Question Type Analysis for Question-Answering Applications in Education

Yang-Yin Lee[a], Chih-Chieh Shao[b], Yen-Pin Chiu[b], Yong-Siang Shih[a], Hsin-Hsi Chen[a,b*], Chieh-Jen Wang[c] and Sen-Chia Chang[c]
[a]*Department of Computer Science and Information Engineering,*
[b]*Graduate Institute of Networking and Multimedia,*
National Taiwan University, Taipei, Taiwan
[c]*Intelligent Analytic Technology Division, Computational Intelligence Technology Center,*
Industrial Technology Research Institute, Hsinchu, Taiwan
*hhchen@ntu.edu.tw

Abstract: In this paper, we present a question-answering (QA) system as a virtual tutor for students in the 5th and 6th grades. Students ask questions and the QA system gives answers to their questions based on a knowledge base. Teaching materials for history and geography are considered as a knowledge source. Because question log is not available in developing QA systems, multiple choice questions (MCQs) in the learning and testing materials are regarded as a training corpus to learn question types, answer types and keywords for retrieval, where an MCQ consists of a stem and a set of options. Options from the same MCQ are grouped into a cluster. Clusters with common elements are merged into a larger cluster. A cluster is labelled with a nominal element selected from the corresponding stems. We also mine question patterns from the stems for question type analysis in the QA system. Because the questions created by instructors in MCQs and the questions asked by students may be different, we develop a procedure to collect possible questions from students in the 6th grade. In the experiments, we first evaluate the question type classification systems using the MCQ corpus and the student corpus with 5-fold cross validation, respectively. Then we train question type classifiers with the complete MCQ corpus, and test them on the student corpus. The student's and the instructor's questions are compared and analyzed.

Keywords: Computer-assisted learning, question-answering systems, student intent analysis, text mining

1. Introduction

QA system can be introduced to the education domain for collaborative learning, online education and distance learning (Wang et al., 2006; Wen et al., 2012). In this paper, we focus on developing QA applications in education as virtual tutors. Students can post questions anytime anywhere after class, and computer-assisted learning systems response answers to their questions.

A typical question answering system is composed of question type analysis, query formulation, passage retrieval, and answer extraction (Ravichandran and Hovy, 2002). Question type analysis, which aims to understand users' intents, provides important clues for the latter tasks. The clues include the question part referencing to the answer, terms indicating the type of entities being asked for, and a classification of the question into some broad types (Lally et al., 2012). Query log which keeps users' information needs provide some prior knowledge for question type analysis, but it is not always available for all the application domains.

This paper utilizes multiple choice questions (MCQs) in learning and testing materials to mine question patterns to support QA systems in education domain. Section 2 specifies an instructors' MCQ corpus and a students' question corpus used in this study. Their question type distributions are shown and discussed. Section 3 evaluates the question type classifiers with these two corpora. Section 4 concludes the remarks.

2. Question Corpora Used in This Study

2.1 Instructors' Multiple Choice Questions (MCQ) Corpus

Question type, answer type, and the keywords to retrieve potential passages containing answers are three major elements in a question. An MCQ corpus based on the learning and testing materials is used to mine the important information. An MCQ consist of a stem and 4 options. The stem describes the questions to be asked and the options list the possible answers. The following shows an example.

What sea area separates Taiwan and South Korea? (*a*) *Bashi Channel* (*b*) *East China Sea* (*c*) *South China Sea* (*d*) *Taiwan Strait*

In this example, the trigger for the question type is "*what*"; the trigger for the answer type is "*sea area*"; and the keyword for retrieval contains the predicate "*separate*" and the related arguments.

We aim to mine entities, entity type, question type, and the relation among entities from MCQs in the teaching materials. A mining algorithm is shown below.

(1) Collect initial clusters from the options in the same MCQ.
(2) Merge clusters with common elements into a larger cluster.
(3) Partition stems in MCQs based on the clusters derived from (2).
(4) Assign a label to each cluster with nominal keywords in the stems.
(5) Find the question patterns for each cluster.
(6) Extract the relations among entities.

2.2 Students' Questions Corpus

The users of the proposed question-answering system are students in the 5th and 6th grades. In this study, we would like to know how children formulate their questions, and the effects of fuzzy questions and incomplete questions on students' question type analysis and question-answering.

Sample topics about history or geography are given first, and then three different methods are proposed to guide students to formulate questions. Method 1, which rewrites the given sentences from declarative to interrogative, is the simplest way. Method 2, which keeps the same answer, but presents different views, is more complex. Method 3, whose answer is different from the original topic, is the most complex. The following lists an example.

Sample Topic: Zheng Chenggong led his troops across the sea to Taiwan, expelled the Dutch, and made Taiwan to be a base for rebelling Qing dynasty and rebuilding Ming dynasty.

Method 1: Please change the above declarative sentence into an interrogative sentence whose answer is "Zheng Chenggong", e.g., "Who expelled the Dutch who occupied Taiwan?"

Method 2: Please list questions whose answer is "Zheng Chenggong", e.g., "Who was called Koxinga?"

Method 3: Please list questions related to "Zheng Chenggong" and their answers, e.g., "Where did Zheng Chenggong land Taiwan?"

2.3 Corpus Annotation and Comparison

There are 11 common question types including (T1) people, (T2) event, (T3) time, (T4) location, (T5) object, (T6) quantity, (T7) application, (T8) theory, (T9) reason, (T10) nickname, and (T11) language. We label a question type for each question. Table 1 lists the distribution of the 11 types in the MCQ corpus (M) and the student corpus (S). These two corpora contain 593 and 395 instances, respectively. The top 5 types in the MCQ corpus are location (45.36%), object (12.48%), theory (11.80%), reason (9.28%) and people (7.92%). They occupy 86.84% of instances. Comparatively, the top 5 types in the student corpus are location (23.04%), time (21.77%), object (17.97%), people (16.46%) and event (8.61%). They cover 87.85% of instances. Location is the most interesting type in the two corpora. Object and people types are also interesting to instructors and students. Surprisingly, only 0.84% of the instances are related to event type in the MCQ corpus. In contrast, 8.61% of students' questions related to this type. Similarly, 11.80% of instructors' questions touch on theory, but only 1.01% of students' questions belong to this type. It may be because this type of questions is difficult to be formulated by students. In both corpora, quantity, nickname and language are the three minority types.

Table 1: Distribution of question types in the two experimental corpora.

	(T1)	(T2)	(T3)	(T4)	(T5)	(T6)	(T7)	(T8)	(T9)	(T10)	(T11)
M	47	5	30	269	74	9	31	70	55	1	2
%	7.92	0.84	5.06	45.36	12.48	1.52	5.23	11.80	9.28	0.17	0.34
S	65	34	86	91	71	3	10	4	23	1	7
%	16.46	8.61	21.77	23.04	17.97	0.76	2.53	1.01	5.82	0.25	1.77

3. Question Type Analysis

In the first set of experiments, we train and test the question type classifiers by 5-fold cross validation with the MCQ corpus and the student corpus, respectively. In the second set of experiments, we train the question type classifiers with the complete MCQ corpus, and test it with the student corpus. L2-regularized L2-loss support vector classification in LIBSVM (Chang and Lin, 2011) is adopted. Two sets of features are explored. The first contains triggers for question types, triggers for answer types and the relation keywords. The second contains bigrams and trigrams features. All the features are binary. Table 2 lists the accuracies of the proposed classifiers. Using triggers and relation keywords is better than using 2-grams and 3-grams in all the evaluation experiments. The classification performance in 5-fold cross validation on the student corpus with the triggers and relation keyword approach is better than that in 5-fold cross validation on the MCQ corpus. There are performance drops from cross validation on the MCQ corpus to testing on the student corpus.

Table 2: Accuracies of the proposed question type classifiers.

	Triggers and relation keyword	2-grams and 3-grams
5-fold cross validation on the MCQ corpus	0.7707	0.7673
5-fold cross validation on the student corpus	0.8608	0.7468
Test on the student corpus	0.5848	0.4734

4. Conclusion

To understand how children formulate their questions, we design a procedure to collect students' questions and analyze the qualities of the questions. We found that 155 of 683 students' questions, i.e., 22.69%, contain some incomplete, vague, ambiguous or erroneous information to QA systems. The experimental results show that the classification system using triggers and relation keywords achieves accuracies 0.7707 and 0.8608 when 5-fold cross validation on the MCQ corpus and the student corpus are adopted, respectively. The system trained with the MCQ corpus and tested on the student corpus has accuracy 0.5848.

References

Chang, C.-C. and Lin, C.-J. (2011). LIBSVM: A Library for Support Vector Machines. ACM Transactions on Intelligent Systems and Technology, 2(3), 27:1–27:27.

Lally, A., Prager, J.M., McCord, M.C., Boguraev, B.K., Patwardhan, S., Fan, J., Fodor, P., & Chu-Carroll, J. (2012). Question analysis: How Watson reads a clue. *IBM Journal of Research and Development*, 56(3/4), 2:1-2:14.

Ravichandran, D. & Hovy, E. (2002). Learning surface text patterns for a question answering system. In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics* (pp. 41-47).

Wang, C.C., Hung, J.C., Yang, C.Y., & Shih, T.K. (2006). An application of question answering system for collaborative learning. In *Proceedings of the 26th IEEE International Conference on Distributed Computing Systems Workshops*.

Wen, D., Cuzzola, J., Brown, L. & Kinshuk (2012). Instructor-aided asynchronous question answering system for online education and distance learning. *The International Review of Research in Open and Distance Learning*. http://www.irrodl.org/index.php/irrodl/article/view/1269/2369.