Solving History Exam by Keyword Distribution: KJP System at NTCIR-11 QALab Task

Yoshinobu Kano
Shizuoka University
kano@inf.shizuoka.ac.jp

ABSTRACT
The QALab task requires to solve the history problems of the Center Exam. Although it seems like a factoid based question-answering problems, we suggest a simple, but fundamentally important, keyword based technique. Regardless of employed methods, the way how to handle the keyword distribution is the fundamental issue to solve the problems. Our system is domain-independent, language-independent, and unsupervised where no training is required. These features would allow direct applications of the system to other types of problems in the future.

Keywords
Yes-No question, question answering, keyword extraction, keyword distribution, knowledge search

Team Name
KJP

Subtasks
Japanese

1. INTRODUCTION
Question-answering techniques could include logic, reasoning, syntactic and semantic analysis. However, almost all of techniques rely on keyword extraction. Given a yes/no answering proposition, we determine yes or no using knowledge sources. The first step of such a determination process is always keyword extraction. Other deeper analyses are based on the keyword extraction results. This implies that keyword distribution in knowledge sources play a critical role. In this paper, we assume that keyword distribution is sufficient to perform the Center Exam question answering tasks to a certain extent of performance.

2. PREVIOUS WORKS AND QALAB
The Todai-Robot project [1] aims to solve university entrance examinations automatically as a challenging task of artificial intelligence. The target examinations include the Center Test, which is the very problems used in this QALab task (Shibuki et al., 2014). We participated the Todai-Robot project and performed the best result for the History subjects in Yozemi Mock Exam Challenge 2013 (Kano, 2014). Our system is almost same as the one used in the Yozemi Mock Exam Challenge.

QALab's dataset was developed from the World History subject of the past Japanese National Center Test questions for University Admissions (Center Test). The Center Test is the common examination for Japanese students when applying to universities. The Center Test asks students multiple-choice style questions. Most frequent types of questions are "select the correct choice" type, "select the wrong choice" type, and "combination" type.

Figure 1 illustrates an example set of choices in the Center Exam. In this example, one of the four choices is the correct one. We added English translations to the original Japanese sentences.

As shown in the figure, domain, location and age could be different in the choices of the same question. Thus, it is not clear for which domain the system should search for the knowledge source. In addition, the expressions used in the questions and those used in the knowledge source are usually different and may be described in several sentences. Furthermore, in the case of wrong choices, there should be no corresponding part in knowledge source. These observations demonstrate that this task is difficult for machines to solve.

The History problems of Center Exam includes other types of questions. We call chronological event sorting type of problems simply as chronological problems. We did not implement solvers for problems where image processing is required.

3. SYSTEM
3.1 Knowledge Source
In the QALab task, preprocessed data of the whole Wikipedia text and high school textbook texts were provided. We only participated to the Japanese subtask where all of propositions are in Japanese. Although we only used Japanese knowledge sources, our system architecture is language independent as described in later sections. We simply call the QALab Japanese subtask as QALab in this paper.

Wikipedia is a typical web sourced knowledge source. However, we decided to use only the textbook data in our system. The reasons are as follows.

First, the questions in the QALab task were taken from the Center Exam questions. Since the Center Exam tries to measure how much the students can solve the questions learned in their high school, the questions are composed of knowledge that is learnable from the high school textbooks.

Second, the structure of high school textbooks is clean. That is, the textbooks tends to use only one place (snippet) for one topic. For example, in the case of history textbooks, one historical event tends to be described in only one place.

Third, the sentences of high school textbooks are usually affirmative sentences. In other words, negative expressions (e.g., "an event did not occur") do not usually appear.

These observations allow us to construct a high-precision system with simple techniques as described in the next section. These observations also answer the issue raised in the previous section, "where to search for knowledge source". The answer is "searching the textbook data for the most relevant part using keyword distributions".

[1] http://21robot.org/

1.	ポルトガルは 12 世紀，神聖ローマ帝国から独立した。 Portugal attained independence from the Holy Roman Empire in the 12th century.
2.	スペイン国王カルロス 1 世は，ポルトガル王を兼ねた。 The King of Spain Carlos I also hold the King of Portugal.
3.	スペインの作家セルバンテスが，『ドン＝キホーテ』を著した。 A Spanish writer Cervantes wrote “Don Quijote”.
4.	グラナダに，ロココ様式を代表するアルハンブラ宮殿が建設された。 At Granada, the Palace of Alhambra was built, which is the hallmark of the Rococo style.

Figure 1. An example of history problems of Japanese Center Exam. No.3 is the correct answer in this example.

3.2 Domain Independent Scoring

Our algorithm is based on the second and third observations. That is, we assume that the answer of a question is described in only one place (snippet) in the textbooks with affirmative expressions.

We also design our system in a domain independent way. While the data we use in our system are small as described above, the high school textbooks have a lot of domains. In order to ensure that our system is applicable to different domains, we design our system to use an unsupervised method for QA.

Our QA system outputs a confidence score w.r.t. the input in the case of select correct/wrong type questions. In other words, let x be the given choice, our system output S(x) as the confidence of x.

Roughly speaking, our system performs (1) keyword extraction from the input, (2) keyword weighting of the input, (3) textbook search and scoring.

(1) Keyword extraction

Because Japanese texts are concatenation of characters not having spaces between words, we apply a morphological analyzer Kuromoji[2], which is based on Mecab[3], to the input. We augmented the dictionary of Kuromoji with the headings of the entries of Japanese Wikipedia. We extract all strings that match with the Wikipedia entries by longest match in the input as the keywords.

(2) Keyword weighting

Let c_i be the frequency of i-th distinct keyword in the input, then the weight of i-th keyword is

$$w_i = 1/(c_i z)$$

In this equation, $z = \sum_i 1/c_i$ is a normalizing constant, where i is defined over the distinct keywords in the input. The frequency c_i was counted over the textbook data.

(3) Textbook search and scoring

We divided the textbook data into snippets. We tried three types of snippets as described later. We search for the snippet that has the highest score w.r.t the input keyword set $\boldsymbol{K}$, which consists of the keywords in the input.

Let $\boldsymbol{R}$ be the word set extracted from a snippet, then the score of $\boldsymbol{R}$ is

$$s_{\boldsymbol{R}} = \sum_{l \in \boldsymbol{R} \cap \boldsymbol{K}} w_l - \sum_{m \in \boldsymbol{K} - \boldsymbol{R}} w_m$$

This expression means that the score of the snippet is the sum of the weights of the input keywords included in the snippet minus that not included in the snippet. If a given choice is correct, keywords in the choice should be densely included in a specific snippet of the textbook; if a given choice is wrong, its keywords should be scattered across snippets. The above equation penalizes such a scattered keyword distribution.

Finally, we regard the maximum $s_{\boldsymbol{R}}$ among all of snippets as the confidence score of the corresponding input.

Our proposed method above does not depend on any domain specific information, even on any specific language.

We solved chronological problems partially using the same system. After finding a relevant snippet, our system tries to find a chronological expression. Then our system sorts the events in the chronological order of the extracted expressions.

4. EXPERIMENTS AND CONCLUSION

Experiments were conducted on the QALab test dataset. Since our system is unsupervised, we did not use the development set.

The QALab dataset was taken from the World History subject. Thus, we used World History’s textbook s as a single knowledge source. As described in the previous section, we used three types of snippets, section, subsection and paragraph, larger to smaller in this order. These text sections were originally explicitly marked in the textbooks.

Our system performed 58 points in the end-end run of Phase I, 53 points in the end-to-end run of Phase II. These results correspond to the subsection type of snippets. These results are higher than other systems, implying our approach is effective to this extent.

5. ACKNOWLEDGMENTS

This work was partially supported by JST PRESTO, MEXT Kakenhi and the Todai Robot Project.

6. REFERENCES

Kano, Y. (2014). Solving History Problems of the National Center Test for University Admissions (in Japanese). *Proceedings of the Annual Conference of JSAI.*

Shibuki, H., Sakamoto, K., Kano, Y., Mitamura, T., Ishioroshi, M., Itakura, K., Wang, D., et al. (2014). Overview of the NTCIR-11 QA-Lab Task. *NTCIR-11 Workshop*.

[2] http://www.atilika.org/

[3] https://code.google.com/p/mecab/