Interactive Document Expansion for Answer Extraction of Question Answering System

Junichi Fukumoto\textsuperscript{a}, Noriaki Aburai\textsuperscript{b}, Ryosuke Yamanishi\textsuperscript{a}

\textsuperscript{a}Dept. of Media Technology, Ritsumeikan University
\textsuperscript{b}Graduate School of Science and Technology, Ritsumeikan University
1-1-1 Nojihigashi, Kusatsu, Shiga 525-8577 JAPAN

Abstract

In this paper, we propose a method to navigate getting a correct answer for Question Answering (QA) system using user interaction. QA is a technology which extracts appropriate answer strings for a given question sentence from huge documents such as Web, newspaper articles etc. If a given question is ambiguous, answers will be various ones according to its possible understandings and retrieved documents with query words of the question sentence will consists of various types of information. In order to focus on intended topic, it is necessary to provide more information to narrow down search area for a question. In our approach, a QA system selects a clue word to decide an appropriate topic from topics in retrieved documents and interacts with a user whether this clue word is appropriate one or not. Then, search space will be reduced using this clue word. However, such narrowing down reduces the number of answer candidates because the number of target documents will be decreased. We will re-retrieve documents using this clue words and expand search space to increase possibility of getting correct answer candidates.

Keywords: Question Answering, User interaction, Named Entity extraction

1. Introduction

Question Answering (QA) is a technology which extracts appropriate answer strings for a given question sentence from huge documents such as Web, newspaper articles etc. For example, when a question sentence “Who is a prime minister of Japan?” is given, QA system will provide answer string “Mr. Abe” from huge documents. There have been a number of work on QA and the most of them are on some evaluation workshops such as TREC QA track \cite{1, 2}, CLEF\textsuperscript{2} and NTCIR QAC \cite{3, 4}. In these tasks, task participants are required to answer for given questions and answers are prepared for these questions for evaluation. However, in practical areas, there are several cases that questions are ambiguous because they do not have enough information to focus on a specific question topic.
If a given question is ambiguous, it will be difficult to get one intended specific answer. In this case, obtained answers will be various ones according to possible understandings of a question and retrieved documents with query words of a question sentence will be various types of information. In order to focus on intended topic, it is necessary to provide more information to narrow down search area for a question. HITIQA system [5] is an interactive open-domain question answering system which reports to satisfy a given scenario template, and this system obtains information interactively, but interaction is based on some scenario. Therefore, their interaction strategy will not be flexible. Shirai et. al. [6] proposed a method to extract an appropriate information to understand an ambiguous question. However, they only extracted some information to narrow down a question topic and did not realize actual user interaction in question answering dialogues.

In our approach, a QA system selects a clue word to decide an appropriate topic from topics in retrieved documents and interacts with a user whether this clue word is appropriate one or not in user interaction. Then, search space will be reduced using this clue word. However, such narrowing down reduces the number of answer candidates because the number of target documents will be decreased. We will re-retrieve documents using this clue words and expand search space to increase possibility of getting correct answer candidates.

In order to select such a clue word, our QA system [7] firstly extracts modifying words to query words of an input question because modifying words will be constraints to query words. Then, the system classifies the extracted modifying words according to their Named Entity types such as city names, sports names and so on and the most frequent type will be selected. Words in this type will be used for narrowing down and re-retrieving documents. In order to select an appropriate word, the system retrieves documents using each word in this type and chooses the document set which includes the largest number of answer candidates. QA system provides the word as the best clue word which can retrieve the largest number of answer candidates using user interaction. QA system will continue to the above interaction until a user get an appropriate answer to the given question.

In the following sections, we will firstly describe a brief overview of QA system and how user-interaction works in QA system. Then, we will show extraction method of clue words from retrieved documents and their classification for implementing user-interaction. Next, we will show how to use the clue word for re-retrieval of documents with interaction. Finally, some experimental results of our method using some question sentences and interaction will be presented.

2. Overview of QA system with user-interaction

We will firstly describe how a QA system provides answers for a given question. Then, we will show an overall process of user interactive question answering in our QA system.

2.1. An overview of Question Answering

In a QA system, a given question will be analyzed, then question type and queries are determined. Question type is information which a given question is asked for and queries are used to retrieve documents from information source. For example, when a question sentence “Who is a prime minister of Japan?” is given, the question type is person and queries are “prime,” “minister” and “Japan.” The word “who” will not be extracted because it is a question term and is not appropriate for document retrieval. Then, a QA system will retrieve target documents from information sources using IR engine such as Google engine. Usually, the number of documents will be ten to thirty or fifty although it depends on QA system setting. The more the number of retrieved documents is, the more time and noise an answer extraction module will have. Answer candidates will be extracted from the retrieved documents using named entity recognition [8] module. If a given question requires a person name, possible person names in retrieved documents are recognized using named entity recognition module. The extracted answer candidates will have their weight for ordering their certainty of the correct answer. There have been many ways of calculating their weight such as word distance using query words. According to weight calculation, possible correct answer will be given with their ordering. For the above sample question, a QA system will provide answer string “Mr. Abe.”
2.2. User interaction to focus on intended topic

If a given question has few information and is ambiguous, retrieved documents for such a question consist of several topics. It is necessary to provide some more information to choose an appropriate topic from the retrieved documents. Yamauchi et. al. [9] proposed a method to focus on intended topic using user interaction. Their approach extracts an appropriate word from retrieved documents and utilizes the word to narrow down search space which is obtained by document retrieval in QA. Such a word is chosen from modifying words to query words in a question sentence.

Figure 1 shows the process of narrowing down search space using a question sentence “オリンピックで金メダルを取ったのは誰ですか？(Who won gold medal in Olympic?)” 3 There are many topics in the retrieved documents and QA system will choose topic about Olympic cities. The left figure indicates that the retrieved documents contains topic about Olympic game cities: Athens, Beijing and London. When a user responds Beijing in user interaction, QA system can selected documents about Beijing Olympic. The focused document still contains several topic and QA system will choose sports event at Beijing Olympic. The center figure indicates topic about sports event such as Judo, swimming and marathon. Then a user required to give information about it and responds Judo. Documents about Judo event at Beijing Olympic will be selected and the right figure shows that there are no more topics in the selected documents. QA system provides answer “Mr. Satoshi Ishii” for the answer of the given question.

Fig. 1. Narrowing down search space in QA

In their approach, QA system can narrow down search space to get appropriate answer using user interaction, however, the number of documents to extract answers will be getting decreased and the possibility of getting a correct answer will become lower.

2.3. Search space expansion with user interaction

In our method, QA system extracts such information from the retrieved document and ask a user whether this information is appropriate or not using user interaction. If QA system can get additional information to focus on intended topic, system will retrieve document again using a clue word obtained by interaction.

We will show an overall process of user interactive question answering on our system in the followings.

1. QA system analyzes a given question sentence to get a question type and query terms for document retrieval.
2. Query terms are used to retrieve target documents using IR engine such as Google engine.
3. QA system extracts modifying words to query words in the retrieved documents and classifies the modifying words according to their information types determined by Named Entity recognizer.
4. The number of documents for each information type will be counted and the most frequent one will be chosen as target information type.

3In this paper, we have implemented our QA system for Japanese question answering, and all the example questions and answer expressions are in Japanese. We leave these Japanese expressions in original expressions. However, we put their pronunciation and English meaning in the next Parentheses.
5. For each clue word in the chosen information type, IR engine retrieves documents together with the first query words, and the number of answer candidates of the analyzed question type is counted.
6. The clue word which has the most frequent answer candidates in the retrieved documents will be selected for interaction.
7. QA system asks a user whether an appropriate topic is related to the selected clue word or not.
8. If a user replies “yes,” the target document set for answer extraction will be the selected one and go to the above process 3. If “no,” QA system will provide answer candidates of the current document set.
9. If there is no clue words in the selected information type, the second information type will be selected. Then go to the process 5 until there is no information type.

For the selection of information type of clue words in the above process 4, the more words exist in some information type, the more effective QA system can focus on an appropriate topic. This is because the system explores answer candidates in the most popular domain. In the process 6, our method prefers a situation which has more answer candidates. It increases a possibility to find a correct answer for QA system.

3. Extraction of clue words from retrieved document

Target documents are retrieved with query words which are extracted from a given question. Query words appear in these documents and are used to extract answer candidates in QA system. A query word sometime is composed of compound noun with some other words. For example, in a question sentence “Who is a gold medalist at Olympic game?,” The word “Olympic” in this question sentence might appear with the other words “London,” “Beijing,” “2012,” “winter” in the form of compound nouns. These adjacent words put semantic constraints on the word “Olympic,” that is, London Olympic is Olympic which is in London. In another case, modifying words to query words sometimes appear with Japanese particle “の (of).” This modifying phrase with the particle “の (of)” will also make semantic constraints on query word. For example, in “オリンピックの水泳: Olympic no suiei (swimming of Olympic),” sport event name “水泳 (swimming)” is described using Japanese particle “の (of)” and modifies the word “水泳 (Olympic)” as semantic constraints. We will extract such modifying words using the following patterns shown in Table 1. In this able, the expression <extracting word> is a word which will be extracted and the expression <query word> is a query word in a question sentence.

<table>
<thead>
<tr>
<th>Num.</th>
<th>pattern</th>
<th>example</th>
<th>example (English)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>&lt;extracting word&gt; &lt;query word&gt;</td>
<td>ロンドンオリンピック</td>
<td>London Olympic</td>
</tr>
<tr>
<td>2</td>
<td>&lt;query word&gt; &lt;extracting word&gt;</td>
<td>オリンピック2012</td>
<td>Olympic 2012</td>
</tr>
<tr>
<td>5</td>
<td>&lt;extracting word&gt; · &lt;query word&gt;</td>
<td>2012·オリンピック</td>
<td>2012-Olympic</td>
</tr>
<tr>
<td>6</td>
<td>&lt;query word&gt; · &lt;extracting word&gt;</td>
<td>ワールドカップ·サッカー</td>
<td>Worldcup-soccer</td>
</tr>
<tr>
<td>7</td>
<td>&lt;extracting word&gt; の &lt;query word&gt;</td>
<td>柔道のオリンピック</td>
<td>Olympic of Judo</td>
</tr>
</tbody>
</table>

The extracted words will be classified into some information types using Named Entity recognizer. We used Named Entity Recognize call iNExT, which is an improved version of the original NExT system [10, 11], which has 77 Named Entity categories. For example, “Beijing,” “London,” “Athens” and “Tokyo” are classified as category CITY and “Swimming,” “Judo,” “Tennis” are as category SPORTS.

The extracted words are classified into several groups according to identified Named Entity categories. The category which includes the most different kinds of words will be selected. In the above example, there are four different words in the category CITY and three in the category SPORTS, then the category CITY will be selected. We will count the number of different words, not the sum of the words.
When a category is selected, our method retrieves documents using each word in the selected category and the original query words, then checks the number of answer candidates in the retrieved documents. In the above example, when the category CITY is selected, our system retrieves documents using the clue words “Beijing,” “London,” “Athens” and “Tokyo” and the number of answer candidates in each retrieval results will be counted. If the retrieval results using “Beijing,” “London,” “Athens” and “Tokyo” includes 5, 3, 6 and 1 answer candidates in the retrieved results, respectively, the clue word “London” will be selected.

We will show a diagram of re-retrieval of documents of each category in Figure 2. The middle circle indicates category classification of the first retrieved documents. For three categories, clue words “Beijing,” “Athens” and “London” are used for re-retrieval and document set will be expanded in each category.

![Diagram](image)

Fig. 2. Explanatory diagram of hierarchical coordination

The newly retrieved documents are also classified according to identified category in each document set. In the retrieved document using clue word “Beijing,” the selected category is SPORTS, which clue words are “Boxing,” “Judo,” “marathon” and “swimming.” In the retrieved document using clue word “London,” the selected category is also SPORTS, which clue words are “Boxing,” “Judo,” “archery” and “swimming.” These new categories depend on their document set.

4. Sentence generation for user-interaction

When a clue word is selected, a user is required to judge whether focusing of topic indicated with this clue word is appropriate or not. For this inquiry, a sentence will be generated using the following pattern.

<clue word> の <query word> ですか？ (Is <query word> of <clue word> ?)

The <clue word> means the selected clue word and the <query word> means the query word which the selected clue word modifies. A user replies “yes,” the clue word will be used for re-retrieval of documents with the first query words. For example, when the selected clue word is “Beijing” which modifies the query word “Olympic,” a sentence “北京のオリンピックですか？ (Is Olympic of Beijing?)” will be generated. Then, QA system can understand the word “Olympic” in a question sentence is actually intended to express
“Beijing Olympic” and try to find answer candidates in the newly retrieved documents using the clue word “Beijing.” If a user responds “no,” the second most clue word in the selected category will be selected. If a user responds “no” more than three times in the selected category, the second category will be chosen. New clue words extracted from the newly retrieved documents as the same approach as the first category.

After selecting a clue word, QA system extracts answer candidates from the newly retrieved documents. If no new answer candidate cannot be found in the newly retrieved documents, then system will present the last answer candidates in some order and all the process will terminate.

5. Experiments of question answering

In this section, we will show some experimental results of interaction on our QA system. It might be difficult to show numerical performance of our method, so we will present actual interactions and answer lists of some question sentences. In these results, “question” shows a given question with English translation in the brackets and “query” indicates a list of query words extracted from this question sentence. English translation sometimes consists of several English words although Japanese is one word. There are top ten answers with their score. After the answer list, interaction started. When interactions terminated, top ten answers from re-retrieved documents are presented.

Figure 3 shows an example of interaction of a question ”大河ドラマの主演は誰ですか? (Who is the main cast in Taiga drama?).” In the first three interactions, YEAR category is selected and clue words are “fifteen,” “2013” and “2012.” The answer list is not intended one for a user, so s/he replied “no.” Then, the new category PERSON/ACTOR will be selected and user intended one “Atsuhime princess,” which is used for an additional query word, is selected in the fifth interaction. The correct answer is at 10th in the first answer list but is at 1st and 8th in the second successfully.

In the second example shown in Figure 4, category TITLE is selected and clue words used in interaction are “Oshin” and “Teacher Umechan.” A user replied “yes” for the second word, then this clue word is used for re-retrieval of documents. The intended answer is “Maki Horikita”, then 4th and 5th in the first answer list are correct and 1st and 3rd are in the second answer list. After the interaction, correct answers are ranked in higher order.

In Figure 5, the selected query word for interaction is “Japan” and its modifying words are “football,” “soccer” and “F1” (SPORTS category). In this case, the generated sentences for interaction are not good because combination of these words is not natural. If system can select query word “Grand Prix” or “Japan Grand Prix,” sentence generation might work well. However, the correct answer in the first list is ranked in 10th and the one in the second list is in 7th, then fortunately the result was not bad.

6. Discussions

In the experiments, we prepared 10 questions. User interaction occurred in all questions after presenting the first answer list. We will discuss how our method works for interaction in QA system from several points.

• Extraction of clue words

In the experiments, we have succeeded to extract modifiers for query words because many query words have modifying words in documents. This means that selecting modifiers is effective to identify ambiguity of query words in documents. The extracted modifiers tend to be categorized into more than one category and there are several words in each category. Therefore, modifiers for query words can be useful clue information to narrowing down information in documents.

• Query words in compound noun

In our current implementation, a word which has a lot of modifiers will be selected for interaction, so a frequent word tends to be selected. In the third example shown in Figure 5, “Japan Grand Prix” is a compound noun. The word “Japan” is frequent and then, modifiers to this word are used to make
Input question: 大河ドラマの主演は誰ですか？ (Who is the main cast in Taiga drama?)

Intended answer: 宮崎あおい (Aoi Miyazaki)

query = 大河ドラマ (Taiga drama), 主演 (main cast)

***** SCORE and ANSWER *****

1st 11.504 / 漫画 (Haruka Ayase)
2nd 10.779 / 平清盛 (Kiyomori Taira)
3rd 10.566 / 岡田 (Okada)
4th 9.396 / 松山ケン (Ken Matsuyama)
5th 5.913 / 官兵衛 (Kanbei)
6th 5.122 / 新島八重 (Yae Niijima)
7th 4.828 / 黒田官兵衛 (Kanbei Kuroda)
8th 4.555 / 豊臣秀吉 (Hideyoshi Toyotomi)
9th 4.197 / 秀吉 (Hideyoshi)
10th 3.880 / あおい (Aoi)

1 5週年の大河ドラマですか？ (yes | no) (Is Taiga drama of 15th?)
>no.
2 013年の大河ドラマですか？ (yes | no) (Is Taiga drama of 2013?)
>no.
2 014年の大河ドラマですか？ (yes | no) (Is Taiga drama of 2012?)
>no.
平清盛の大河ドラマですか？ (yes | no) (Is Taiga drama of Kiyomori Taira?)
>no.
篤姫の大河ドラマですか？ (yes | no) (Is Taiga drama of Atsuhime princess?)
>yes.

query = 大河ドラマ (Taiga drama), 主演 (main cast), 篤姫 (Atsuhime princess)

***** SCORE and ANSWER *****

1st 17.282 / あおい (Aoi)
2nd 12.285 / 天智院 (Tenshouin)
3rd 9.715 / 徳川家定 (Isada Tokugawa)
4th 9.299 / 徳川 (Tokugawa)
5th 9.035 / 阪 (Masato)
6th 7.795 / 松坂慶子 (Keiko Matsuzaka)
7th 7.681 / 和宮 (Kazunomiya)
8th 7.550 / 宮崎あおいさん (Aoi Miyazaki)
9th 7.496 / 場口幸子 (Kazunomiya)
10th 7.461 / 小松野田 (Tatsumi Komatsu)

category to interaction. The expression “Japan Grand Prix” had to be used in one query word in this case. It might be necessary to handle a compound noun as one query word but we need further analysis for improvement.

• Determination of categories

Words in the same Named Entity type will be used for interaction; therefore, it is important to set appropriate detail level of Named Entity types. We used our Named Entity recognition tool which has 77 types in two levels. Upper types are person name, organization name, place name etc. and there are several subtypes for upper types. Upper and lower types are recognized as different, for example, athlete name and person name are treated in different. It is difficult to identify Named Entity types correctly for many elements. If there are many athlete names in documents, all of them cannot be recognized as athlete names in recognition of Named Entity. This is because NE tool has little information to identify them or such names have no clue to identify to be athlete name.

• Re-retrieval of documents

In the experiments, there was a case that re-retrieval of documents failed. The question is to ask
Input question: 電視小说の主演は誰ですか？ (Who is the main cast in TV novel?)
Intended answer: 堀北真希 (Maki Horikita)
query = テレビ (TV), 小説 (novel), 主演 (main cast)

***** SCORE and ANSWER *****
1st 9.080 / 酒 (Jun)
2nd 7.427 / 梅ちゃん先生 (Teacher Umechan)
3rd 6.567 / 夏 (Natsu)
4th 6.186 / 堀北真希 (Maki Horikita)
5th 4.668 / 堀北真希さん (Ms Maki Horikita)
6th 3.766 / ファン (fun)
7th 3.603 / 玲奈 (Reina)
8th 3.142 / 福士 (Fukushi)
9th 3.000 / 滝沢秀明さん (Hideaki Takizawa)
10th 2.918 / 北真希 (Maki Kita)

おしかん的小説ですか？ (yes | no) (Is novel of Oshin?)
> no.
梅ちゃん先生的小説ですか？ (yes | no) (Is novel of Teacher Umechan?)
> yes.
query = テレビ (TV), 小説 (novel), 主演 (main cast), 梅ちゃん先生 (Teacher Umechan)

***** SCORE and ANSWER *****
1st 15.021 / 堀北真希 (Maki Horikita)
2nd 9.551 / 梅ちゃん先生 (Teacher Umechan)
3rd 6.550 / 堀北真希さん (Ms Maki Horikita)
4th 6.247 / 梅子 (Umeko)
5th 4.135 / 北真希 (Maki Kita)
6th 4.112 / 高橋克実 (Katsumi Takahashi)
7th 3.966 / 南果步 (kaho Minami)
8th 3.848 / 川井憲次 (Kenji Kawai)
9th 3.801 / 竹村 (Takeo)
10th 3.760 / 小田状 (Megumi Koide)

Fig. 4. Sample interaction in QA system (2)

a certain period of time. System focused on a topic and re-retrieved related documents with user interaction but there are more answer candidates then the previous documents. Then, scoring of answers did not work well. In our current implementation, selection of appropriate document set is determined by the number of answer candidates but it is necessary to evaluate this method using more examples.

• Sentence generation for interaction
We applied simple approach in generation of interaction sentence. In user interaction, we use only one clue word in sentence generation. It might be possible to ask a user more than two clue words in one time for effective focusing.

7. Conclusion and Future Work
In this paper, we propose a method to navigate getting a correct answer for QA system using user interaction. When a question is given, QA system will narrow down search area to focus on user’s intended topic of documents. QA system selects a clue word to decide an appropriate topic from topics in retrieved documents and interacts with a user about whether this clue word is appropriate one or not. System will re-retrieve documents using this clue word and expand search space to increase possibility of getting correct answer candidates. We have conducted experiments using ten questions and shown effectiveness of our interaction method in question answering.
Input question: 日本グランプリで優勝したのは誰ですか? (Who won Japan Grand Prix?)
Intended answer: ミハエル・シューマッハ (Michael Schumacher)
query = 日本 (Japan), グランプリ (Grand Prix), 優勝 (won)

***** SCORE and ANSWER *****
1st 14.592 / 小林 (Kobayashi)
2nd 9.207 / セバスチャン (Sebastian)
3rd 7.083 / ルノー (Renault)
4th 6.865 / ファン (Fan)
5th 6.463 / ワリアムズ (Williams)
6th 5.723 / 佐藤琢磨 (Takuma Sato)
7th 5.587 / 鈴木 (Suzuki)
8th 4.790 / マーク (Mark)
9th 4.655 / ポール (Paul)
10th 4.554 / シューマッハ (Schumacher)

フットボールの日本ですか? (yes | no) (Is Japan of football?)
> no.
サッカーの日本ですか? (yes | no) (Is Japan of soccer?)
> no.
F 1の日本ですか? (yes | no) (Is Japan of F1?)
> yes.
query = 日本 (Japan), グランプリ (Grand Prix), 優勝 (won), F 1 (F1)

***** SCORE and ANSWER *****
1st 14.261 / 小林 (Kobayashi)
2nd 6.923 / ファン (Fan)
3rd 6.180 / アaron (Aaron)
4th 5.651 / セバスチャン (Sebastian)
5th 5.305 / ルノー (Renault)
6th 4.378 / 鈴木 (Suzuki)
7th 4.318 / シューマッハ (Schumacher)
8th 4.068 / ジェン (Jen)
9th 4.030 / セナ (Senna)
10th 3.949 / テル (Tel)

In the future, it is necessary to continue experiments using more questions and evaluate our method for improvement. According the experiments, there are some points to improve in selection of query words and categorization of clue words. Moreover, sentence generation method for interaction is simple, then we have much space to improve sentence generation such as inquiry of several clue words. It will be also possible to apply this question answering mechanism for some applications.

References