Research and development of the clinical thinking mining and discovery system based on artificial intelligence

Li Peng, Hongwu Zeng*
School of Medical Information, Chongqing Medical University, Chongqing 400016, China

ABSTRACT

An AI-based clinical thinking mining discovery system that can be used on PCs and mobile terminals (cell phones, pads, etc.). The system utilizes artificial intelligence technology to obtain massive medical case data and case discussions from the Internet for data extraction and organization, and intelligently classifies them by disease type and clinical presentation. Using natural language processing technology and intelligent mapping mechanism of medical terms, it mines the information of clinical features, tests, inspections, disposal measures and reasons (drugs, surgeries, etc.) of the extracted cases, abstracts and visualizes the diagnostic rules and clinical treatment channels of the cases. The medical case data processed on the Internet will be combined with the typical case data in the HIS system to form a clinical knowledge repository of diseases, guide junior doctors and students to conduct clinical thinking training and consolidate medical knowledge.

Keywords: Artificial intelligence, clinical thinking, mining

1. INTRODUCTION

1.1. Telemedicine education is an effective means to improve the clinical skills of primary medical care and medical students

The low capital investment, low training opportunities for medical personnel, relatively low education and relatively lack of medical hardware resources in grassroots hospitals have led to a lower level of diagnosis and treatment in grassroots hospitals. Although medical students in school have rich theoretical knowledge, clinical skills are not available. In recent years, the rapid development of artificial intelligence and mobile Internet has provided convenient conditions for telemedicine education. The construction of an AI-based clinical thinking mining and discovery system, which provides a convenient medical knowledge re-education platform for grassroots medical personnel and medical students in schools, is a crucial measure to achieve the goal of “adjusting the imbalance in the distribution of medical resources, accelerating the construction of grassroots medical and health service system, and promoting the equalization of urban and rural medical and health services”, and is an effective approach to achieve lifelong learning in the era of knowledge economy.

1.2. Cultivating clinical diagnostic thinking is the basis for diagnosing diseases

Clinical thinking is the integration of knowledge from four disciplines: medical sciences, humanities and social sciences, natural sciences and behavioral sciences. Clinical thinking takes the patient as the central point, and after sufficient communication and exchange, gets first-hand information in history taking, physical examination and auxiliary inspection, combined with available evidence and information such as the patient’s family background and humanistic context. The clinician then critically analyzes the above information and makes a different diagnosis, and finally completes a personalized plan for diagnosis, treatment, rehabilitation and prevention and puts it into practice.

With the speedy development of medical science and technology, various high-tech examination techniques have been introduced one after another, providing an important clinical ground for diagnosis. There is a tendency to be reliant on high-tech examination technology to diagnose diseases in the current clinical work, but in reality, a doctor can only form a specific thinking to understand medical activities in a targeted and conscious way. The ability to reflect on clinical problems independently is not only a requisite quality for doctors, but also a manifestation of clinical thinking ability. In order to improve the diagnosis and treatment of diseases, it is imperative to cultivate the ability of doctors to think autonomously about clinical problems, and it is extremely valuable to train doctors in strong clinical thinking.

*dowsam8@sina.com

International Conference on Computer Application and Information Security (ICCAIS 2021),
edited by Yingfa Lu, Changbo Cheng, Proc. of SPIE Vol. 12260, 1226021 · doi: 10.1117/12.2637407

2. CURRENT SITUATION AND SIGNIFICANCE OF THE RESEARCH

2.1. Research status

2.1.1. Serious decoupling of medical education applications from user needs. In recent years, the utilization of WeChat applets and WeChat public numbers for telemedicine education has been well developed, and the WeChat applets and WeChat public numbers of some societies, colleges and hospitals provide abundant, high-quality and free teaching materials, so that grassroots medical personnel and medical students in schools have the access to facilitate medical education resources. Nevertheless, their shortcomings are that the teaching content lacks interaction, mostly presented in a static form, the resources provided are limited, and the content is often not what users need most. I confirmed the author's idea by visiting some township health centers and medical schools. Doctors in primary hospitals can learn relevant theoretical knowledge through various books, and medical students want to improve themselves through practical learning as soon as possible after learning theoretical knowledge. Doctors and medical students prefer to have a medical education program that is easy to operate and can quickly increase their practical business skills by using fragmented time.

It can use artificial intelligence technology to accurately push systematic and rich case materials related to their own profession for users, taking micro-video, micro-case explanation, micro-interaction, online training sessions and so on. Users can make use of their fragmented time to participate with PCs and mobile terminals, which are more accessible, and the instruction format will be richer and the experience will be better.

2.1.2. Lack of case-type medical professional teaching. With the overall teaching hours in universities being tight, case-based classroom teaching for students is unlikely. Clinical medicine is a very practical science, and primary care physicians have relatively little exposure to surgical cases, and most of them are currently learning through self-study, conferences and training courses, with limited effect. In order to investigate the development of the relevant platforms, we searched the App Store of Apple, the Android application market and Baidu, but could not retrieve the relevant applications, indicating that the practical activities of case-based medical distance learning using PCs and mobile terminals are most likely not carried out. In addition, clinical conferences and training courses at various levels in China usually charge a conference fee, and there is rarely a free remote broadcast. Even if there are a number of training of public interest, they suffer from a wide range of problems, such as the scarcity of training sessions per year, short duration, and conflicting schedules, which obviously fall far short of meeting the educational needs of primary care physicians.

2.1.3. Scarce internet medical data mining. With the growth of information technology, telemedicine and online medicine have expanded rapidly, and medical data on the Internet has exploded. Medical data on the Internet includes discussions by doctors on medical websites and various forums about a disease (the possibility of the existence of the disease), descriptions by patients of a disease (signs, symptoms, etc.), and responses by doctors to patients' diseases (diagnosis, disposition recommendations, etc.). Such data bear great clinical and learning value, but is currently underutilized.

Case data (doctors and patients) is collected from the Internet, and natural language processing technology is used to intelligently classify these data, and the case discussions are extracted and combined with case data in the HIS system to form a case knowledge base. This knowledge base can provide guidance to junior doctors and students for clinical thinking training and consolidation of medical knowledge.

2.2. Significance of the research

With the penetration of the mobile Internet, medical information on the Internet is explosively growing, and the effective use of medical information on the Internet is conducive to the advancement of medical affairs.

China's distance learning resources are very limited, which obviously cannot meet the market requirements of primary care physicians and medical students. Primary care physicians for case-based training, especially in the western region are almost missing, and artificial intelligence-based clinical thinking teaching resources is a gap. It is imperative to increase investment.

With artificial intelligence as the means, the Internet as the carrier, and medical knowledge as the basis, this project aims to improve the diagnosis and treatment level of primary care doctors and provide a learning platform for medical students in school to cultivate clinical thinking. The implementation of the project will benefit the vertical flow of high-quality medical resources, and the use of the platform can promote the fast development of graded diagnosis and treatment.

3. SYSTEM IMPLEMENTATION

3.1. System prototype

The system uses Web crawler technology to collect relevant data on the Internet, applies natural language processing technology and medical terminology intelligent mapping mechanism, mines and extracts information on clinical features, tests, examinations, disposal measures and reasons (drugs, surgery, etc.) of cases, abstracts and visualizes the treatment rules and clinical treatment paths of cases. The processed medical case data on the Internet and typical case data in the HIS system are combined with the clinical knowledge base of diseases to form the knowledge map of treatment thinking. Figure 1 below shows the general architecture of the system:

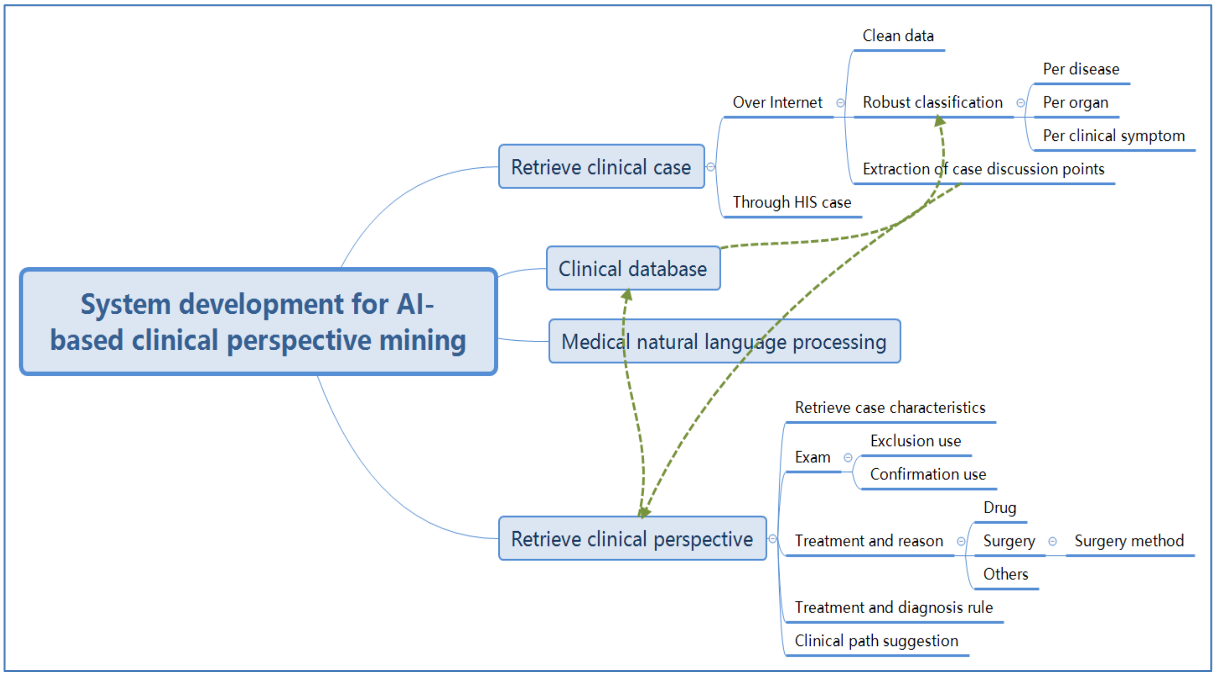

Figure 1. Shows the overall architecture diagram of the system.

3.2. Platform architecture

Figure 2 shows the schematic diagram of the system platform. PC, smart phone, tablet PC and other devices can access the system through mobile network or WIFI for training.

Figure 2. Platform architecture diagram.

3.3. Platform framework design

On basis of the platform architecture design, fully into account the interactivity and convenience of the functions, the platform framework structure comprises five interconnected hierarchies: system resource layer, data storage layer, data extension layer, business, management layer and application service layer (Figure 3).

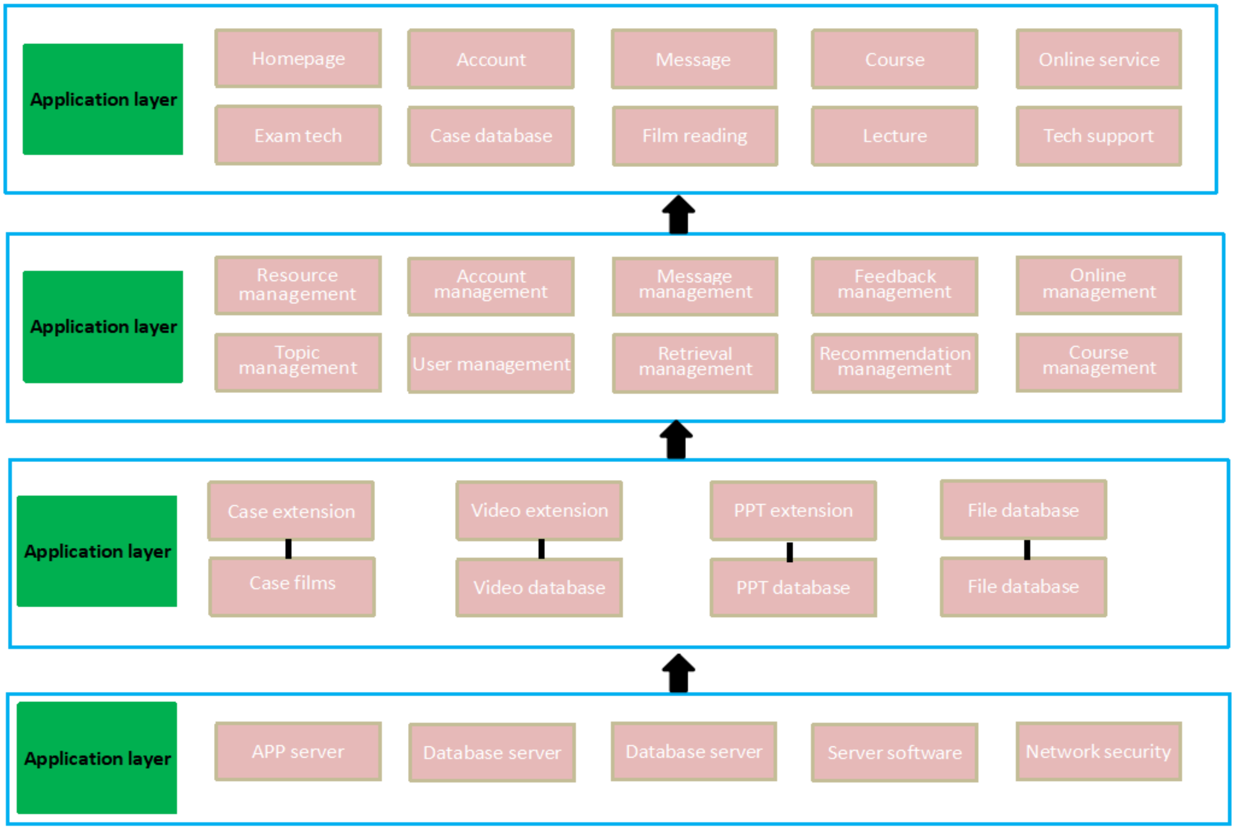

Figure 3. Platform frame structure.

(1) System resource layer: the foundation and base of the platform, including App server, database server, operating system, server software, and network and security.

(2) Data storage layer: store and manage data such as video, pictures, PPT, documents, etc. in different physical ways.

(3) Data extension layer: The stored data is described according to different data standards for labeling data features, realizing extension, and enabling knowledge discovery by extracting and storing traces generated by users in the learning process.

(4) Business management layer: It is the management interface for platform administrators, covering resource management, account management, message management, opinion management, online communication, topic management, user management, retrieval management, recommendation management and course management.

(5) Application service layer: It is the interface for users to operate when they access the platform, integrating several application services, which include home page, account, message, course, online communication, inspection technology, case database, teaching reading, special lecture and technical service.

3.4. Data collection and processing of the internet

Internet data collection is generally performed through web crawlers, and its basic framework is shown in Figure 4. The framework is designed to meet the real-time system for large-scale Internet data collection. Its input is a URL obtained from a web page, and the required data is output after data cleaning. The URL is first parsed and web resource information is collected, and then the resource information obtained in the first step is cleaned to obtain the target data. This study will use python third-party library requests to complete the data collection. The data cleaning is implemented using the HarK tool, which can be quickly implemented with a data cleaning model. The tool is mainly used to filter the data acquired by the web collector module to get the data in the target format. The module includes dozens of sub-modules, which are roughly divided into four categories: generation, transformation, filtering and execution.

Figure 4. Internet data collection framework.

3.5. Design of the recommended system algorithm based on the user behavior

3.5.1. Relevant concepts. A recommendation system algorithm based on user behavior is called a collaborative filtering algorithm. Collaborative filtering means that users can work together to increasingly satisfy their needs by constantly interacting with the website so that their recommendation list can filter out information that does not interest them.

The simplest form of user behavior data that exists is logs. Many Internet businesses aggregate raw logs according to user behavior into session logs, where each session represents user behavior and corresponding services, such as display logs and click logs. User behavior in personalized recommendation system is generally divided into explicit feedback and implicit feedback. Explicit feedback includes behaviors that users explicitly express their preferences for items. Implicit feedback refers to behaviors that do not explicitly reflect the user's preferences. Compared to explicit feedback, implicit feedback is not explicit, but it is substantial.

3.5.2. Algorithm implementation. (a) First find the set of users with the same interests as the target user;

(b) Find the content that users in this collection like and that the target user has not heard of and recommend to the target user. The key to step 1 is to calculate the similarity of interests of two users. Here, the similarity of interests is mainly calculated using the similarity of behaviors. Given users u and v, let N (u) be the set of items for which user u has ever had positive feedback and let N (v) be the set of items for which user v has ever had positive feedback. The calculation is as follows.

$$W_{uv} = \frac{|N(u) \bigcap N(v)|}{\sqrt{|N(u)||N(v)|}}$$

(c) Key code functions

```
#Data reading
def get_data(file):
return data
#Divide the data into a test set and a training set
def split_data (data, M, k, seed):
  return train,test.
# Get user similar collections and eslish user inverted tables
def UserSimilarity (train):
# Seek the user similarity matrix W
W = np.zeros ([NumOfUsers, NumOfUsers], dtype = np.float)
  for u,users in C.items ():
  return W,user_related.
# Generate recommended record to users through a similar matrix W
def recommend (User, train, K, N, W, user_related):
return rank.
```

4. CONCLUSION

The system builds an AI-based clinical thinking mining discovery system that can be used on PCs and mobile terminals (cell phones, pads, etc.), and the system comes in the form of PC WEB and mobile applications (APP).

The system effectively combines artificial intelligence technology, natural language processing technology, and Internet technology. The processed medical case data on the Internet and typical case data in the HIS system are combined with clinical disease ontology to form a clinical disease knowledge base to abstractly and visually present the treatment rules and clinical treatment paths of cases.

The system is based on cases, with small videos and micro-lessons as the main form, combined with relevant professional theoretical knowledge through a variety of forms of interaction.

The AI-based clinical thinking mining and discovery system provides an effective learning platform that can quickly upgrade the business level of primary care doctors and cultivate the clinical thinking ability of medical students, especially in rural hospitals where there are few cases and medical students cannot be exposed to more medical cases in school. The implementation of this project will gradually settle the problem of uneven distribution of medical resources between urban and rural areas, and will boost the development of graded diagnosis and treatment and precision medicine.

ACKNOWLEDGMENT

Foundation item: Intelligent Medicine Research Project of Chongqing Medical University ZHYX202003.

REFERENCES

[1] Ren, J. H., Shen, Y. B., Meng, X. F. and Wang, W., "Document clustering of entry-based associations," Computer Engineering and Application, 52(7), 86-90 (2016).
[2] Chen, L. X., "Current status and development of electronic medical records," Chinese Journal of Misdiagnosis, 31(2), 46-47 (2014).
[3] Liu, L., "Overview and problems of international clinical phenotype," World Science and Technology, 35 (6), 16-18 (2010).
[4] Cao, M., "Study of ontology-based vector space model in electronic medical records," Chinese Mathematics Medicine, 17(11), 35-39 (2018).
[5] Li, J., [Cluster Analysis and Its Application in Text Mining], Xi'an University of Electronic Technology, 33-35 (2015).
[6] Zhu, D., [Research and Implementation of the Ontology-Based Knowledge Visualization System], Chongqing University, 19-23 (2017).
[7] Xu, G. C., [Medical Case Reasoning Research Based on Ontology], Sichuan University of Technology, 78-81 (2014).
[8] Chen, Y., "Status of ontology research," Medical Intelligence Science, 37(7), 23-27 (2014).
[9] WHO Chronic Diseases, http://www.who.int/topics/chronic_ diseases/en/, (2016).
[10] Ren, S. Q. and Aung, K. M. M., "PPDS: Privacy preserved data sharing scheme for cloud storage," International Journal of Advancements in Computing Technology, 4(16), 493-499 (2012).
[11] Zeng, H. W. and Wang, J., "Clustering of electronic medical records based on association relationship," Chinese Medical Books and Intelligence Journal, 9, 73-79 (2018).
[12] Wang, J. and Zeng, H. W., "Design and implementation of intelligent medical system for chronic diseases," American Journal of Computer Science and Technology, 3(4), 86-91 (2020).