A Question Understanding Model Based on Knowledge Points for Chinese Question Answering Service in E-Learning

Zheng-Hong Wu¹, Ming Li¹, and Huamin Feng²

¹ School of Information Science & Technology, East China Normal University, Shanghai 200062, P.R. China

wzh_hongzheng@hotmail.com, mli@ee.ecnu.edu.cn,
ming_lihk@yahoo.com

² Key Laboratory of Security and Secrecy of Information, Beijing Electronic Science and Technology Institute, Beijing 100070, P.R. China
fenghm@besti.edu.cn

Abstract. Question answering service in e-learning environment is an important issue. Chinese semantic analysis is a key bottleneck for question answering service to understand question's content. This paper proposes a question understanding model to understand and process syntactic and semantic structure. In this paper, we analyzed a lot of questions from students, and clustered questions based on knowledge points. The question understanding model is made to get question focus and question type. According to question focus and question type, the question answering service can precisely know the question's answer by locating knowledge point's attribute. This method could more perfectly understand semantic content of questions than using pure Chinese semantic analysis. It is very useful for students to study in a self-learning environment.

Keywords: Question Answering, Knowledge Point, E-Learning.

1 Introduction

Nowadays, question answering (Q&A) has become an important research field with the rapid development of the Internet. However, current Q&A systems are far below human expectations. This is because natural language is not easy to process and interactive Q&A is even harder to model. Q&A research is a multidisciplinary field, and systems are usually integrated from many techniques and resources. Information Retrieval (IR), Natural Language Processing (NLP), Information Extraction (IE), Machine Learning, and even Software Engineering techniques are all needed to build a Q&A system. Most Q&A system would be an ideal way for people to search for information. Since 1999, many international question answering contests have been held, such as TREC (Ellen M. Voorhees 2003, 2004, 2005), CLEF, and NTCIR. Results from the TREC evaluations ((Kwok et al., 2000) (Radev et al., 2000) (Allen et al., 2000)) showed that Information Retrieval (IR) techniques alone are not sufficient

for finding answers with high precision. The Webclopedia project at the USC Information Sciences Institute (Hovy 2000, 2001) pursued a semantics-based approach to answer pinpointing that relies heavily on parsing. They want to provide much more accurate answers for users. Some researchers were working on designing the Q&A methodologies, such as Debra T. Burhans (2005), Dan Moldovan (2005), and Humphreys (2000), etc. Some researchers researched the question stem, e.g. who, where or how much and eventually one of the question concepts. But the stem is ambiguous (for example what), as described in (Harabagiu et al., 2000) (Radev et al., 2000) (Srihari and Li, 2000). There are only a handful research teams working on Chinese QA, which may explain why the first Chinese QA contest was only held at this year's NTCIR workshop (2005). New Chinese question taxonomy respectively from approach and semantic viewpoints was presented (Youzheng Wu et al.). Hybrid architecture for answering Chinese factoid questions from news documents was proposed (Cheng-Wei Lee et al.).

Frequently, question reformulations use different words, but imply the same answer. Moreover, many equivalent answers are phrased differently. That is the main trouble in solving Q&A. Current Q&A systems are far below human expectations. This is because natural language is not easy to process and interactive Q&A is even harder to model. Since natural language is full of noise, it is important to deal with real data. Noisy or inconsistent information could be filtered out to achieve better performance. Q&A also has the same features in e-learning. And Q&A is an important part in e-learning for students. But, Chinese semantic analysis is a key bottleneck in Q&A field.

How to solve these difficulties is very important for Q&A system. Firstly, we collected a lot of questions from students and textbooks. Then, we analyzed these question samples. We found that a question from students in e-learning course always asks content about some knowledge point. Moreover, each question's answer is connected with one attribute of knowledge point. Knowledge is composed by knowledge points. Attributes of knowledge point are the followings: knowledge name, definition, characteristic, function, and so on. The same attribute can reformulate question use different words, namely different question types, but simply the same answer. So, questions have some special features from students in e-learning. Every question has two important features. One is that a question is almost related with some knowledge points. This feature is named question focus. The other is that the question's answer is some attributes of one or more knowledge points. This second feature is named question type or knowledge attribute. At the same time, we found another key feature. Though different questions, they are always related with one or more knowledge points. And their answers are these knowledge points' attributes. So questions can be classified into simple question and complex question based on the number of knowledge points involved in one question sentence. If the question focus relates to one knowledge point, it is called simple question. And its answer can be found from an attribute of the knowledge point. Complex question is defined that its question focus is connected with several knowledge points. Complex question is divided into several kinds according to the relationship of knowledge points (Zheng-Hong Wu et al., 2004). Then, Knowledge base is constructed by ontology. Each

attribute of knowledge point can be found by linking this knowledge point. If these two features are determined, the answer can be retrieved from knowledge base.

Combining early research (Zheng-Hong Wu et al., 2005), we modeled question item from two novel angels based on knowledge points. One is question focus. The other is knowledge attribute. Two novel angels are helpful in locating answer in this paper. Section 2 gives the overview of the architecture of Q&A service. According to Neural Network Classifier (NNC) and knowledge base, Section 3 describes process of generating question understanding model including question item and question panel, and gives the experiment system. Finally, Section 4 makes a simple conclusion and describes some possible future works.

2 Question Answering Service

Figure 1 shows how the question answering service works. The first part is inputting questions by interface. This part classifies question type and question focus through question item. Question item is constructed by Neural Network Classifier (NNC). The second part is setting up knowledge point base, synonymy base and answer base. And, the third part is mining answer.

Fig. 1. Question Answering Service

In the beginning, the service retrieves question from the interface. And, question focus and question type could be created (step I) according to the question item. Then, the service uses data mining to locate the knowledge point in knowledge base (step II). Finally, the service is to generate answer (step III). Students then would be able to see the answer item by simply using their web browser.

3 Operation Flow of Generating Question Panel

This section goes through the question item by clustering abundant question examples. From the knowledge point, Table 1 details question focus and question type. There are three sub-sections in this section, and each sub-section represents a key for different parts in the service architecture as Figure 2 shows. The three keys are the knowledge base (step I-2), question item modeling (step I-3) and question panel (step I-4).

Fig. 2. Flow of Question Panel

Figure 2 shows the flow of generating question panel. The architecture of question panel could be seen as four parts: the first part is sampling questions from students and textbooks (step I-1); the second part is setting up synonymy base (step I-2); and, the third part is to model question item by classing question type and question focus on the base of the synonymy base and Neural Network Classifier (step I-3); the fourth part is generating question panel (step I-4). The repeated flow indicates the question item would be changed from time to time. Finally question panel is made by question item. Students then would be able to see the question panel by simply using their web browser.

3.1 Modeling Question Item

In the beginning, we collected thousands of questions from students and textbooks. Then, we concluded that student inputting a question is to know some content about one or some knowledge points, which is the attribute of knowledge point. Additionally,

a question is composed by question focus and question type. The question focus is knowledge point. The question type is to know attribute of one or more knowledge points. Finally, we found that questions could be modeled according to the number of knowledge points. Therefore, we can classify questions from two angles. One is the number of knowledge. The other is the attribute involved in questions.

Table 1. Question Attribute Clustering and Statistics

<u>Q.C.</u>	<u>Q.T. (K.A.)</u>	<u>Example of question</u>	<u>Q.F. (K.P.)</u>
Simple	What...(Definition/Concept...)	What (<i>definition</i>) is <u>optical fiber communication</u> ?	<u>Optical fiber communication</u>
	What/How...(Function/Use/Meaning/...)	What is the meaning of developing intelligent network?	<u>intelligent network</u>
	How/What...(Mechanism/Theory/...)	How (<i>mechanism</i>) to avoid <u>disturber from the same frequency</u> ?	<u>disturber from the same frequency</u>
	How/What...(Application...)	Describe the method that realizes the <u>global communication using synchronous satellite</u> .	<u>global communication using synchronous satellite</u>
	What/How...(Feature/...)	What are the features of <u>satellite communications</u> ?	<u>satellite communications</u>
	What/How...(Structure/Component/Kind...)	What components compose the <u>optical fiber communication</u> ?	<u>optical fiber communication</u>
	...	...	...
Complex	What/How...(Difference/Comparison/Relating...)	What is the difference between <u>transmission</u> and <u>convection</u> ?	<u>Transmission, convection</u>
	What/How...(Reason/Influence/...)	What special demand for <u>modem</u> in <u>mobile communication</u> ? What is the influence from <u>waste</u> and <u>chromatic dispersion of optical fiber</u> in <u>optical fiber communication system</u> ?	<u>Modem, mobile communication waste, chromatic dispersion of optical fiber, optical fiber communication system</u>
	...	...	...

Q.C.: Question Class; Q.T.: Question Type; Q.A.: Question Attribute; Q.F.: Question Focus; K.P.: Knowledge Point;

Taking questions in Modern Communication Technology for example, Table 1 shows the question type and question focus. The model of question item could be seen as two parts: the first part is simple question (from Q.C.) which involves one knowledge point (from Q.F. (K.P.)); the second part is complex question (from Q.C.) which involves two or more knowledge points (from Q.F. (K.P.)).

As we can see within Table 1, any question is about some attribute of knowledge points. One attribute may be expressed by similar words. For instance, mechanism, theory and other synonymy standard the same attribute for knowledge point, e.g.

Q1: How does the signal microwave communication work?

Q2: What is the working mechanism/theory of microwave communication?

Q1 and Q2 use different words, but imply the same answer, namely the same attribute of knowledge point.

So, synonymy words are ranged (Table 1) to express the same attribute after studying those question examples. Using sampling more questions, more synonymy words could be found by clustering. And more exact attribute kinds could be set up.

Simple question is focused on one knowledge point's attribute. These are several attributes in a knowledge point, including concept, function, feature, and others. The answer of question mainly aims at these attributes (from Q.T. (K.A.)).

Complex question's answer focused on the attribute among several knowledge points regarding one knowledge point as prerequisite. Complex question can be classified into two main sub-classes. One is the comparison of these knowledge points to know the difference or respective strongpoint. The other is that which influences one knowledge point to other knowledge points, or how does one knowledge point function at others.

According to the above features, NNC is introduced in this paper to set up question item. NNC implemented through neural network method. During a learning session, each example of the question, representing a single question feature vector, must be pre-classified by a human expert. The NNC processes training examples and induces rules for extracting question features from further unclassified examples, to be used by the question item. According to the synonymy base, NNC can discriminate question focus and question type from questions for the question item.

3.2 Knowledge Base

Knowledge base is built based on ontology. It includes knowledge point base, synonymy base and answer base. Knowledge point base is designed according to question's focus. Synonymy base is designed according to question's type. Answer base is designed by knowledge attribute.

First, the question is put. Use logic calculation to get question item. Question focus and question type are determined. According to question focus, the service could know exact knowledge point. Then the knowledge attribute is determined by question type. Finally, the service matches the answer by pitching the attributes of knowledge point.

3.3 Question Understanding Model

Based on question item and knowledge base, we set up a question panel using Visual Studio.NET to implement the question answering service. Figure 3 shows the question panel. When a student logs on and asks a question, the service can present this question panel for the student.

In this question panel, if you click knowledge attribute, a new windows will tell you the examples of attribute, such as definition, mechanism etc. If the question includes several knowledge points, you should input underline among several knowledge point names. For instance, the question: What (definition) is optical fiber communication? This question is generated in Figure 3. Through the panel, the service can precisely know his/her question's content to avoid trouble from semantic analysis using Natural Language Understanding. This question panel is very simple and effective for students in Q&A.

Fig. 3. Question Panel for Students

4 Conclusions and Future Directions

This paper focuses on setting up a question understanding model to support question answering for self-learning in e-learning course. To reach the goal, we collected a lot of questions and analyzed them from the number of knowledge point and attribute of that. Questions were clustered by using NNC. Subsequently, question panel was built to get question focus and question type. Then, knowledge base was set up. Finally, the question could be precisely understood through question understanding model. The service could provide the student answer. This model was tested in the network college in CCNU. This method has some efficiency in Q&A, especially in simple question. But this model should be improved in deal with complex question.

Question answering service becomes more and more important in e-learning field. We should have a more accurate method to reach accurately and adaptively present answer. Furthermore, because the clustering attributes should be more precise, we have to improve the method of clustering question type, knowledge attribute and synonymy base. Besides, we also need to establish a perfect and friendly question model.

Acknowledgements. This work was supported in part by the National Natural Science Foundation of China under the project grant numbers 60573125 and 60672114, and by the Key Laboratory of Security and Secrecy of Information, Beijing Electronic Science and Technology Institute under the project number KYKF 200606 of the open founding.

References

1. Voorhees, E. M.: Overview of the TREC 2005 Question Answering Track. Proc., the Thirteenth Text Retrieval Conference (TREC 2005), 2006
2. Voorhees, E. M.: Overview of the TREC 2004 Question Answering Track. Proc., the Thirteenth Text Retrieval Conference (TREC 2004), (2005) 52–62
3. Voorhees, E. M.: Overview of the TREC 2003 Question Answering Track. Proc., the Twelfth Text Retrieval Conference (TREC 2003), (2004) 54–68
4. Burhans, D. T., Shapiro, S. C.: Defining Answer Classes using Resolution Refutation. Special Issue of the Journal of Applied Logic on Questions and Answers: Theoretical and Applied Perspectives, Available online 18 Jan. 2006, (2006) 279–295
5. Shapiro, S. C.: A Logic of Arbitrary and Indefinite Objects, in: Principles of Knowledge Representation and Reasoning. Proc., the Ninth International Conference (KR-2004), AAAI Press, (2004) 565–575
6. Fleischman, M., Hovy, E., Echihiabi, A.: Offline Strategies for Online Question Answering: Answering Questions Before They Are Asked. Proc., 41st Annual Meeting of the Association for Computational Linguistics, (2003) 1–7
7. Ferret, O., Grau, B., Hurault-Plantet, M., Illouz, G., Jacquemin, C., Masson, N., Lecuyer, P.: QALC- the Question-Answering System of LIMSI-CNRS. Proc., the Text Retrieval Conference (TREC-9), (2000) 316–326
8. Moldovan, D., Rus, V.: Logic Form Transformation and its Applicability to Question Answering. Proc., ACL 2001
9. Moldovan, D., Pasca, M., Harabagiu, S., Surdeanu, M.: Performance Issues and Error Analysis in an Open-Domain Question Answering System. Proc., ACL 2002
10. Wu, Y.-W., Wu, Z.-H., et al.: Generating Personalized Answer by Constructing Question Situation. Springer LNAI, (2006) 731–739
11. Wu, Y.-W., Wu, Z.-H., et al.: Personalized Intelligent Question Answering Algorithm in E-learning. The Fourth International Conference on Machine Learning & Cybernetics (ICMLC-2005) 3299–3303
12. Wu, Y.-Z., Zhao, J., Xu, B.: Chinese Question Classification from Approach and Semantic Views. Information retrieval technology (Second Asia Information Retrieval Symposium, AIRS 2005, Jeju Island, Korea, Oct. 2005, proceedings) Asia Information Retrieval Symposium No2, (2005) vol. 3689, 485–490
13. Lee, C.-W., et al.: Perspectives on Chinese Question Answering Systems. Proc., the Workshop on the Sciences of the Artificial (WSA 2005), Taiwan, (2005)
14. Lee, C.-W., et al.: ASQA: A Hybrid Architecture for Answering Chinese Factoid Questions. http://www.iis.sinica.edu.tw/~thtsai/paper/Rocling2005_Demo.pdf