Intelligent Assistant Medical Diagnosis Based on Online Interactive Information

Youzhe Liu[a], Wenxuan Jiang[a], Tianqi Ren[a] and Weihui Dai[a]

[a]School of Management, Fudan University, Shanghai 200433, China
Tel: +86-21-25011241, E-mail: whdai@fudan.edu.cn
[blank line]
[blank line]

Abstract
[blank line]
The rapid development of the Internet industry has given rise to a large number of online medical communities and medical information websites, providing patients with a wide range of medical information and Accessible channels. These sites focus on health knowledge, disease information, medical news, etc., but also provide users with online interactive disease consulting. A large number of patients are consulting online on these online medical websites every day. Faced with a large number of consulting patients, online medical consultation faces a resource crunch. In this paper, we explore the textual interaction information between doctors and patients, improve the existing interaction model, enrich the interaction process and methods. We also introduce machine This is a learning method to intelligently extract diagnostic patient characteristics to guide the physician's focus and provide a diagnosis decision support to improve the efficiency and simplicity of interactions and help physicians make more efficient and accurate decisions about patient conditions.
[blank line]

Keywords:
Online Medical, Medical Diagnosis, Machine Learning, Semantic Analysis
[blank line]
[blank line]

Introduction

[blank line]
Online medical care is an Internet-based, outpatient medical treatment for the existing shortcomings. It uses the form and technology of online questioning and answering, instant messaging tools, etc. Internet-based medical activities, in which professional doctors provide patients with diagnosis of medical conditions and health advice. Currently, the mainstream online medical platforms include Chunyu Doctor and Good Doctor Online and so on. [1] According to data, the number of online medical users in China in April 2019 was 45 million, with a penetration rate of 6.6%. Influenced by the epidemic in 2020, China's online healthcare user penetration rate is expected to reach 7.9% in 2020. [2] In order to meet the increasing demand and improve the efficiency of communication between doctors and patients, it's meaningful and necessary to take good use of technology to extract key information from the Interactive mode and Q&A process. In this way, we can help to guide doctor to focus on the important part, that is, answer the exact question from patients instead of the complex and fuzzy way to get to the question. What else, we can also provide some technology support for the doctors in this way.

The existing online healthcare interaction model is clear enough, but not efficient enough to face the increasing demand for online healthcare and the market's progressive Expanding. Improving the efficiency and intelligence rate in online healthcare has become a pressing issue. One of the major pain points is that many online Medical questions and issues are too slow to be answered, and distractions such as wellness-related, chronic illnesses and unrelated issues cannot be addressed. Intelligent sequencing may be inefficient for some acute conditions.

This study seeks to optimize and simplify the design of the consultation process, in addition to intelligently extracting information about the semantics of patient consultation and machine learning classification.

In the process design part, the main focus is on easy acceptance of messages and recall mechanisms, taking chat messages for each patient's message. The mechanism, which takes the form of a one-to-one conversation for each physician, ensures that the physician has a better understanding of each patient's condition. The entity classification method is used in the classification of diseases, so that the general classification of diseases can be pushed. In addition, on the management side of the physician, the patients who ask questions are ranked in order of urgency, and the semantics of the questions are analyzed by machine learning methods. The post-partition training allows the physician to distinguish between urgent and smooth patient questions more precisely, thus making the physician more efficient in message processing.

On the one hand, key information was extracted from the description of the patient's condition by word segmentation, semantic analysis, entity recognition and other techniques. On the other hand, while using the entity in the statement to excavate, judge the anxiety and urgency of the patient. Providing results from these two dimensions facilitates the identification and classification of disease entities and the ranking of emotional and etiological urgency, which facilitates rational Matching physicians to different faculties and diagnosing them in the appropriate order can improve the efficiency of online medical implementation. Finally, a feasible intelligent medical assisted diagnosis technology and system process based on online interactive information is given to improve the online medical consultation efficiency and user experience.

The results of this study will serve as a complementary approach to online healthcare, which can effectively optimize the interaction pattern of online medical questions and answers and improve the doctor-patient Communicate efficiently and provide quicker and higher quality answers to patient questions.

[blank line]

2. Overview of online health care

[blank line]

2.1 Definition of online healthcare

[blank line]

Online medical care is the use of the Internet or mobile Internet to provide medical services, that is, the use of any part of the provision of medical services The Internet or mobile Internet is online health care. Online health care includes online health care, online diagnostic and treatment services provided to the general public or to patients, and the services associated with these services. The business of providing medicines, medical appliances; and services and tools such as social networking, expertise and online consultation platforms to doctors.

[blank line]

2.2 Classification of services in online healthcare

[blank line]

There are two main categories of online healthcare.

The first category is health care, for the general public, and the services are divided into health information, health management, health care drugs, and massage, in descending order. There are six categories of acupuncture, health screening and chronic disease management. The second category is diagnosis and treatment, which can be divided by user into two categories: patient-oriented and doctor-oriented, with patient-oriented services divided by the process of seeking medical care. Includes registration, consultation, diagnosis, treatment and payment; physician-facing services are divided into social, professional and online, in descending order. Questioning. As shown in Figure 1.

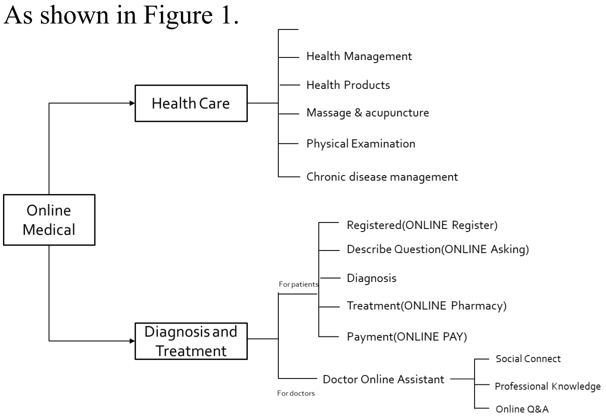

Figure 1

[blank line]

2.3 Current problems

In addition to the current lack of laws and regulations and the conservative wait-and-see attitude of some healthcare sectors that restrict online healthcare, most online Doctors are still staff members with establishment status, and "sharing" online may mean "losing" offline.[4] Online communication also means more communication barriers and inefficient allocation of medical resources. Therefore, effective use of information technology is needed to help doctors and patients improve the effectiveness of communication and thus the use of health care resources. Efficiency.

3. Interaction Pattern Design

3.1 Patient Side Implementation Process

The basic patient-side implementation is shown in the following figure 2.

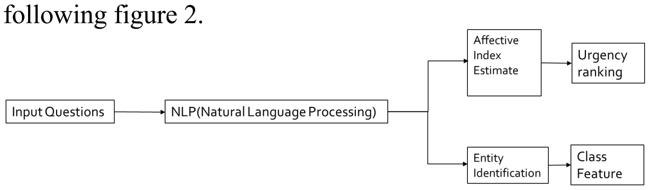

Figure 2

Relative to the existing interaction model, the main focus is on some degree of classification of patient descriptions. On the one hand, we need to try to process the named entity recognition on the disease classification. While most medical treatments for the online Q&A format will require the patient to select the category of illness they wish to consult. Considering actually many users are unaware of the attribution and cause of their condition, entity identification of the condition described by the user is not always a good idea. Correct categorization and secondary training for misjudgments to better match hospital departments. This method can reduce the barriers for patients to use online care applications and also improves the usage experience and diagnostic efficiency.

In addition, a two-level classification of the emotional characteristics of the questions posted by users will be developed. During online interactions, a large number of non-disease-related questions and non-urgent questions about health are often asked. For relatively urgent questions, the user's emotions are categorized into two levels. There are easily overlooked phenomena in the patient's condition. By means of natural language processing such as Chinese partitioning and matching, a suitable machine learning model is used to train the patient's questions to the polarized classification. We can rank the questions by patients according to the resulting urgency index. In this way, we can optimize the order of physician responses and improves the reasonableness and practicality of patients' consultation.

The flow chart on the doctor-patient interaction side is as follows.

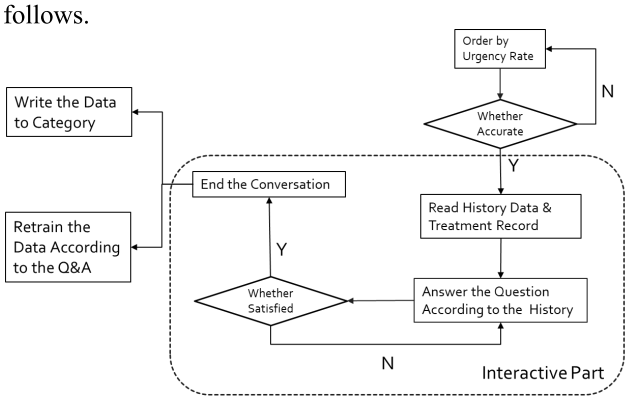

Figure 3

For the doctor-patient interaction part, the focus is on the innovation and optimization of the interaction mode. Based on the traditional one-time dialogue, we introduce the mechanism of the circular conversation, which means creating a small one-to-many chat room for the physician side, constantly push questions based on the above rules for the patient side. The conversation ends when appropriate answer is given. Repeat the cycle again if the patient still wants to pursue the question. At the same time, all questions and answers are written to a database, so that the next question can be prompted to the doctor as soon as it is asked. The patient's prior information to achieve a more effective diagnosis.

4. Introduction to the algorithm

4.1 Introduction to entity recognition algorithms

Named entity recognition is an important area of research in the field of natural language processing. The Sixth Conference on Message Understanding, held in 1995 (MUC-6) formally proposes the task of naming entity recognition. Its main task as a first step in text mining is to identify the words in the text that represent its body of knowledge.

There are three main types of work related to named entity recognition: lexicon-based approaches, heuristic rule-based approaches, and machine learning-based Method.

(1) The solution based on dictionary-method realizes entity identification through string matching, but has a strong dependence on the dictionary. In foreign countries, the English medical entity identification is becoming more mature. The available reference materials are more detailed. The most famous dictionaries include the International Disease Classification ICD 10 (International Classification of Disease-10), the Unified Medical Language UMLS (Unified Medical Language System) and the Medical Thesaurus MeSH (Medical Subject Headings). In the Chinese language, domestic research is still scarce and available resources are relatively scarce. [5]

(2) In terms of heuristic rule-based approaches, Kraus and others address the clinical records of university health systems by constructing Regular expressions that identify the drugs, dosages and dosage information mentioned therein.

(3) The more popular approach is based on machine learning. Named entity recognition can be viewed as a classification problem, using classification methods like support vector machines, Bayesian models, etc., and also It can be viewed as a sequence labeling problem using machine learning methods such as hidden Markov, maximum entropy Markov, and conditional random fields. [6] Sondhi et al. have conducted a study of information on disease topics on the medical forum HealthBoards, using SVM and CRF methods for superficial information extraction. The CRF methods have reached some good results in recent medical studies.

The specific processing of the CRF algorithm for words in a sentence is implemented as follows:

$$P(Y|X;\theta) = \frac{1}{Z(X;\theta)}\exp\left\{\sum_k \theta_k \psi_k(Y,X)\right\} \quad (1)$$

$$Z(X;\theta) = \sum_{Y'}\exp\left\{\sum_k \theta_k \psi_k(Y',\ X)\right\} \quad (2)$$

where θ represents the model parameter, which is an arbitrarily defined eigenfunctions of the parameter with respect to the observed sequence X and the labeled sequence Y. The model is based on an arbitrarily defined function of the parameter. It can also be seen as a normalization factor.

Named entity identification using the CRF model can be considered as a sequence annotation problem. Each sentence to be identified is treated as a sequence of observations. Each word in the sentence is treated as a symbol and each symbol is assigned a category label. One of the simplest structures of the CRF model is the chain structure.

The process for performing entity identification is shown in following Figure 4:

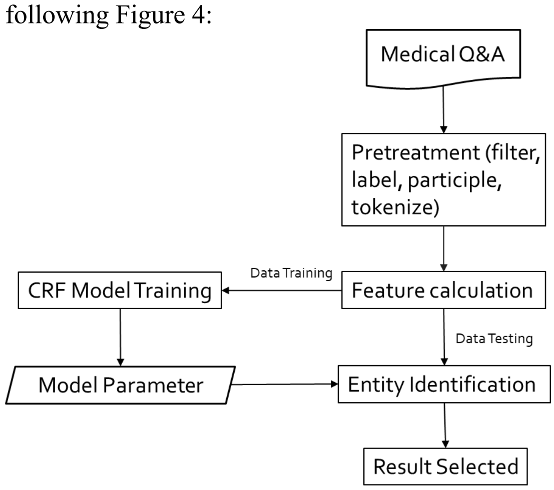

Figure 4（According to[5]）

4.2 Sentiment Index Analysis Algorithm

The dimension of mindfulness is specifically analyzed through the analytic model of frame semantics for the dimension of mindfulness.

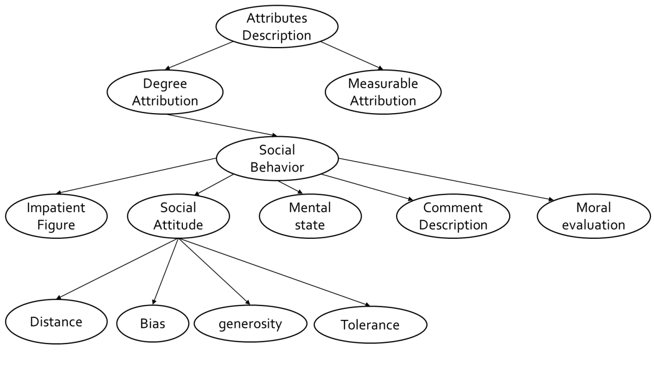

Figure 5(According to [7])

Specific implementation of recurrent neural networks:
The first is the labeling of the data. In the example, for each question from patients, the words are labeled with emotions on the anxiety/relaxation dimension, where 0 is the emotion rather flat and 1 represents a very urgent emotional condition. After marking 1 for urgent onset queries and 0 for health care and chronic illness queries, the Preservation of the text is classified.
For all data after labeling, the LSTM cyclic neural network model was used for training.
First of all we design the word vector from. In this word vector model, each word is an index corresponding to a vector of length 300. We need to construct the LSTM neural network model does not directly process the Chinese text. It needs to be subdivided and converted into a word vector first, using the Index the vocabulary after the split tool. Since the length of each statement is different, if we simply take the longest statement and fill the other comments with the same length. This is very wasteful of computational resources, so we take a compromise length. Now let's prepare the word vector matrix for the model. According to the requirements of the keras deep learning interface library, we need to prepare a The dimension is a matrix of (numwords, embeddingdim), with num words representing the number of words we use and emdedding dimension is 300 in our current use of pre-trained words in the vector model Each vocabulary is represented by a vector of length 300. The text at the end of these indexing is then filled and constructed for better training.
It's worth noting that we chose to use only the top 50k most frequently used words. Since for a total of 2.6 million in this pre-trained word vector model vocabulary, it would be a waste of computational resources to use it all on classification problems as our training sample is very small (only 4k in total). If we have 100k, 200k or even more training samples, we can consider reducing the use of the vocabulary. Training was then conducted using the long-short term memory (LSTM) with the following basic model equations:

$$f_t = \sigma_g(W_f x_t + U_f h_{t-1} + b_f) \quad (3)$$

$$i_t = \sigma_g(W_i x_t + U_i h_{t-1} + b_i) \quad (4)$$

$$o_t = \sigma_g(W_o x_t + U_o h_{t-1} + b_o) \quad (5)$$

$$c_t = f_t \circ c_{t-1} + i_t \circ \sigma_c(W_c x_t + U_c h_{t-1} + b_c) \quad (6)$$

$$h_t = o_t \circ \sigma_h(c_t) \quad (7)$$

Ultimately, it enables crude affective value calculations and
predictions of affective tendencies from input data.

5. System working model

Based on the information extracted from the user's Q&A data, we added the sentiment information embodied in the user's questions and answers to the analysis as well. Combining the departmental structure of the medical website and the implied emotions obtained from the analysis, we perform word segmentation and entity recognition of the user's Q&A data. To help the doctor make further diagnoses. The process is shown in Figure 6:

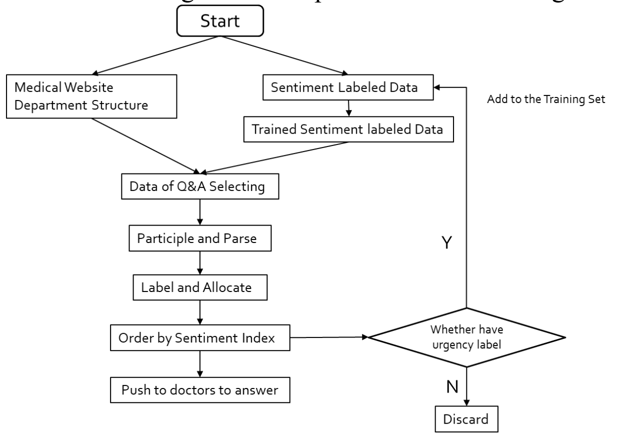

Figure 6

5.1 General structure of the system

The overall structure of the online semantic analysis system described in this paper consists of the identification of disease entity family lineage attributions, semantic sentiment index estimation, output the information function is composed of 3 modules. Among them, the attribution identification of disease entities simulates and realizes the assignment of departments to patients when registering, and realizes the function of automatic classification; emotional the estimation of the index module achieves the level of patient anxiety and thus the ranking of patients with different levels of urgency. The output module can automatically send the patient's statement questions and the corresponding tags to the appropriate department for the physician to diagnose. As shown in Figure 7.

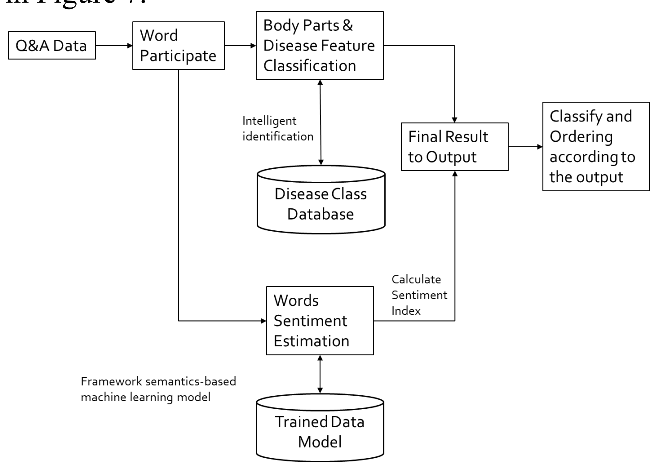

Figure 7

5.2 Framework for system implementation

The framework of the online interactive semantic analysis method described herein consists of four steps: entity data pre-processing, labeling and training of sentiment index data, the matching and labeling of the entities and the assessment of the

sentiment index, which are implemented as follows.
Step 1: Design and read the departmental structure from some large hospitals or large online medical interaction platforms, according to different diseases. This kind of data is relatively messy and can basically realize the attribution analysis and labeling of the major categories of diseases.
Step 2: Collect enough questions and answers about online healthcare to perform some degree of data cleansing for all user questions, rendering them meaningless. and irrelevant data for initial screening, and then pure questioning data for sentiment labeling, with 0, 1, respectively, indicating sentiment on the negative and positive, where positive means very urgent, and negative corresponds to the type of illness that is milder or asks the question The tone is more soothing for chronic diseases or health care issues. For all the data after labeling, the LSTM cyclic neural network model was used for training
Step 3: Use the CRF model to classify entities with the aid of the CRF depth commonly used in python. A learning library that automates the learning of already labeled data. The trained model is stored, and after entering a text, specific identification of body parts and disease names can be achieved.
Step 4: Output the entity classification results and sentiment index size and training data for the sentiment index from the sentiment index that the user responded to Conduct regular updates.

Summary

This paper focuses on the innovation of the existing interaction process and Q&A model for online healthcare and improve the efficiency and get optimized for the process. Separating the doctor-patient interaction side and the doctor management side of the process, we achieve the embedding of a one-to-many chat room based on the simulation of a classifiers and sorting push, as well as a history recall mechanism, resulting in a significant increase in interaction efficiency. In addition, in the specific algorithmically, the current methods for entity recognition in medicine are investigated, and the CRF model is selected as the system implementation methods. Unlike most existing research, we extracted the affective elements of the textual information entered by patients, using the current more effective LSTM cyclic neural network to output metrics of prioritize more urgent needs to push to physicians. This order can give a more effective of medical resource allocation. We present a holistic vision of the program. But currently it’s difficult to extract large amounts of data due to the privacy policy of medical websites, which makes it difficult to train the model. The accuracy of the solution implementation is still challenging.

Acknowledgments

[blank line]

The research was funded by School of Management, Fudan University. We thank for their valuable database contributions.

[blank line]

References

[blank line]

[1] Mao Ying, Jing Pengpeng, Zhu Bin, Liu Jinlin, Wu Jingxian, Li Yichen, “Studying on the complementation of telemedicine and online medicine” *Chinese Health Service Management*, *2016, 33(2).*

[2] The China Epidemic 2020: An Analysis of the Development and Investment Logic of the "House Economy" Report. https://www.iimedia.cn/c400/69247.html.

[3] Ai Rui Dennison: Mapping the Online Healthcare Industry Chain. *Ai Rui Report.*

[4] Opinions on Promoting the Development of "Internet + Medical Health", *The State Council of China.*

[5] Su Ya, Liu Jie, Huang Yalou. Research on entity identification in online medical texts[J]. *Journal of Peking University (Natural Science Edition),2016,52(1):1-9. DOI:10.13209/ j.0479-8023.2016.020.*

[6] Zheng Qiang, Liu Qijun, Wang Zhenghua, & Zhu Yunping. (2010). Research and progress in biomedical named entity recognition. *Computer Application Research, 27(3), 811-815.*

[7] Yoo Li-Ping, He Ling-Ling. “A Sentiment Analysis of Online Medical Reviews Based on Frame Semantics” Modern Intelligence, *Mar. 2020,Vol.40, No.3*.

[8] YE Feng, CHEN Ying-Ying, ZHOU Gen-Gui, et al. Intelligent recognition of named entities in electronic medical records[J]. *Chinese Journal of Biomedical Engineering, 2011, 30(2)*

[9] Lu Jiafa, Zhang Guoming, Chen Anqi. Deep learning-based disease diagnosis[J]. *Journal of Medical Informatics,2017,38(4):39-43.*

[10] Wang Ran Ran. Research on disease-oriented machine learning method for intelligent diagnosis[D]. *Jiangxi:East China Jiaotong University,2015.*

[11] ZHANG Ming-Huan, CHEN Ying, SHEN Ying, et al. Comparison of the performance of artificial neural network and support vector machine and its application in DMD disease recognition[J]. *Journal of Shanghai Institute of Technology,2016,38(4)*

[12] C. Yao et al., "A Convolutional Neural Network Model for Online Medical Guidance," *IEEE Access, vol. 4, pp. 4094-4103, 2016*

[13] J. Wang, C. Man, Y. Zhao and F. Wang, "An answer recommendation algorithm for medical community question answering systems," *2016 IEEE International Conference on Service Operations and Logistics, and Informatics (SOLI), Beijing, 2016*