Constructing a Hybrid Automatic Q&A System Integrating Knowledge Graph and Information Retrieval Technologies

Yang Liu, Bin Xu, Yuji Yang, Tonglee Chung, Peng Zhang

Department of Computer Science and Technology, Tsinghua University, China
yliu16@mails.tsinghua.edu.cn, xubin@tsinghua.edu.cn,
{yangyujiyyj,tongleechung86,zpjumper}@gmail.com

Abstract. Question-answering (QA) system provides a friendly way for human-computer interaction, which has become an important research direction of smart learning. It provides an easy and individual way for the learner to acquire knowledge. This paper focuses on K-12 education and constructs a hybrid automatic question-answering system which integrates Knowledge Based Question Answering (KB-QA) and Information Retrieval-based Question Answering (IR-QA). The system is built based on Chinese textbooks and a Chinese K-12 knowledge graph (edukg.org). Our QA system covers 9 subjects in K-12 education field, including mathematics, Chinese, geography, history, etc. We evaluate our system on more than 9,000 questions, and achieve average accuracy over 70%. The system could provide effective assistance for teachers' teaching and students' learning.

Keywords: Smart Education. Knowledge Base. Question Answering

1 Introduction

Smart education [1] has attracted a large amount of attention in research. The essence of smart education is to build a smart learning environment with technology, where students can acquire knowledge and solve problems faster and better. The automatic question answering (QA) system is a very effective way which can help students get answers in a timely manner during the daily learning process. So far the existing QA systems are not designed for K-12 education. Therefore, the challenge is to build a question answering system specifically for K-12 education, which can accurately understand the students' questions and give precise answers quickly.

The early stage QA system [2] was a template-based expert system that was designed to manually construct rules for specific domains. The disadvantage is that it can only process a small amount of data in a specific field. With the search technique developing, the information retrieval question answering (IR-QA) [3]

M. Chang et al. (eds.), *Foundations and Trends in Smart Learning*,
Lecture Notes in Educational Technology, https://doi.org/10.1007/978-981-13-6908-7_9

appeared. It extracts the answers in a large number of texts based on the keywords and semantic relationships, e.g., IBM Watson [4], TREC, etc. These QA systems solved the previous narrow coverage limitation to a certain extent, but due to the uneven quality of the text, the accuracy was unsatisfying at the same time, the Internet QA community gradually emerged such as Yahoo Answers, Stack Overflow, etc. These QA communities only provide an aggregation platform for users, the correctness of the answers depend on the user's own judgment.

"Knowledge Graph" [5] was proposed by Google to define the structure of knowledge from a new perspective. It attempts to transform unstructured data into structured data from the data itself and connect the various data together to form a graph model. This structured graph data provides a new direction for the development of the question answering system, namely the knowledge graph based question answering system (KB-QA), which can provide users with the structural data in the knowledge base. Since KB-QA can provide very simple and precise answers, it gradually becomes an important research direction of the question answering system. KB-QA can also effectively help with the development of "next generation intelligent retrieval" and "humanoid robot".

The core idea of KB-QA is to semantically analyze the question statement and transform it into a structured query language, then get the final answer by querying and reasoning the knowledge graph. Berant J [6] uses semantic analysis methods, such as Combination Category Grammar (CCG) and Dependent Combination Semantics (DCS) [7,8,9] to directly obtain the semantic representation of the question to query the knowledge base, and then get the answer. These semantic parsing methods are not good for questions containing multiple instances and multiple relationships, which limits the system's performance. Bast H [10] classifies questions according to the number of entities involved in the question, and formulates three corresponding templates, and then queries the knowledge graph to get the answer. The advantage of this method is that it has a certain reasoning ability from the question. Given multiple entities, it is easy to find the common relationships and nodes between the entities. However, this requires the connectivity and data quality of the knowledge graph to be very high.

So far, the main challenges of KB-QA are as follows: 1. the semantic analysis caused by the diversity of natural language problem is difficult. For example, "Who is the author of Hamlet?" and "Who wrote Hamlet?" have the same meaning. 2. The accuracy of entity recognition in the domain affects the accuracy of KB-QA. However, KB-QA has the advantage of high precision, while the recall is fairly low.

The IR-QA system depends on a reliable retrieval strategy. Its core technology is information extraction and matching. IR-QA contains three steps normally. 1. Extract keywords from the question. 2. Retrieve relevant documents from the corpus. 3. Filter requested information from the documents. The advantage of IR-QA is that it can retrieve the relevant documents that most likely contain the answer. The recall of IR-QA is fairly high, while the precision is fairly low.

This paper proposed a hybrid question answering system in the K-12 education that combines KB-QA and IR-QA [11, 12], taking advantages of both. We focus on the factoid questions which can be directly answered by one entity or one property in the knowledge graph. For example, "Where is Stephen Zweig from?", we can get the answer, "Germany". Our contributions in this paper can be summarized as follows:

- Proposed an automated question answering system which can efficiently answer questions in the K-12 education field.
- Combine KB-QA with IR-QA to improve the accuracy and coverage of answers.
- Full coverage of 9 subjects in the field of K-12 education, including Chinese, math, English, history, geography, biology, physics, chemistry, politics.

2 Related Works

Recent years, the question answering system has made great progress in the field of smart education. A wide variety of systems and related hardware devices are emerging. The emergence of early education robots and smart speakers provides a new way for children to learn knowledge through entertainment; exam answering robots are constantly challenging the limits of human examinations. These robots or applications are inseparable from a QA system that supports them in the background. The robots and AI devices based on automatic QA have effectively supplemented the traditional one-to-many education model that has not been overcome in the field of smart education, and opened up a one-to-one education model for human-computer interaction.

At the heart of IBM's Watson robot [4] is an automated question answering system based on machine learning. The robot has already defeated human players in the television quiz show "Jeopardy!" In the field of smart education, IBM launched the Watson Education Classroom to provide intelligent solutions for teachers and students.

In China, the AI-MATHS [13] and Aidam [13] robot challenged the Chinese college entrance examination and scored 105 points and 134 points in the mathematics volume (out of 150); Todai Robot [14] from the University of Tokyo in Japan has been able to pass the local university entrance examination, and exceeded the average score in subjects such as mathematics, physics, and English.

In recent years, Deep Learning (DL) is rapidly developing. Many researchers try to build KB-QA system using deep learning model. Dong [25] used a convolutional neural networks (CNN) to understand the question. Ming [26] structured an end-to-end KB-QA model based on a Long Short-Term Memory (LSTM) model. Salman [27] focused on the factoid question and divided the task into entity detection, entity linking and relation prediction, on this basis they build an LSTM+RNN model. These models get better results than the traditional methods, but they are lack of interpretation and controllability.

Yet all these methods are not focused on factoid question answering for K-12 education. But these questions' relevant knowledge is important for K-12 students, while we combined the advantages of KB-QA and IR-QA and tried to build a hybrid factoid question answering in K-12 education.

3 System Structure

In this section, we will introduce the structure of our automatic question answering system. For QA system in K-12 education [15] field, accuracy is the most important prerequisite. The system integrates template matching [16], similarity calculation [17] and text retrieval [18], it tries to cover as many concepts as possible in each subject. We developed the QA system in Java and deployed it on a Tomcat server. The query language that we search knowledge graph is SPARQL. The architecture of the system is shown in Figure 1.

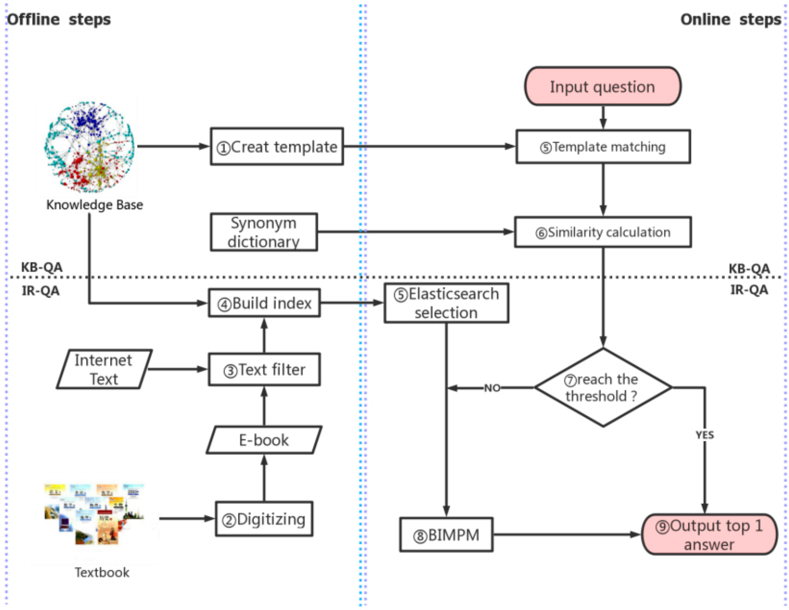

Fig. 1. The Question Answering System Structure

3.1 Offline Steps

In offline steps, we formulate the feature rules of questions manually, and generate templates by the computer automatically. A number of high-quality question templates for K-12 education are constructed as the basis of the KB-QA module. We digitized more than 1,300 physical textbooks, and acquired more than 10,000 electronic books from digital libraries. Also, we filter and clean the amount of text data on Internet. Then we got a large amount of high-quality text data, these data can be used for the IR-QA module.

Templates are very important for the whole system. They have a certain influence on entity and predicate recognition. Therefore, how to construct a high-quality template generation system is very important. In our system, the templates are constructed using regular expressions. Each template has a plurality of fields, such as attributes and priorities, corresponding to the specific regular expression template, e.g., "(?<title>(.*)?)作者(.*)?". The Chinese word "作者" means "author" in English. This template is to search the author of a book or an article. A detailed look of the template's regular expression is shown in Table 1.

Table 1. Structure of the templates

Field	Example	Type
content	(?<title>(.*)?)地理位置(.*)?	varchar(200)
subject	true	tinyint
value	false	tinyint
type	diliweizhi	varchar(100)
class	null	varchar(50)
usage	data	varchar(20)
priority	1	int

- **Content**: The content row is the template content which is written in regular expressions. For example, if the question matches the template, it will recognize "地理位置" as a possible predicate of a question, while the Chinese word"地理位置" means "geographical location". "(?<title>(.*)?)" is a named group capture that was used to determine the location of the subject. For example, "Where is the geographical location of Mount Tai", the template got the subject is "Mount Tai".
- **Subject/Value/Type:** This field indicates whether the template subject is determined, the default flag is true. If the subject of the question is unknown, it will be marked as false. Such as "who is called Tian Khan", "value" indicates whether the object is certain or not, "type" indicates the label of the template.
- **Class**: It indicates the class of the subject, it was used to qualify the subject of certain special questions and can be ignored.
- **Usage**: When questions can't be answered through a simple SPARQL query, "usage" is used to identify these questions. It is optional.
- **Priority:** This field indicates the priority of the template. It is mainly used to calculate the score of the predicate.

3.2 Online Steps

The answering process of the problem is completed by KB-QA and IR-QA. The implementation of each part will be introduced separately below.

KB-QA. KB-QA includes the following five sub-modules:

- **Entity recognition [19] and entity linking [20] module:** In this module, we adopt template segmentation, synonym word forest query, similarity calculation, longest common substring matching, etc., and set the priority according to the confidence of each method. Then we build a set of candidate entities.
- **Predicate and relationship identification module:** This module uses template matching, similarity calculation, etc. The method also sets the priority according to the method to obtain a set of candidate Predicate.
- **Query statement generation and query module** [21]**:** According to the candidate sets, this module generates a SPARQL [22] query statement, and query the corresponding result from the knowledge base, then add to the candidate answer set. For example, "Who is the author of Hamlet?", the subject is "Hamlet" and the predicate is "author". We get a SPARQL query statement like this:

```
SELECT distinct ?subject ?author
from<http://edukb.org/Chinese>
WHERE {
?subject rdfs:label ' Hamlet '.
?property rdfs:label 'author'.
?subject ?property ?author.
}
```

- **Answer screening module:** According to the priority of the entity and the predicate, we calculate the score of each candidate answer. Then, we choose the top1 as the answer.

IR-QA. IR-QA will be used when KB-QA cannot provide a credible answer. We will select a text sentence from the text corpus. IR-QA is composed of the following two parts:

- **ElasticSearch [18] rough selection:** (1) We use the rules to special treatment of the question, find the most advanced key such as "highest", "third" and other words. (2) Splitting the question into words, and assigning different vocabulary weights according to the established part of speech priority. (3) Using the combined query and similarity matching strategy to query the selected keywords.
- **BiMPM [23] Featured:** Use the BIMPM model to further screen ElasticSearch's rough selection results, then choose the top1 answer.

4 Experiments

The system is an automatic QA system based on the K-12 education knowledge base and a large number of electronic texts. The K-12 education knowledge base [24] contains more than 22 million triples, 1.62 million instances, 1,000 concepts, and 4,000 attributes. The source of knowledge includes the annotation library and the external source library. The annotation library is obtained from the knowledge which is tagged in the textbook, the external source library is extracted from the encyclopedia and Internet data. These textbooks are provided by The National Library of China for research. It covers all knowledge points of the 9 subjects in K-12 education. The electronic text mainly includes 1,300 elementary education textbooks and 10,011 electronic extracurricular reading materials.

4.1 Test Dataset Generation

In the preliminary preparation work, a great number of test questions was obtained from the teacher resource books, the Internet and the textbooks. All of these questions have exact answers. The questions types mainly are fill-in-the-blank questions, multiple-choice questions, reading comprehension questions, essay questions, etc. These questions cannot be directly parsed by the KB-QA system. So we need to convert them to the questions that our system can analyze. For example, “The ratio of land to sea in the world is about ( )” is converted to “What is the ratio of land to sea in the world?”.

Accuracy is the main evaluation index. We test each subject’s question dataset separately. When we input the test cases, answers are recorded. The test cases are designed for each subject. The subjects include Chinese, math, English, physics, chemistry, history, geography, biology, and politics. A total of 9,020 test cases were written by experts. The details of the questions in each subject are shown in Table 2.

Table 2. Quantity of test cases under each subject

Subject	Number of test cases	Percent of all test cases (%)
Chinese	1,007	11%
math	926	10%
English	1,033	12%
physics	1,000	11%
chemistry	1,001	11%
history	1,040	12%
geography	1,017	11%
biology	1,000	11%
Politics	996	11%
Total	**9,020**	**100%**

4.2 Experimental Result

Totally 9,020 test cases were executed, and the result of each test case was recorded in detail. Test case execution statistics and results are shown in Table 3 and Table 4. Table 3 shows the accuracy of each subject. Mathematics got the highest accuracy, because most of its subject knowledge are described precisely. Geography got the lowest accuracy, because a variety of geographical questions need to ratiocinate or summarize the answer from attribute value or text data. For example, "What does crustal movement lead to?". The expansion of the knowledge graph and text data is helpful to increase the accuracy.

Table 3. Accuracy of Our system on test cases

Subject	Number of test cases	Correct	Wrong	Accuracy(%)
Chinese	1,007	787	220	78.15%
math	926	862	64	93.09%
English	1,033	887	146	85.87%
physics	1,000	911	89	88.40%
chemistry	1,001	897	104	89.61%
history	1,040	904	136	83.17%
geography	1,017	739	278	72.66%
biology	1,000	860	140	85.50%
politics	996	885	111	88.86%
Total	**9,020**	**7,732**	**1,288**	**85.72%**

Table 4. Example test cases

No	Subject	Question	Answer
1	**Chinese**		
1.1	Who is the author of "Shi Ji"		Sima Qian
1.2	What are the elements of the message?		Person, time, place, cause, event
1.3	Where is Stephen Zweig from?		Germany
2	**Physics**		
2.1	How to calculate the active power?		P=W/t
2.2	What is the function of infrared ray?		heat energy
2.3	What is the applied object of gravity?		earth
3	**Geography**		
3.1	What temperature zone is Europe in?		North Cold Zone, North Temperate Zone
3.2	What's the climate in eastern China?		monsoon climate
3.3	What are the three elements of a map?		Scale, direction, legend

5 Conclusion

With the advancement of artificial intelligence, smart education is developing towards intelligence and individualization. The question answering system is a new hotspot in this field. We construct an automatic question answering system that integrates KB-QA and IR-QA, combines traditional template matching with semantic calculation, and achieves fairly good accuracy. The system can help many primary and middle school students who lack counseling resources, they can get answers in a timely manner during their daily study. At the same time, it can be integrated with various educational hardware or mobile applications to improve their ability to provide automatic QA for students. It has positive significance to construct the smart education system in the Internet era.

6 Acknowledgement

This work is partly supported by National Engineering Laboratory for Cyberlearning and Intelligent Technology. Beijing Key Lab of Networked Multimedia also supports our research work.

References

[1] Zhu, Z.T., He,B.: 智慧教育:教育信息化的新境界(Smart education: The new realm of education informatization). E-education Research. 12, 7-15(2012)
[2] Zadeh, L.A.: The role of fuzzy logic in the management of uncertainty in expert systems. Fuzzy Sets & Systems. 11(1), 197-198(1983)
[3] Andrenucci, A., Sneiders, E.: Automated question answering: review of the main approaches. In: International Conference on Information Technology and Applications, IEEE,pp. 514-519 (2005)
[4] High, Rob.: The era of cognitive systems: An inside look at IBM Watson and how it works.IBM Corporation, Redbooks(2012).
[5] Knowledge Graph, https://en.wikipedia.org/wiki/Knowledge_Graph
[6] Berant, Jonathan.: Semantic parsing on freebase from question-answer pairs. Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing(2013)
[7] Xu, K.: Question answering via phrasal semantic parsing. In: International Conference of the Cross-Language Evaluation Forum for European Languages. Springer, Cham(2015)
[8] Yih, Scott.Wen-tau.:Semantic parsing via staged query graph generation: Question answering with knowledge base (2015)
[9] Xu. K.: Question answering on freebase via relation extraction and textual evidence. arXiv preprint arXiv:1603.00957 (2016)
[10] Bast. H, Haussmann. E.: More accurate question answering on freebase. Proceedings of the 24th ACM International on Conference on Information and Knowledge Management, ACM (2015).
[11] Bordes. A, Chopra. S, Weston. J.: Question answering with subgraph embeddings. arXiv preprint arXiv:1406.3676 (2014)

[12] Yih, Scott Wen-tau, et al. Semantic parsing via staged query graph generation: Question answering with knowledge base. (2015)
[13] What is “AI-MATHS” and “Aidam”, http://www.caigou.com.cn/news/2017061496.shtml
[14] Robot beat 80% of students on University of Tokyo entrance exam, https://www.businessinsider.com/robot-beat-most-students-on-university-tokyo-entrance-exam-2017-9
[15] What is K-12? , https://whatis.techtarget.com/definition/K-12
[16] Abujabal. A, Yahya. M, Riedewald. M.: Automated template generation for question answering over knowledge graphs. Proceedings of the 26th international conference on world wide web. International World Wide Web Conferences Steering Committee (2017)
[17] Tous. R, Delgado. J.: A vector space model for semantic similarity calculation and OWL ontology alignment. In: International Conference on Database and Expert Systems Applications, pp.207-216. Springer, Berlin, Heidelberg(2006)
[18] Gormley.C, Zachary.T.: Elasticsearch: The Definitive Guide: A Distributed Real-Time Search and Analytics Engine. O’Reilly Media, Inc(2015)
[19] Mohit.B.: Named entity recognition. In: Natural language processing of semitic languages, pp: 221-245. Springer, Berlin, Heidelberg(2014)
[20] Shen.W, W.J.Y.: Entity linking with a knowledge base: Issues, techniques, and solutions. IEEE Transactions on Knowledge and Data Engineering(2015)
[21] Resource Description Framework (RDF), https://www.w3.org/RDF/
[22] SPARQL Tutorial, https://jena.apache.org/tutorials/sparql.html
[23] Wang, Z.G., W. Hamza., Radu. F.: Bilateral multi-perspective matching for natural language sentences. arXiv preprint arXiv:1702.03814 (2017)
[24] K-12 education knowledge graph, http://www.edukg.org
[25] Dong, L., et al. Question Answering over Freebase with Multi-Column Convolutional Neural Networks. In: Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing, vol. 1, pp. 260-269 (2015)
[26] Tan, M., Cicero, D.: Lstm-Based Deep Learning Models For Nonfactoid answer Selection. In: Computer Science(2015)
[27] Mohammed, Salman, Peng. S, Jimmy L.: Strong Baselines for Simple Question Answering over Knowledge Graphs with and without Neural Networks. arXiv: 1712.01969 (2017).