

Research on the Application of BiLSTM-CRF Model in the Field of Chinese Modern History Question-and-Answer

Xuehe Zhuang^{1, a}, Yuanhui Yu^{1, b, *} and Suyu Lan^{1, c}

¹ School of Computer Engineering, Jimei University, Xiamen, Fujian, China

^a 202021125013@jmu.edu.cn, ^{b, *} 18350123715@163.com, ^c 202021123098@jmu.edu.cn

Abstract. Knowledge graph is the key technology of knowledge engineering in the era of big data. Using the powerful semantic understanding and knowledge organization ability of knowledge graph, it can be a better solution to the problems such as the disordered and over-wide coverage of knowledge related to modern Chinese history. The core of this paper is to use high-quality machine learning and deep learning algorithms with the support of big data knowledge graph to obtain the problem analysis result through natural language processing, and then match the problem analysis result with the question template to generate relevant query statements in the constructed knowledge graph to query relevant content through the knowledge graph rich semantic relations. The close relationship between the entities returns the most appropriate information for the user. The experimental results show that the designed question-and-answer system of modern Chinese history fills the gap in this field to a certain extent.

Keywords: Knowledge graph; Question answering system; Natural language processing.

1. Preface

In today's information age, knowledge graph, as an important knowledge representation method, has become a hot research topic in information retrieval, natural language processing, intelligent question answering and other fields. In the past few years, with the rapid development of big data and artificial intelligence technology, the research and application results of knowledge graph have become more and more abundant, covering many aspects such as entity link, relationship extraction, knowledge reasoning, question answering system and so on. For example, Li Weigang[1] et al. adopted a domain ontology-oriented question classification method and a structured semantic information extraction method to implement an aviation question answering system based on ontology knowledge base. Li Daiyi[2] et al. designed a military weapon question answering system based on knowledge graph, which has achieved good performance in domain question answering and has high application value. Du Zeyu[3] et al. built an intelligent question-and-answer system for e-commerce industry based on knowledge graph, which was deployed and used in real scenarios. According to the above research, it is found that the question answering system based on knowledge graph may have a variety of different question answering methods for the same type of questions, but still cannot be absolutely accurate in the question parsing level. At the same time, how to accurately transform the knowledge association between question and answer pairs into the query statement on the knowledge graph needs further research.

With the development of deep learning technology, more and more researchers try to apply deep learning technology to knowledge graph question answering system. In this paper, high quality deep learning technology framework is integrated with knowledge graph question answering system, and it is applied to the current knowledge graph involved in the shallow field of education. Under the current condition of science and technology and the gradual perfection of theoretical knowledge, it is feasible to construct a Q&A system of modern Chinese history based on knowledge graph by comprehensive use of developed natural language processing technology. Based on the knowledge data of modern Chinese history collected from multiple channels, this paper constructs a knowledge graph of modern Chinese history to provide data support for the question-answering system. Neo4j graph database is used to store and manage the knowledge graph, and BiLSTM-CRF deep learning technology is combined to process natural question statements. The questions are classified by

SVM algorithm, parsed into cypher query statements, and the relevant information is queried in the knowledge graph and returned to the user.

2. Construction of knowledge graph

2.1 Data acquisition and preprocessing

The quality of a knowledge map depends on the data quality and data scale of the knowledge map. For the knowledge map of modern Chinese history, how to obtain high-quality and large-scale data is a challenging problem. The data in this paper come from two sources, one is from CN-DBpedia. CN-DBpedia, developed and maintained by the Knowledge Atlas team of Fudan University, is the largest generalized structured encyclopedia knowledge atlas in the current Chinese field. In this system, the knowledge in the field of Chinese modern history is screened and extracted, and finally processed into relevant data. On the other hand, it obtains relevant knowledge manually in official textbooks and historical documents to achieve the authority of data.

2.2 Use Neo4j to store knowledge map of modern Chinese history

At present, there are three kinds of knowledge graph storage, the first is based on graph database storage, the second is based on RDF triplet storage, and the last is based on traditional relational database storage[4]. Due to the special dynamic change characteristics of knowledge graph, the entity relationship and content of knowledge graph need to be adjusted in time, and traditional relational databases can no longer adapt to it, and the retrieval efficiency is low under a large amount of knowledge. Therefore, there are only the first two knowledge graph storage methods chosen by the industry. Table 1 shows the differences between the two mainstream storage methods.

Table 1. The difference between the two main storage methods

Storage type	Form	Typical database of this type
based on RDF triplet storage	triplet	gStore,Blaze-Graph,RDF-3X
based on graph database storage	Entity, node, edge, attribute	Neo4j,Drgraph,Tiger-Drgraph

Graph database can display knowledge view simply and intuitively, can adapt to the dynamic change of knowledge graph, and has high query efficiency[5]. This paper chooses the storage method based on graph database, and stores the constructed knowledge graph in Neo4j platform, a typical representative of graph database. Figure 1 below is part of the display interface of the constructed knowledge graph of modern Chinese history in Neo4j platform.

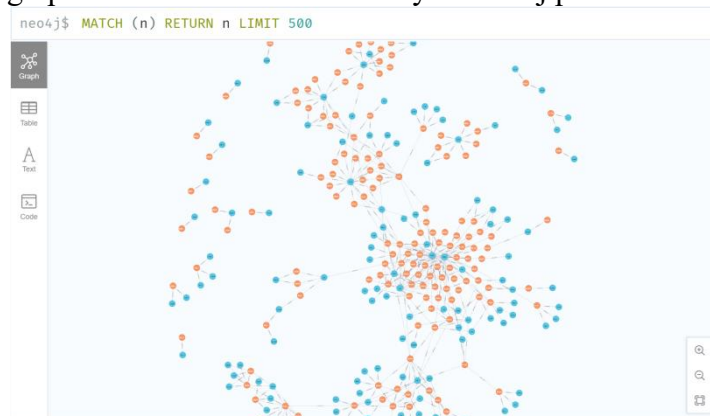


Figure. 1 Part of Chinese modern history knowledge map display

3. An applied study on the Q&A of modern Chinese history

3.1 Overall system architecture

The overall architecture of the modern Chinese history question-and-answer system based on knowledge graph constructed in this paper is shown in Figure 2 below. The system is mainly composed of three modules: knowledge graph construction module, question parsing module and front-end interactive interface module. Knowledge graph construction is to convert the data set of modern Chinese history into knowledge graph through data model and store it in Neo4j graph database. The front-end interface accepts questions raised by users and accesses the question parsing module. The questions are classified by natural language processing technology, and the question template is matched. Finally, the question template is converted into Cypher query statement to search for relevant information in the neo4j graph database. Finally, the found information is modified and returned to the front-end interface and fed back to the user.

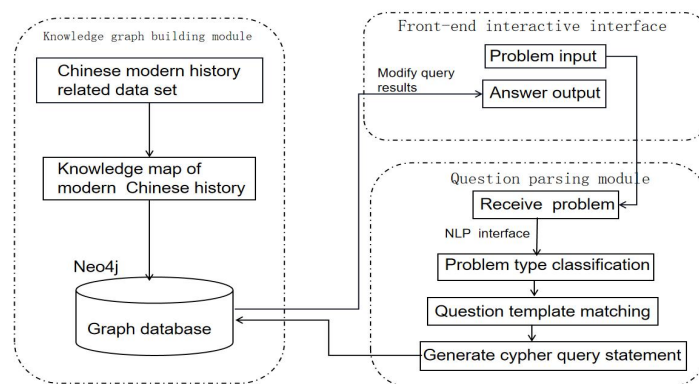


Figure. 2 The overall structure of the question-and-answer system of modern Chinese history

3.2 Preliminary filter question based on AC automata multi-mode matching algorithm

Among the multi-pattern matching algorithms, AC automata multi-pattern matching algorithm is the most classic algorithm, which is based on the improvement of traditional dictionary tree Trie. Through the feature corpus of modern Chinese history, the AC automaton algorithm is used to match whether there are feature words in the questions, and the candidate keywords are identified by actree tree, and the keywords are matched with the feature words in the corpus to achieve the purpose of preliminarily filtering the questions.

3.3 Processing natural language questions based on BiLSTM-CRF model

The mainstream named entity recognition in the industry can be divided into rule-based named entity recognition and named entity recognition based on machine learning and deep learning[6]. BiLSTM-CRF is a commonly used model for named entity recognition based on deep learning. With the development of theory and practice, researchers have improved and proposed many newer models on this basis, such as the enhanced Bert-BiLSTM-CRF Litchi Longan pest and disease model proposed by Zheng Yongzhi[7] et al. In this system, natural questions are inputted into BiLSTM feature extraction layer for training, and then inputted into CRF conditional random field layer for correction, and finally the best candidate words are obtained. The model process diagram is shown in Figure 3.

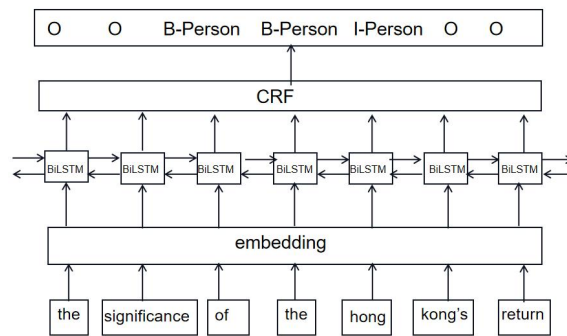


Figure. 3 BiLSTM-CRF model diagram

Based on BiLSTM-CRF model, precision rate, recall rate and F1 value are selected as evaluation indicators to evaluate a model, and the final model effect is shown in Table 2 below.

Table 2. BiLSTM-CRF model evaluated the effect

model	P	R	F1
BiLSTM-CRF	0.87	0.92	0.89

3.4 Question text classification based on SVM algorithm

The purpose of querying natural questions in modern Chinese history is clear. This system constructs SVM multi-classifier to classify natural questions. SVM algorithm has the characteristics of fast classification speed and high accuracy, which is more in line with this system. The system takes events as the main body and constructs a variety of question query templates related to events, as shown in Table 3 below. The SVM algorithm can classify questions according to feature words.

Table 3. Part of the problem type template

Question classification	Question template
	What impact did the Sino-Japanese War have on China
The impact or significance of an event	What was the impact of the Sino-Japanese War
	What was the result of the Sino-Japanese War
Description of events	A brief introduction to the Westernization Movement
The time of the event	When is the return of Hong Kong

3.5 System results display

Based on the above points, the basic logical structure and related technical structure of the system have been roughly explained. The following are the operating examples of part of the system, Figure4,Figure5,Figure6

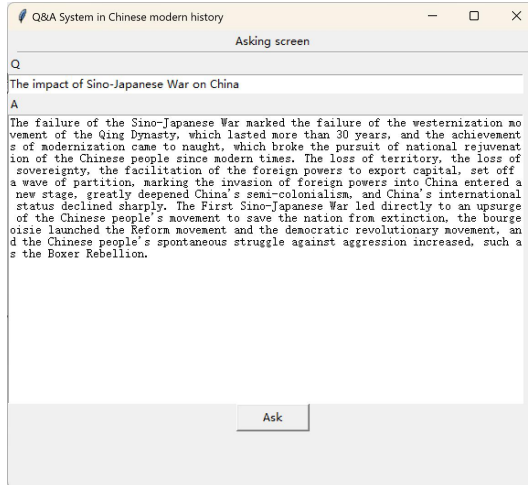


Figure. 4 System running instance 1

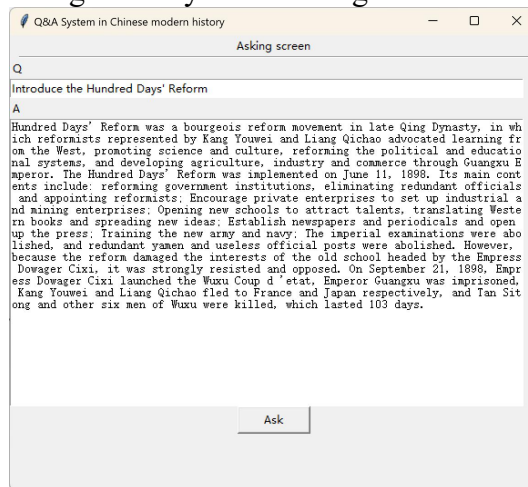


Figure. 5 System running instance 2

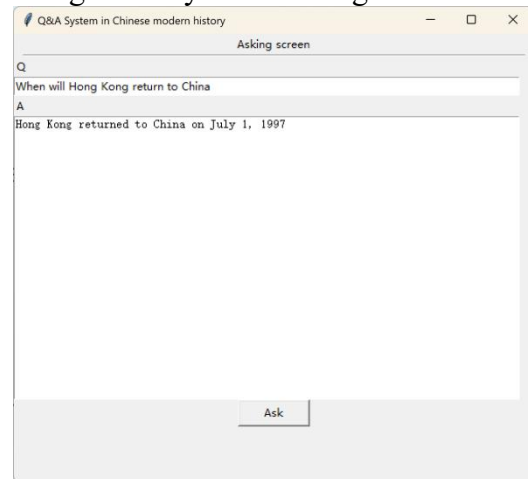


Figure. 6 System running instance 3

4. Conclusion and prospect

A This paper takes the relevant events in modern Chinese history as the starting point, uses multi-source heterogeneous data to build a knowledge map of modern Chinese history of a certain scale. On the basis of the knowledge map, the task of intelligent question-answering is realized through named entity recognition, question classification, question parsing, etc., which fills the field of education to a certain extent. The knowledge map of modern and contemporary Chinese history is vacant. Due to the complexity of knowledge in the field of modern Chinese history, it is difficult

to divide it, and the knowledge introduced in this paper cannot cover all the relevant knowledge of modern Chinese history. The knowledge map of modern Chinese history should be optimized by continuous iteration and increasing the richness between points. In the next research work, how to improve the intelligent semantic analysis and the quality of knowledge graph will be further explored.

Acknowledgment

This paper was supported by the Undergraduate Innovation Project of Jimei University (Project No. J023316) and Xiamen Science and Technology Subsidy Project (Project No. 2022CXY0301).

References

- [1] Weigang Li, Huilan Wang, Yong Li. Aviation question answering system based on ontology knowledge base. In: Proceedings of the 6th National Youth Computational Linguistics Conference. Shanghai, 2012, pp.018-222.
- [2] Daiyi Li, Jie Sheng, Yunxing Liu. Military weapon question answering system based on knowledge graph. 2020, 11(05):58-65. DOI:10.15908/j.cnki.cist.2020.05.011.
- [3] Zeyu Du, Yan Yang, Liang He. E-commerce question answering system based on Chinese knowledge graph. Computer applications and software, 2017, 34(05):153-159.
- [4] Qiao Liu, Yang Li, Hong Duan. Overview of knowledge graph construction techniques. Computer research and development, 2016, 53(03):582-600.
- [5] Zhen Yang, Weiqing Wang. Research and application of graph database. Computer programming skills and maintenance, 2020, (12):91-93. DOI:10.16184/j.cnki.comprg.2020.12.035.
- [6] Lishuang Li, Yuankai Guo. Biomedical named entity recognition based on CNN-BLSTM-CRF model. Chinese information journal, 2018, 32(01):116-122.
- [7] Yongzhi Zheng, Huilin Wu, Dingju Zhu. Question and answer system based on knowledge map of lychee and Longan pests. Computer and digital engineering, 2021, 49(12):2618-2622.