Design of Intelligent Question Answering System for Hospital Online Triage based on Knowledge Graph

Jiahao Sun

Zhongshan College of Dalian Medical University, Liaoning 116085, China

2426420536@qq.com

Abstract. At present, people pay more and more attention to medical and health issues, and more and more Internet users are eager to obtain medical related knowledge from the Internet. However, despite the large amount of medical data spread on the Internet, it is difficult for ordinary Internet users to effectively obtain and use these data through search engines. To solve this problem, this paper designs and implements a medical question answering system based on knowledge map, so that users can quickly obtain the relevant knowledge in the medical field they want by asking questions, and realize the effective use of medical data on the Internet. At present, there are few question answering systems in the medical and health field, so the medical question answering system designed in this paper has a broad application prospect. The medical knowledge map constructed in the system can also be used in more application scenarios such as medical chat robots in the future, further promoting the process of intelligence in the medical field.

Keywords: Webcast; College Students; Consumption Behavior.

1. Introduction

Knowledge map, known as knowledge domain visualization or knowledge domain mapping map in the library and information industry, is a series of various graphs showing the development process and structural relationship of knowledge [1], map and display knowledge and the interconnections between them.

Knowledge graph is a combination of theories and methods of applied mathematics, graphics, information visualization technology, information science and other disciplines with quantitative citation analysis, co-occurrence analysis and other methods, and uses a visual graph to visually display the core structure of the discipline, Develop the modern theory of history, frontier fields and overall knowledge structure to achieve the purpose of multi-disciplinary integration, which can provide practical and valuable reference for disciplinary research[2].

With the arrival of the big data era, the informatization of the medical industry has also ushered in its own big data era. The key of medical big data analysis is how to obtain useful information from the surge of medical data. As one of the technologies of Semantic Web [3], Knowledge Graph provides a means to extract structured knowledge from massive data, and plays an important role in intelligent search, intelligent question answering, recommendation systems and other application fields. Knowledge mapping in medical field is the focus and hotspot of domestic and foreign scholars [4]. At present, well-known high-quality medical knowledge maps include IBM Watson, domestic Sogou AI medical knowledge map, Ali Health's "Yi Zhi Lu" and Shanghai Shuguang Hospital's TCM knowledge map.

The core of the question answering system based on knowledge atlas is to understand the semantics of the user's questions, that is, how to make the computer understand the diversified and unstructured expression of human questions. There has been some research on question answering system related to semantic web at home and abroad [5]. Some scholars proposed to construct SPARQL query by reducing dependency algorithm to improve the recognition rate of questions, thus completing the design of question answering system in e-commerce field. Some scholars have also designed and implemented a question and answer system for biological knowledge. By improving the R-WMD flat H logical regression model to generate subject entity recognition and entity link, compared with the R-WMD algorithm [6], the accuracy of entity recognition and entity link has been effectively

improved. AquaLogtSl is a portable question answering system that was put forward earlier. It uses queries expressed by natural language and ontology as input and integrates disambiguation and sorting methods to understand natural language queries. TBSL proposes an automatic question answering method based on templates. It generates SPARQL templates by parsing the natural language input by users, and then instantiates the templates to obtain query results, and obtains good results. The research of question answering system has developed from Turing test to today. Most of the systems with excellent performance are based on English, and few are oriented to Chinese medical field. Therefore, it is necessary to build a Chinese medical knowledge map for the medical field and build a question-and-answer system on this basis [7].

2. System Design

The whole system is mainly composed of knowledge map and intelligent question answering. Among them, intelligent question answering is divided into three steps. First, medical entities such as diseases and drugs in the question are identified by part of speech tagging. Then use the convolutional neural network based on the shared layer to obtain the question vector of the user's input question, and match it with the preset question template for similarity. At the same time, combine TF-IDF algorithm to obtain the question template with the highest score. After the question template is obtained, the real intention of the user's question is understood by obtaining its question type, and the natural language question input by the user is understood by template matching; finally, cypher query statements are constructed according to the obtained entities and question types to retrieve answers from the knowledge map and return them to the user.

First of all, the construction of the Chinese medical knowledge map. The data of the knowledge map comes from the vertical medical website, which focuses on diseases and mainly on structured data. It crawls the introduction, causes, complications and other contents in the disease page of the website, and stores the data in the MangoDB database. Through the analysis of the data in the database, various medical entities are obtained, and the concept layer of the medical knowledge map is extracted according to this, and then the entity relationship between nodes is constructed according to these entities. There are mainly two types of nodes in the knowledge map: one is the disease category, which includes disease attributes such as complications and symptoms; The other type is general entity nodes such as medicine and food, which do not contain attributes. Among them, for two different disease nodes, the first disease may be a complication of the second disease, so "company with" is marked as the entity relationship between them.

In the knowledge map, the organization form of knowledge is graph structure. As a native graph database, Ne04j uses a storage back-end that is customized and optimized for the storage and management of the graph database. Ne04j is selected to store the knowledge map, and the parsed entities and relationships between entities are imported into Ne04j. The data storage form of Ne04j is mainly to organize data by node and edge, where node represents entities in the knowledge map and edge represents relationships between entities. At the same time, Ne04j uses the query statement cypher designed by itself to retrieve the data in the knowledge base. Its characteristics are similar to those of SQL. Among them, match, where, and return are the most commonly used keywords: match clause is equivalent to select in SQL, which is used to specify the search mode; The where clause adds constraints to the match pattern.

In addition, another important function of this system is medical entity recognition. Traditional entity recognition refers to the recognition of entities with special meaning in text, mainly including person names, place names, organization names, etc. In this study, the method of part of speech tagging is used to tag all medical entities obtained by parsing. For example, the medical entity included in the question "What are the symptoms of cold" supported by this study is "cold", and its category is marked as disease; The medical entity included in "How to treat night sweats" is "night sweats", and its category is labeled as symptoms. After medical entities are classified and labeled, there will be no problem that the same entity may have different forms of expression. Entities will

only appear in the form of words. Load these labeled entities into the medical domain dictionary, and then judge the part of speech of the words after word segmentation. The words belonging to the above categories are the medical entities that need to be recognized.

3. Summary and Prospect

At present, a large number of Internet users have a need to obtain relevant knowledge in the medical field from the Internet. Although a large amount of medical information is spread on the Internet, it is difficult for ordinary Internet users who lack professional guidance to extract such knowledge from the Internet only by conventional search engines. To solve this problem, this paper constructs a knowledge map in the medical field based on relevant public data, and designs and implements a medical question answering system on this basis.

Although the medical question answering system designed and implemented in this paper has basically achieved the expected functions, there are still many areas that can be improved.

3.1 The Source of Medical Knowledge Map Data

The source of medical knowledge map data used in this paper is relatively single. On the one hand, the data crawled by crawlers cannot support a larger medical question answering system in quantity. On the other hand, there are many parts of the data that need to be cleaned, which makes entity extraction difficult. In the process of data processing, the rule-based method is mainly used to extract entities and relationships, with a large workload. In the future, more sources of data can be introduced to enrich the types of entities and relationships in the medical knowledge map. In the process of knowledge extraction, the method based on deep learning can be used to achieve automatic entity extraction and relationship extraction.

3.2 Considering the Speed of System Construction and the Number of Actual Users

Considering the speed of system construction and the number of actual users, the medical question answering system in this paper is built based on Python's lightweight Web service framework Flask. Although it can meet the current needs, in the future application scenarios facing more users and more needs, it is necessary to use a more mainstream and stable server framework to ensure the system's stress tolerance.

3.3 The Training Data Sources

Although the named entity recognition method and semantic similarity calculation method proposed in this paper have achieved good performance on several Chinese datasets, the training data sources are relatively single in general, the Chinese datasets used in the experiment are small in scale, and the types of models used for comparison are also small. In the future, comparative experiments should be conducted on higher quality datasets and more other types of models based on deep learning, To further verify the performance of the model.

3.4 The High Accuracy of the Question Answering System

The final question answering system has a high accuracy for the user's questions that only contain a single entity and relationship. However, if the user's questions are complex questions that contain multiple entities and multiple relationships, the answer selection mechanism adopted in this paper is difficult to achieve ideal results. Another classic complex scenario faced by the question answering system is the multi hop question answering scenario. If the user asks questions continuously and the question object in the latest question is hidden in the question sentence before the last sentence or even more, the question answering system in this paper does not have the ability to handle such questions. In the subsequent work, we can further study these complex question and answer scenarios.

References

[1] Zhen W, Zhang J, Feng J, et al. (2014). Knowledge Graph Embedding by Translating on Hyperplanes. National Conference on Artificial Intelligence. AAAI Press.

[2] Zhang L. (2002). Knowledge graph theory and structural parsing. university of twente.

[3] Hakkani-Tur D, Heck L , Tur G . (2013). Using a knowledge graph and query click logs for unsupervised learning of relation detection. IEEE International Conference on Acoustics. IEEE.

[4] Wang C , Ming G , He X , et al.(2015). Challenges in Chinese knowledge graph construction. IEEE International Conference on Data Engineering Workshops. IEEE.

[5] Zhang J . (2014). Knowledge Graph Embedding by Translating on Hyperplanes. AAAI - Association for the Advancement of Artificial Intelligence.

[6] Bao J , Duan N , Yan Z , et al. (2016).Constraint-Based Question Answering with Knowledge Graph. International Conference on Computational Linguistics. The COLING 2016 Organizing Committee.

[7] Sande V , Miel, Vocht D , et al.(2016). Triple Pattern Fragments: A low-cost knowledge graph interface for the Web. Journal of web semantics: Science, services and agents on the world wide web.

[8] Xiong W , Hoang T , Wang W Y . (2017). DeepPath: A Reinforcement Learning Method for Knowledge Graph Reasoning.

[9] He H , Balakrishnan A , Eric M , et al. (2017).Learning Symmetric Collaborative Dialogue Agents with Dynamic Knowledge Graph Embeddings.

[10] Li Z , Yang Q , Wei L L , et al. (2015).Application of the Intelligent Dispensing System in Outpatient Pharmacy in Our Hospital[J]. China Pharmacy.