Multi-Strategy Question Answering System for NTCIR-7 C-C Task

Han Ren[1], Donghong Ji[1], Yanxiang He[1], Chong Teng[1], Jing Wan[2]

[1]*School of Computer Science, Wuhan University, China*
[2]*Center for Study of Language & Information, Wuhan University, China*

cslotus@mail.whu.edu.cn, dhji@ i2r.a-star.edu.sg, yxhe@whu.edu.cn
tchong616@126.com, jennifer.wanj@gmail.com

Abstract

In NTCIR-7, ACLIA task focus on more complex questions than factoid ones. To participate in Monolingual Chinese task, we propose a multi-strategy QA system that can handle questions with different types. Basic steps of this system are described as follow: first, pattern-based question type analysis is performed to classify questions of four types, then an Information Retrieval method is utilized to retrieve and re-rank documents that may contain correct answers; after that, we deal retrieved results with different strategies including heuristic rules, pattern matching and web knowledge bases; finally, nuggets are extracted from answer candidates and formed as the results. The evaluation result shows that our system achieves 43.29% average F-score(beta=3).

Keywords: *NTCIR-7, Multi-Strategy Question Answering, pattern matching, web knowledge*

1. Introduction

Current research in Question Answering(QA) mainly concern more complex questions than factoid ones. In NTCIR-7, questions in CCLQA task are defined as four types(EVENT, DEFINITION, BIOGRAPHY and RELATIONSHIP), which is different from pervious NTCIR QA tasks. Correct answers are thereupon defined as nuggets, which are minimum units containing correct response for questions.

We participate in simplified Chinese-Chinese(C-C) QA subtask in NTCIR-7, which is the first simplified Chinese monolingual task in the series of NTCIR workshop. According to requirements of the subtask, we implement a multi-strategy QA system that can handle questions with different types. Basic steps of this system are described as follow: first, question type analysis is performed to classify questions of four types, then an Information Retrieval method is utilized to retrieve and re-rank documents that may contain correct answers; after that, we deal retrieved results with different strategies according to the type of each question; finally, answer candidates are ranked and formed as the results.

The rest of the paper is organized as follows. In Section 2, we discuss the processing mechanism of our system. In Section 3, we describe the multi-strategy QA System in detail. In Section 4, we give the evaluation results and error analysis. Finally, the conclusion and future work are given in Section 5.

2. System Description

The processing mechanism of our system is show in Figure 1, which contains five modules: question type analysis, text annotation, document retrieval, answer extraction strategist and answer ranking. The system carries out the following steps:

1) for question type analysis, questions are matched with some rules to decide their types.

2) after preprocessing, documents are retrieved with annotated questions; here we utilize a free NLP tool for text annotation.

3) according to question types, different strategies, which include heuristic rules, pattern matching method, similarity calculation and web-based knowledge, are performed to select sentence candidates.

4) after extracting nuggets, answer candidates are re-ranked; after post-processing, the result is given.

Details of each module will be introduced in the following section.

3. Multi-Strategy QA System

3.1. Question Type Analysis

To deal with Chinese-Chinese QA, questions are first analyzed to obtain its question type. Unlike factoid questions, all question types cannot be detected by named entities or interrogative words in complex questions. Through analyzing questions in this CCLQA task, we find that some hint words frequently appear in questions; in particular, some hint words only appear at

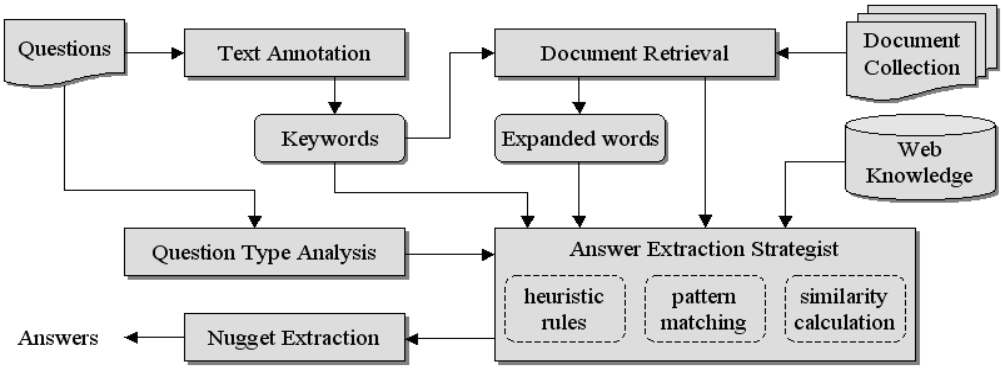

Figure 1. QA system processing mechanism

one certain type of questions. For instance, the Chinese word '列举' is only appeared at questions with 'EVENT' type. Therefore, a question potentially belongs to 'EVENT' type if it contains the word. The hint words are shown in Table 1. On the other hand, hint words are insufficient to identify the type of a question. We utilize some rules to classify questions. The rules are as follow:

1) if a question contains the word '关系', we use a template 'NE+和+NE+的+关系', to compare with the question; if they are matched, the question is assigned to 'RELATIONSHIP' type;

2) if a question starts or ends with hint words mentioned in Table 1, the question is assigned to corresponding type.

Table 1. Hint words for each question type

type of questions	hint words
DEFINITION	什么是 是什么
BIOGRAPHY	谁是 是谁
RELATIONSHIP	关系
EVENT	列举 列出 举出 说出 哪些

3.2. Text Annotation

Chinese Text Annotation module analyzes questions and documents to obtain named entities and part-of-speech for each word. We utilize ICTCLAS[1], a free Chinese lexical analysis tool which contains Chinese Word Segmentation, Part-Of-Speech(POS) tagging and Name Entity Recognition(NER), to deal with questions. Although the tool can recognize a limited number of name entity types(Person, Location, Organization and Time) which are insufficient for documents, it still works well in processing questions because named entities in questions mostly belong to these four types. After text annotation, a keyword list is built by annotated words, and a query for the retrieval is composed of these keywords. For retrieval purpose, unavailable words should be removed from the query. We build a filter to remove these words which include hint words, stop words and punctuations of Chinese and English.

3.3. Document Retrieval

For QA purpose, retrieval granularity is supposed to be specified according to the corpus. By analyzing the corpus of ACLIA task in NTCIR-7 we find that, most documents are composed of short passages; some passages have only one sentence, and some passages even have several words, not a sentence. These passages may not be considered in passage retrieval in most cases or their ranks are very low, whereas some of them are potentially relevant to the questions. In the system we utilize document retrieval to deal with the corpus.

We implement a retrieval module which is proposed by Yang, et al.[5] at NTCIR-6 workshop. The module utilizes bi-grams and single Chinese words as index units and OKAPI BM25 as retrieval module. In addition, the module utilizes a label-propagation-based learning algorithm[3, 9] to re-rank the documents before and after query expansion.

3.4. Keyword Expansion

Keywords in the initial query is insufficient for answer extraction. We propose an expansion method to augment relevant words to the keyword list, that is: in top 20 re-ranked documents after initial retrieval, we extract named entities(including person, location and organization) for each question and calculate their frequency; named entities above a threshold are added to the keyword list and utilized in answer extraction with initial keywords together.

3.5. Answer Extraction

After procedures mentioned above, keywords, expanded keywords, results of document retrieval which contain top 50 documents, and results of question type analysis are integrated as the input for answer extraction module. Since CCLQA task in NTCIR-7 is composed of complex questions, answer types for these questions are

[1] http://ictclas.org

not explicit. On the other hand, nuggets, which represent the minimum unit of correct information in answer sentences, are required as answer candidates in the task. For answer extraction, we utilize different strategies to deal with different type of questions.

3.5.1. 'DEFINITION' type

To a definition question, the answer refers to the explanation or description for it. Research has been made which is mainly based on English questions[3], whereas few of them focus on Chinese ones. Being a language of parataxis, Chinese is more complicated than English to identify whether a sentence belongs to 'DEFINITION' type or not. Most systems utilize various pattern matching approaches to acquire answers of this type[1, 4]. For this task, we produce a simple pattern-based parser to extract targets from documents with regular expressions. If a sentence which contains named entities in keywords or expanded keywords matches any one of the patterns, it is treated as an answer candidate. The patterns are shown in Table 2.

Table 2. Patterns of 'DEFINITION' type

pattern	bindings
np_1+*v*+np_2	*v* ∈ definition_verbs, $np_1 = t \wedge np_2 = t$
np+*v*	*v* ∈ affiliation_verbs
np_1+*c*+np_2	c ∈ description_words $np_1 = t \wedge np_2 = t$
$np_1(np_2)$	$np_1 = t \wedge np_2 = t$

t is the named entity in keywords or expanded keywords. By analyzing Chinese sentences of this type, we compile some words which are commonly used to them and organize as three sets: 'definition_verbs', 'affiliation_verbs' and 'description_words'. Sample words of these sets are shown in Table 3.

Table 3. Sample words of three sets

sets	words
definition_verbs	是, 就是,叫, 叫做, 称, 称为,称作
affiliation_verbs	属于, 包括
description_words	或, 或者, 即, 换言

If a sentence which contains keywords or expanded keywords does not matches any one of the patterns, it is difficult to identify whether the sentence is an explanation or description for the question. Zhang et al.[8] propose a method to identify and select definitional answers by web knowledge bases, and their experiments in TREC2004 indicate that web knowledge bases are effective resources to definitional question answering. In our system, we use three web knowledge bases, which are Baidupedia[2] , Wikipedia[3] and Hudong encyclopedia[4], to identify the probability that it becomes an answer candidate. All of them provide simplified Chinese version, so that we can easily use them for the C-C task.

We define a sentence set $S=\{s_1, s_2, \ldots, s_m\}$, in which s_i ($1 \le i \le m$) represents a sentence which contains keywords or expanded keywords but not matches any one of the patterns, and another sentence set $K=\{ k_1, k_2, \ldots, k_n \}$, in which k_j ($1 \le j \le n$) represents the definition of keywords in the *k*th web knowledge base. Then the score of each sentence in *S* is calculated by the following formula:

$$Score(s_i) = \sum_{j=1}^{n} Sim(s_i, k_j) \quad (1)$$

$Sim(s_i, k_j)$ is the similarity of sentence s_i and the *k*th knowledge base. We utilize *cosine* similarity to deal with $Sim(s_i, k_j)$, and s_i and k_j are treated as a bag of words. After ranking scores of all sentences, ones above threshold are picked out and added into answer candidates.

3.5.2. 'BIOGRAPHY' type

To a biography question, the answer actually refers to the description of a person. For this type, the processing method is similar to the definition questions which are based on pattern matching and Web knowledge bases. Patterns are generalized in Table 4, and the matching method is the same as definition questions.

Table 4. Patterns of 'BIOGRAPHY' type

pattern	bindings
np_1+*v*+np_2	*v* ∈ biography_verbs, $np_1 = t \wedge np_2 = t$
np+*v*	*v* ∈ feature_words
np_1np_2	$np_1 = t \wedge np_2 = t$

Some words are frequently appear at answers of this type, and they are compiled into feature_verbs set. Samples words of sets are shown in Table 5.

Table 5. Sample words of two sets

sets	words
biography_verbs	是, 就是,叫, 叫做, 称, 称为,称作
feature_words	生于, 出生, 现年, 籍贯

If a sentence which contains named entities(almost are person names) does not matches any one of the patterns, we also use web knowledge bases to evaluate whether it becomes an answer candidate.

3.5.3. 'RELATIONSHIP' type

By analyzing questions of this type we find that, answers mostly refer to the relationship between two named entities in a question. Since the answers are not defined clearly, we utilize heuristic rules to identify the potential targets, these are:

[2] http://baike.baidu.com/
[3] http://zh.wikipedia.org/
[4] http://www.hudong.com/

1) if a sentences contains two named entities which exist in a question, it could also contains the description of relationship between them;

2) phrase between two named entities potentially describes the relationship between them.

According to the first rule, we utilize a pattern matching method to extract answer candidates, that is: for each question, we retrieve documents and find out sentences, which contains all keywords of the question, as answer candidates. The second rule indicates nugget extraction method of this type and we will describe it in Section 3.6.

3.5.4. 'EVENT' type

Event questions often seek different types of information and we do not suppose that patterns could match all of the answer sentences. In fact, unrestrictive answers of this type indicate that users may not have a clear definition to information requirements. Andrew Hickl, et al.[2] propose a web count list strategy based on term frequency, which is obtained from online search engines such as Google, to identify the association between an answer candidate and the series target. In our system, we use it to deal with sentences which are not contained in web knowledge bases. First, we find sentences which contains keywords or expanded keywords of an event question. Then, if keywords in the question appear at web knowledge bases, we utilize the method mentioned in Section 3.5.1 to handle them; else we utilize Google to search keywords in the question and each sentence, and get a web count for the question and each sentence. After ranking, sentences above a threshold for the counts are added in answer candidates.

3.6. Nugget Extraction

For this task, the nugget aims at minimum units containing correct response for questions. We use pattern-based nugget extraction method to handle answer candidates. For definition questions, sentences which satisfy the first, third and fourth pattern in Table 2 can extract nuggets. For example, if $np_1 = t$, then np_2 is the nugget; and vice versa. For biography questions, the method can also use with sentences satisfying the first and third pattern in Table 4. In particular, a phrase matches the third pattern must appear at the head of a sentence. For relationship questions, according to the second rule mentioned in Section 3.5.3, we utilize a pattern-based rule to extract nugget, that is: if only noun phrase exist between the keywords, it is extracted as a nugget; else the entire sentence is considered as a nugget. For event questions, because information requirements is still indefinite, we treat sentences as the nuggets.

4. Results and Error Analysis

We perform two experiments which are based on question, narrative and both of them, and submit three runs for question type analysis. Due to the same results of type analysis, we submit one run for answers which is based on question.

In question type analysis, we achieve 96% overall accuracy and the detail is shown in Table 6.

Table 6. Results of Question Type Analysis

Type	Correct	Wrong	Total
DEFINITION	20	0	20
BIOGRAPHY	19	1	20
RELATIONSHIP	29	1	30
EVENT	28	2	30
Overall(%)	96%	4%	

Rules in our question type analyzer are immature, so if a question such as the question '举出通货膨胀与经济的关系' contains two or more hint words, the system probably identifies it to the wrong type. Besides, rule-based question type analysis is improper when facing more complex question like '中国对台湾问题的态度是什么'. Our system can not judge it as an event question, for the question has not a definite hint word. As a total, rules should be revised to improve the accuracy of question type analysis.

In NTCIR-7 CCLQA evaluation, nugget pyramid evaluation method is adopted and F-score(beta=3) is used to calculate response score of each question. Figure 2 shows the average score of each type questions.

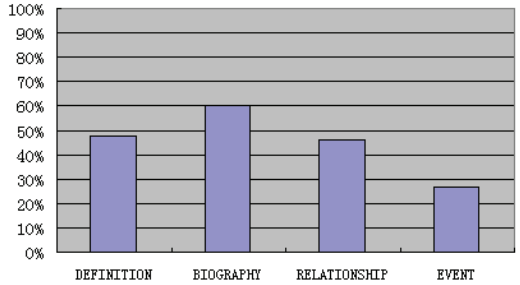

Figure 2. Average score of each type questions

From Figure 2 we can indicate that, our system has a better performance to handle the former three type questions than event questions. For these type questions, targets are relatively clear than event questions, thus we utilize pattern-based rules to extract answer candidates and nuggets, and scores of three type questions are universally higher than those of event ones. For example, one of our response for question 'ACLIA1-CS-T317' is '极光是太阳风在地球两极"绘出"的美丽图案', which is easily extracted according to patterns of 'DEFINITION' type in Table 2. On the other hand, error cases in answer extraction can be generalized as four reasons which are shown as follow:

1) retrieved documents which we utilized to answer extraction don't refer to more information of a question, such as the question 'ACLIA1-CS-T86'. Named entities in the question don't appear at retrieved documents so that the correct answers are difficult to be extracted through these documents. Therefore, re-ranking method could be adjusted to promote ranks of documents which

contains more keywords or expanded keywords in of question.

2) bias exists between the answers and the intent of a complex question, such as the question 'ACLIA-CS-T55'. Although web knowledge bases and search engines could find information related to questions, exact targets are not identified due to the complexity of them. For this purpose, more training should be performed to obtain patterns for each type of questions.

3) multiple sentences answer extraction should be considered in answer extraction. For example, the fourth nugget of system response in question 'ACLIA-CS-T84' is composed of three sentences, which is not extracted by our system. In order to deal with it, context of a sentence should be considered to get a better performance.

4) patterns are not insufficient, such as the question 'ACLIA-CS-T379'. Therefore, the investigation should be continued to find more patterns to adapt more types of questions.

In addition, pattern matching can identify simple answers; when the questions are more complex, the patterns are also more complex. In order to solve more complex questions, syntactic and semantic parsing should be utilized to identify whether a response is the answer to a question. All of these should be considered in the future work for a better performance of question answering system.

5. Conclusion

We present a multi-strategy question answering system for NTCIR-7 Monolingual Chinese subtask by using heuristic rules, pattern matching method and web knowledge bases. The system can handle questions with different types and achieves 43.29% average F-score(beta=3). Experiment shows that our system has a better performance to handle definition, biography and relationship questions for this task. In the future, we aim at syntactic and semantic methods, which could achieve a better performance in dealing with complex question answering.

Acknowledgments

This work is supported by the Natural Science Foundation of China under Grant Nos.60773011.

References

[1] Hang Cui, Keya Li, Renxu Sun, Tat-Seng Chua, Min-Yen kan. National University of Singapore the TREC-13 Question Answering Main Task. *In Proceedings of the Thirteenth Text REtreival Conference*. NIST, Gathersburg, 2004

[2] Andrew Hickl, Kirk Roberts, Bryan Rink, Jeremy Bensley, Tobias Jungen, Ying Shi, and John Williams. Question Answering with LCC's CHAUCER-2 at TREC 2007. *In: Proceedings of Text Retrieval Conference 2007*, 2007

[3] Wesley Hildebrandt, Boris Katz, Jimmy Lin. Answering Definition Questions Using Multiple Knowledge Sources. *In Proceedings of the 2nd HLT-NAACL*, Massachusetts, 2004

[4] Milen Kouylekov, Bernardo Magnini, Matteo Negri, Hristo Tanev. ITC-irst at TREC-2003: the DIOGENE QA system. *In Proceedings of the Twelfth Text REtreival Conference*. NIST, Gathersburg, 2003

[5] Yang Lingpeng, Ji Donghong, Nie Yu. Information Retrieval Using Label Propagation Based Ranking. *In: Proceedings of NTCIR-6 Workshop Meeting*. Tokyo, 2007

[6] T. Mitamura, E. Nyberg, H. Shima, T. Kato, T. Mori, C.Y. Lin, R.H. Song, C.J. Lin, T. Sakai, F. Gey, D.H. Ji and N. Kando. *Overview of the NTCIR-7 ACLIA: Advanced Cross-Lingual Information Access*. NTCIR-7, Tokyo, 2008

[7] Z.Y. Niu, D.H. Ji, and C.L. Tan. Word Sense Disambiguation Using Label Propagation Based Semi-supervised Learning. *In Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics* (ACL05), Ann Arbor, Michigan, US, pp.395-402. 2005

[8] Zhushuo Zhang, Yaqian Zhou, Xuanjing Huang, and Lide Wu. Answering Definition Questions Using Web Knowledge Bases. *In: Proceedings of IJCNLP 2005*, Korea, 2005

[9] X. Zhu, Z. Ghahramani. Learning from Labeled and Unlabeled Data with Label Propagation. CMU CALD technical report CMU-CALD-02-107, 2002