Using IS-A Relation Patterns for Factoid Questions in Question Answering Systems

Bojun SHIM†, Youngjoong KO††, Nonmembers, and Jungyun SEO†††, Member

SUMMARY This paper describes a flexible strategy to generate candidate answers for factoid questions in Question Answering (QA) systems. Most QA systems have predefined the conceptual categories for candidate answers. But if the conceptual category of answers to any question is not prepared in the QA system, it is hard to extract correct answers to that question. Therefore, we propose an extraction method for IS-A relation patterns which describe relations between the nominal target concepts of question and candidate answers. The extracted IS-A relation patterns can be used for questions with an unexpected target concept.

key words: Information Retrieval, Question Answering System, IS-A Relation Pattern, Factoid Question

1. Introduction

Until now, most existing QA systems have generally concentrated on improving the ability to deal with limited and typical target concepts. In the factoid questions from TREC, there are a lot of questions with typical target concepts such as 'city' and 'country' as follows:

"What city is Disneyland in?"
"What country did Catherine the Great rule?"

In order to extract candidate answers for these questions in QA systems, the finite conceptual categories of expected answers are previously predefined and Named Entity (NE) recognizers are widely used. However, if any question has an unexpected target concept or the target concept cannot be captured by the NE recognizer, the QA systems must have an additional module for handling the exception to generate candidate answers. For example, if a QA system does not redefine the target concept 'book' as an unexpected target category, it is highly difficult to extract candidate answers that are semantically hyponyms of 'book'.

"What book did Rachel Carson write in 1962?"

Whenever an unexpected target concept occurs in questions, the target concept must be added to the set of conceptual categories in QA systems. In addition, since every common noun word can be the target concepts, it is highly difficult or nearly impossible to build up conceptual categories for all the common noun words. As you can observe from the TREC question sets, the number of questions with unexpected target concepts has been increasing. In this paper, we suggest a flexible and unlimited method that can easily extract candidate answers for questions with unexpected target concepts. Since there are always IS-A relations between candidate answers and target concept, we can extract Lexico-Semantic Patterns (LSP) [4][6] of their IS-A relations from the World Wide Web (WWW) documents and assign a confidence score to each pattern. Using these IS-A relation patterns, we can extract candidate answers corresponding to the unexpected target concepts. The rest of this paper is organized as follows. Section 2 describes related works in existing QA systems. Section 3 explains the new definition of target concepts for QA systems using IS-A relations. In section 4, we describe the processes to extract IS-A relation patterns and to generate candidate answers using the extracted patterns. Section 5 is devoted to the analysis of experimental results. The final section presents conclusions and future works.

2. Related Works

Here, we simply describe the distinctions between the proposed system and other related works. First, previous QA systems such as Harabagiu [3] and Echihabi [2] in TREC 2003 adopted predefined limited numbers of answer types. While these systems tried to resolve questions with typical target concepts, our study focuses on questions with unexpected target concepts. Second, Ravichandran [9] and Riloff [10] exploit bootstrapping algorithms to extract lexical patterns. Ravichandran extracted the lexical patterns from phrases including answers, corresponding to each answer type. To get general patterns from the small number of samples, Ravichandran used the bootstrapping algorithm similar to Riloff's one. We applied the algorithm to extract LSP of IS-A relations. Finally, Mann [8] and Hearst [5] extracted hyponyms from large text corpora with manually prepared patterns. While Mann and Hearst used IS-A relation patterns to be manually defined, we automatically extract IS-A re-
Table 1  The examples of target concepts categorization in typical methods

<table>
<thead>
<tr>
<th>Questions</th>
<th>Target Concepts (Answer Types)</th>
</tr>
</thead>
<tbody>
<tr>
<td>What actress has received the most Oscar nominations?</td>
<td>PERSON</td>
</tr>
<tr>
<td>What beach was &quot;I Dream of Jeannie&quot; filmed on?</td>
<td>LOCATION</td>
</tr>
</tbody>
</table>

Table 2  The examples of target concepts categorization in the proposed method

<table>
<thead>
<tr>
<th>Questions</th>
<th>Target Concepts (Answer Types)</th>
</tr>
</thead>
<tbody>
<tr>
<td>What actress has received the most Oscar nominations?</td>
<td>ACTRESS</td>
</tr>
<tr>
<td>What beach was &quot;I Dream of Jeannie&quot; filmed on?</td>
<td>BEACH</td>
</tr>
</tbody>
</table>

relation patterns from web documents and assign them confidence scores.


3.1 New Target Concepts for Questions

Table 1 shows a classification method with typical target concepts which has used in many QA systems.

In this method, it is important not only to properly categorize target concepts but also to thoughtfully redefine the categories of target concepts. To make an answer about a question with an unexpected target concept, the QA systems must add a new target concept category to the set of expected target concepts. However, even though a QA system has very specific and well-defined categories, it is impossible to prepare all the anticipated target concepts. Thus most of QA systems generally have 'others' category which may be a blind point to classify the target concepts for answers [7]. Therefore, our method makes use of common noun words as target concepts directly as shown in Table 2. Table 2 shows the proposed method of target concepts categorization. By using this method, target concepts of all the questions can be automatically detected without any exceptions.

3.2 Candidate Answers Extraction from the IS-A Relation Patterns

As you can see in Table 2, the candidate answers must be one of 'X'es which satisfy "X is-a ACTRESS", "X is-a BEACH", and "X is-a BOOK". To efficiently extract these 'X'es, we use the IS-A relation patterns represented by LSP. We generate IS-A relation patterns in web documents from a WWW search engine. In the next chapter, we will describe how to create the IS-A relation patterns and how to utilize them for extracting candidate answers. In fact, NE recognizers can be also used to extract IS-A relation patterns but it has several limitations. First, the NE recognizers have only limited categories to extract hyponyms of the concepts. Second, some categories of general NE recognizers have too broad semantic variations (e.g. ORGANIZATION can have many semantic variations such as a government office, NGO, a political party, etc.).

4. Extracting and Utilizing of IS-A Relation Patterns

4.1 Lexico-Semantic Patterns Representation

In order to represent the patterns of a phrase including IS-A relation, LSP is used in the proposed method. In LSP, a used tag set consists of function words, punctuation marks, Part Of Speech (POS) tags, and NE tags; function words and punctuation marks are used as lexical tags and other words are represented by NE tags or POS tags. Each word is uniquely denoted as one of three tag types. Fig. 1 shows an example of our LSP representation.

4.2 Using Web Documents to Extract IS-A Relation Patterns

Web documents have been used as a knowledge resource for lots of QA systems due to their abundance and accessibility. Brill [1] used a web search engine to aggregate web documents and he acquired n-grams anticipated as answers. In the proposed method, we use the snippets returned by the Google web search engine for extracting IS-A relation patterns, assigning confidence scores to the extracted patterns, and generating candidate answers.

4.3 Extracting IS-A Relation Patterns and Assigning Confidence Scores

The first step to extract patterns is to select the pairs of seed words obviously with IS-A relationship. We manually picked out 110 <candidate answer(X), target-concept(Y)> pairs from the QA data between TREC 1999 and TREC 2002. Table 3 shows a part of the selected pairs.

In the next step, patterns with these <X, Y> pairs
Table 3  Examples of word pairs, \(<X, Y>\) means \(X\) 'is-a' \(Y\)

<table>
<thead>
<tr>
<th>(X)</th>
<th>(Y)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Oilers</td>
<td>Drink</td>
</tr>
<tr>
<td>Iberia</td>
<td>Peninsula</td>
</tr>
<tr>
<td>Napoleon</td>
<td>Ruler</td>
</tr>
</tbody>
</table>

are extracted from snippets of the web search engine. Fig. 2 shows an example of the process to extract a pattern from the word pair of \('<\text{calcium}', \text{mineral}'>\).

The extracted patterns may have various levels of confidence scores. Fig. 3 shows an example of assigning a confidence score to each pattern.

Fig. 3 shows how to evaluate a pattern "X CC JJ Y/NN" extracted from a pair of seed words '<calcium', 'mineral'>; these pairs are denoted by 'XY_pair'. At first, the X term ('calcium') of this pair is chosen as a query for web search. Then the collected documents from the result of web search are filtered into only 1,000 sentences. Since all the filtered sentences include 'calcium', the matched phrases with the pattern of "X (calcium') CC JJ Y/NN' can be extracted from those sentences annotated with LSP representation. If

\[
C.S(Pattern) = \frac{\text{right}_\text{cnt}}{\text{right}_\text{cnt} + \text{wrong}_\text{cnt}} \times \frac{\log(\text{right}_\text{cnt})}{\log(\text{XY}_\text{pair}_\text{cnt})}
\]

where \(\text{XY}_\text{pair}_\text{cnt}\) denotes the number of \(<X, Y>\) pairs from which the pattern is extracted with the process of Fig. 2. For example, if the pattern, "X CC JJ Y/NN', of Fig. 3 occur in many 'XY_pair's, the confidence score of this pattern decrease by formula 1. Especially, although the pattern 'X Y/NN' appears in most sample 'XY-pairs', it is not a useful pattern because it has too general usage. Table 4 shows a part of the patterns and the corresponding confidence scores of them generated by processes of Fig. 2 and Fig. 3. The confidence scores are normalized to between 0 and 1 by being divided by the highest score. Finally, we obtained 26 patterns from our method.

<table>
<thead>
<tr>
<th>Selected Patterns</th>
<th>Confidence Scores</th>
<th>Examples X is-a Y</th>
</tr>
</thead>
<tbody>
<tr>
<td>Y/NN in the NNP</td>
<td>1.000</td>
<td>word in the Oxford English Dictionary is fluorocinacnihilification</td>
</tr>
<tr>
<td>NNP NNP is X</td>
<td>0.962</td>
<td>calcium is the most abundant mineral</td>
</tr>
<tr>
<td>X is the RBS JJ</td>
<td>0.672</td>
<td>Mars is the fourth planet</td>
</tr>
<tr>
<td>Y/NN</td>
<td>0.648</td>
<td>calcium is a mineral</td>
</tr>
</tbody>
</table>

4.4 Extracting Candidate Answers Using the Patterns

Now we can extract hyponyms of target concepts by using the LSP pattern set in the previous section. The web search engine is also used for this process. The process to extract candidate answers is similar to the one to calculate the confidence score, but there is one significant difference; in Fig. 3, we already know what an X term of the "X is-a Y" relation is, but we here know what Y as a target concept is. Fig. 4 shows the process of extracting candidate answers.

As shown in Fig. 4, candidate answers are the hyponyms of the target concept, 'river'. A query for web search is generated from the target concept word and the content words of the question. Then the collected...
documents from the result of web search are filtered into only 1000 sentences that must include 'river'. We try to match each pattern with the sentences annotated as the LSP representation. Then X is picked out as a candidate answer from matched sentences and its candidate score is calculated by a pattern score and a sentence score (formula(3)); the pattern score is the confidence score calculated by a pattern score and a sentence score (formula(3)); the pattern score is the confidence score of a matched pattern and the sentence score is calculated by formula (2):

\[
\text{Sentence\_Score}(ans_{ij}) = \frac{NQ}{\text{Max}D} \tag{2}
\]

\[
\text{Candidate\_Score}(ans_{ij}) = \text{Sentence\_Score}(ans_{ij}) \times \text{Pattern\_Score}(ans_{ij}) \tag{3}
\]

where \(NQ\) denotes the number of query words included in \(j^{th}\) sentence and \(\text{Max}D\) is the maximum distance between query words and the candidate answer. \(ans_{ij}\) denotes \(i^{th}\) answer candidate that occur in \(j^{th}\) sentence. The final score of any candidate answer \(S_p(ans_i)\) is calculated by accumulating the candidate scores of each sentence as follows:

\[
S_p(ans_i) = \sum_{j=1}^{n} \text{Candidate\_Score}(ans_{ij}) \tag{4}
\]

where \(n\) is the number of sentences which include the candidate answer, \(ans_{ij}\).

4.5 The Hybrid Methods to Combine the NE Method and the Pattern Method

In general, the coverage of the proposed method using patterns is wider than the method using a NE recognizer. However, each method shows the weak point in different cases of target concepts. The following examples show the cases that only one method can generate the correct answer.

**Table 5** The process of the Hybrid 1

<table>
<thead>
<tr>
<th>Target concept</th>
<th>Answers by NE</th>
<th>Answers by Pattern</th>
<th>Final Answers</th>
</tr>
</thead>
<tbody>
<tr>
<td>President</td>
<td>Abraham, Lincoln (0.65), Librarian (0.7)</td>
<td>Abraham, Lincoln (0.65)</td>
<td>Abraham, Lincoln (0.65)</td>
</tr>
<tr>
<td>gas</td>
<td>NONE Oxygen (0.7), Hydrogen (0.6)</td>
<td>Oxygen (0.7)</td>
<td>Oxygen (0.7)</td>
</tr>
</tbody>
</table>

"What gas is 78 percent of the earth's atmosphere?" (By only the Pattern method)

"What president served 2 nonconsecutive terms?" (By only the NE method)

In the first example, the target concept, 'gas', can be easily detected by patterns while it cannot be covered by NE recognizer, and 'president' as the target concept of the second example can be easily captured by a 'PERSON' by the NE recognizer. If two methods can be efficiently combined, their strong points can compensate for their weak points. Therefore, we constructed the hybrid methods as follows:

**Hybrid 1**: if the NE method can detect candidate answers, the pattern method extracts answers among the candidate answers previously detected by the NE method. Otherwise, the hybrid 1 method acts like the pattern method; only the pattern method is applied to candidate answers selection. Finally, the scores are calculated by the answer score of the pattern method.

Table 5 shows the examples of the process of Hybrid 1.

**Hybrid 2**: We combine the confidence scores from two methods by a linear combination as the following formula:

\[
S(ans_i) = S_p(ans_i) \times \log(n + 1) \\
+ S_{ne}(ans_i) \times \log(m + 1) \tag{5}
\]

\[
S_{ne}(ans_i) = \sum_{j=1}^{n} \text{Sentence\_Score}(ans_{ij}) \\
\times \text{NE\_Score}(ans_{ij}) \tag{6}
\]

Where \(S(ans_i)\) denoted the score of the candidate answer, \(ans_i\). \(S_p(ans_i)\) is the answer score by the pattern method, \(S_{ne}(ans_i)\) is the score by the NE method, \(n\) is the number of sentences which include the candidate answer in the pattern method, and \(m\) is the number of sentences that include the candidate answer in the NE method. Formula (6) shows how to calculate \(S_{ne}(ans_i)\). Here the NE score is 1.0 if \(ans_i\) is detected by the NE recognizer in \(j^{th}\) sentence, otherwise 0.0 and the sentence score is calculated by the same method as one of the pattern method. \(m\) denotes the same number as \(m\) in formula (5).
Table 6  Experimental results of each method

<table>
<thead>
<tr>
<th>Method</th>
<th>MRR</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>NE</td>
<td>0.202</td>
<td>0.127</td>
</tr>
<tr>
<td>Pattern</td>
<td>0.258</td>
<td>0.218</td>
</tr>
<tr>
<td>Hybrid1</td>
<td>0.356</td>
<td>0.324</td>
</tr>
<tr>
<td>Hybrid2</td>
<td>0.427</td>
<td>0.359</td>
</tr>
</tbody>
</table>

5. Experimental Evaluation

To verify the effectiveness of the IS-A relation patterns in QA systems, we first selected 142 questions with a type of "What Noun-Phrase Verb-Phrase" in TREC 2003 test questions because the head noun of a noun phrase in this type of questions can be obviously resolved as the target concept. As performance measures, we followed the standard definition of MRR and Accuracy measures for TREC evaluations. According to the TREC 2003 judgement set, we determined whether the extracted candidate answer is a correct answer.

5.1 Experimental Results

We conducted experiments for the NE method using the NE recognizer with 32 NE tags, the Pattern method, and two hybrid methods. Table 6 shows the result of each experiment. The accuracy values mean the proportion of the correct answers when the system submits only one answer per each question. In this experiment, the pattern method reported higher performance than the NE method and the hybrid 2 showed the best performance among all the methods. This result means our proposed pattern method properly works to capture correct answers and the hybrid methods are also effective to make up for the weak points of both the NE and Pattern methods.

6. Conclusions and Future Works

In this paper, we proposed a flexible method which can easily find answers with unexpected concepts. As shown in our experiments, the IS-A relation patterns with confidence scores are effectively used when users ask questions with unexpected answer types that cannot be easily captured by the NE recognizer. This study awaits several further researches. First of all, we need an additional study to extract the main target concepts from questions with the complicated type. And we have a plan to apply the IS-A extraction technique to other information extraction systems.

Acknowledgments

This research was performed for the Intelligent Robotics Development Program, one of the 21st Century Frontier R&D Programs funded by the Ministry of Commerce, Industry and Energy of Korea.

References