data, which stores all medical data crawled and preprocessed by reptiles. In order to support the writing of business code in the business logic layer to access the system database, and to realize the query, aggregation, statistics and other operations of database content according to the needs of users. In addition, it has been mentioned that the source and format
of the data obtained in this paper may not be uniform. Therefore, before the data access layer obtains the data, the data is standardized and processed into JSON format. This is the data format required by the platform and the lowest cost data conversion format. After the data is accessed, it is fed back to the view layer through the business logic layer as mentioned above.

2.2. Platform module design

The visualization part is divided into operable and non-operable visualization. Non-operational visualization means to display directly on the visualization map or in the form of various graphs and charts, generally displaying information such as the number of hospitals, the proportion of doctors, the number of departments, the favorable comments and poor comments of hospitals. In addition, the map shows the densely distributed areas, and other graphs show the multi-index analysis results of major hospitals. The data used in this paper has obvious regional characteristics. Through the visual display of map data, we can clearly analyze the distribution and proportion of hospitals, doctors and expert doctors in different regions, so as to provide a clear and intuitive understanding for the development of medical industry in different regions, while the graph curve helps us understand the correlation between different indicators in key hospitals. It is expected to innovate the medical industry according to the relevant relationship. The completion of this part is that the background data is processed by the logical business layer to obtain specific appropriate data indicators, and the results are directly displayed to the user. The operation of the data is completed by the background business logic code.

Based on crawler technology, data mining technology, Vue front-end technology, node service technology, visualization technology and database technology, this paper realizes the effective integration of data presentation layer, business logic layer and data access layer of the platform. On this basis, the medical big data analysis platform is developed.

Operable visualization display mainly feeds back the results in the form of tables, that is, the platform receives data user instructions, processes data and feeds back the results to users in the form of graphs or data tables. In terms of function, it is a query, actually a simplified ranking recommendation system. The original purpose of this part of function design is to enable the staff to obtain the distribution of medical resources in different regions and the analysis results of relevant hospital indicators, while using machine learning method to provide authoritative regional hospital impact ranking, as well as the recommendation of regional experts and doctors. It can make the platform rich in accurate positioning functions on top of statistical analysis functions. The main processing of the expert doctor recommendation module is to preprocess the data according to the existing data features combined with the Jieba segmentation, establish the feature thesaurus, and use node2vec for doctor attribute features Construct the feature vector, include all doctor features in the feature network and map them to a high-dimensional space, use the doctor vector accumulated and standardized by the same doctor attribute feature vector in the network, at this time, input the disease in the feature network database, use cosine similarity calculation to get the ranking of doctor's field similarity, and expert doctors are ranked TOP.

3. INBO model

3.1. Improvement strategy

**Define 1** (Levy flight path) the individual motion path of the new bionic optimization calculation (Dragonfly algorithm) obeys the Levy flight path expressed by formula (1) and formula (2), so that the algorithm motion path can maintain good inertia.

\[
L_{e V}^{y} (\beta) = \frac{\mu \times \phi}{|V|^{\frac{1}{\beta}}}
\]  

(1)
\[
\delta_n = \left\{ \frac{\Gamma(1 + \beta) \sin\left(\frac{\pi \beta}{2}\right)}{\Gamma\left[\left(1 + \beta\right)/2\right] \times \beta^{3/2}} \right\}^{\beta^{-\lambda/2}}
\]

Define 2 (multi-step surrounding disturbance) new bionic optimization calculation (Dragonfly algorithm) in which the individual increases the multi-step surrounding disturbance, and increases the multi-step surrounding disturbance operation to the dragonfly individual, which can increase the diversity of the dragonfly individual and is conducive to the global optimization. Formula (3) is a disturbance operation.

\[
\Omega = \text{Cauchy}(0,1) \times \Omega_1 + N(0,1) \times \Omega_2 + T(0,1) \times \Omega_3
\]

3.2. Big data classification model of cancer based on INBO

Using INBO as the classification model of cancer big data, because the cancer data is big data, the high precision and fast speed of INBO solution meet the needs of big data classification. In the process of classification, the similarity value between data is used, and those with large similarity value are divided into one class. The specific process of tumor big data classification model is as follows:

Step 1: use INBO to calculate the key class set. The key class set can represent the global characteristics, so the selection quality of the key class set affects the classification results.

Step 2: the weight estimation of each key class in the key class set must be balanced to avoid excessive inclination.

Step 3: calculate the similarity value between the data according to the key class set, and the similarity value is the standard to measure whether the data is divided into one class.

Step 4: using INBO to construct the classifier, the classifier must have certain robustness, and be able to handle some special values to avoid the negative impact of special values on the overall classification results.

Step 5: output the classification results.

The experimental verification is carried out by using the experimental database of tumor and cancer big data, and the comparative analysis is carried out by using three algorithms (A: the method in this paper, B: Bayesian classification, C: K-mean classification). The experimental comparison is carried out from the accuracy of classification and classification time. The experimental results are shown in the figure below.

![Figure 1. Comparison results of classification accuracy](image-url)
4. Conclusion
In recent years, China's medical information industry shows a high-speed growth trend, which promotes the rapid development of medical information. The vigorous development of medical information industry has given birth to a considerable scale of medical platform. These medical platforms are committed to providing convenient medical services for ordinary users, which not only facilitates people's access to medical resources, but also accumulates a lot of data. This paper will mainly focus on the construction and application of medical big data platform, including medical resources of various dimensions such as doctors, diseases, hospitals, etc. the data needed to develop this platform is benefited from the rich data of such websites. Then, taking the upper gastrointestinal cancer screening based on Chinese population as an example, this paper discusses the future prospects and comprehensive utilization of a large number of resources and information such as basic information and risk factor information collection, registration, statistics, treatment, follow-up and so on. The purpose is to further integrate the comprehensive digitalization of early diagnosis and treatment database link resources, and explore the cancer in the era of big data the necessity and importance of information construction of early diagnosis and treatment resources.

References
[1] Liu Zhi, Zhang Quan Ling. A review of big data technology research [J]. Journal of Zhejiang University (Engineering Edition), 2014, 48 (6): 957-972
[2] Tu Xinli, Liu Bo, Lin Weiwei. Review of big data research [J]. Computer application research, 2014, 31 (6): 1612-1616
[3] Yu Guopei, Bao Xiaoyuan, Huang Xinting, et al. Types, properties and related issues of medical and health big data [J]. Journal of medical informatics, 2014, 35 (6): 9-12