Question Answering System with a vector similarity scoring method

Takuya Kosugi, Hiroya Susuki, Hiroyuki Okamoto, Hiroaki Saito
Department of Information and Computer Science
Keio University
3-14-1, Hiyoshi, Kouhoku-ku, Yokohama, Kanagawa 223-8522, Japan
{kosugi,motch,hxs}@nak.ics.keio.ac.jp

Abstract

We implement and evaluate a question answering system with a vector similarity scoring method. Our question answering system consists of four modules. The question analyzer classifies questions with manually created regular expressions. The document retrieval engine uses vector space retrieval method. The Named Entity extractor finds answer candidates in retrieved articles. Answer selector uses simirality score calculated by the document retrieval engine to decide final answers to present user.

The result of evaluation on NTCIR Question Answering Challenge 2 is 0.242 *in recall,* 0.095 *in precision,* 0.137 *in F-measure and* 0.231 *in MRR*

Keywords: *Question Answering, Vector space search*

1 Introduction

Question Answering(QA) is to retrieve the exact answer for a natural language question rather than presenting the document for artifical keywords. Thus, the purpose of QA is real information retrieaval.

Many researchers have interested in QA. NTCIR held Question Answering Challenge(QAC) task to evaluate QA systems[4]. QA systems are usually consist of some modules. One of these modules is a search engine to get articles which have a correct answer. Many systems use full-text search engine as this module[5, 3, 6]. However, to use full-text search engine for QA is difficult. Making a combination of keywords for search engine are usually done in ad-hoc way.

A vector space method is a document retrieval method which calculats simirality of two documents. Thus, It doesn't have to use ad-hoc way to combinate keywords. We thought this feature is suitable to QA, and similarity score can be used for answer selction.

2 The Structure of our QA System

Our QA system consisits of four modules. Question analyzer, Article retrieval engine, Named Entity(NE) extractor, and Answer selector. The relation of these 4 modules is drawn in Figure 1.
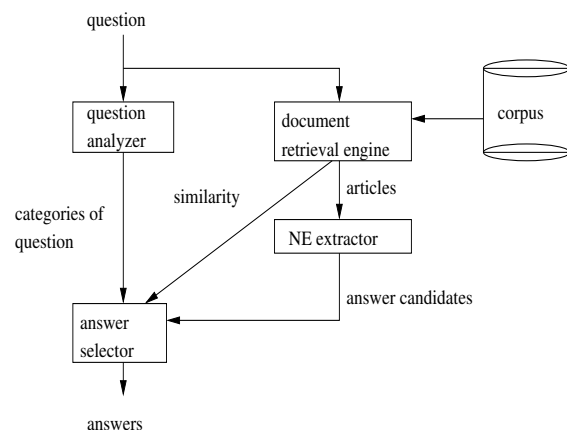

Figure 1. The structure of QA system

2.1 Question analyzer

This module classifies a question into predefined categories. The number of categories are nine. Categories are based on the NE categories on IREX[1]. The NE categories of IREX is 8 category(ORGANIZATION, PERSON, LOCATION, ARTIFACT, DATE, TIME, MONEY, and PERCENT). In Addition to these categories, ninth category NUMBER for general numerical expression are used.

Question classification is done by manually created regular expression rules. If more than two rules match the question, categories which related to matched rules are assigned to a question. When no rule matches a question, it assigins all categories.

The regular expressions for this module are shown in Figure 2.

Figure 2. Regular expressions for question analyzer

2.2 Article retrieval engine

This module retrieves articles which probably have answer expressions. The method of article retrieval is a vector space method. This method is a way of compute similarity of articles. One doucment constructs one vector in word-article space, thereofore this method can compute the similarity of a question vector and a To construct article vector, article analyzed by morphological analyzer

We use ChaSen to analyze article[8]. For Vector Space searching method, weighting vector is important. We used wighting by $tf\text{-}idf$, since it is most popular way to weight vector.

The similarity of a question vector and an article vector is measured by cosine measure. And top 5 articles ranked by cosine measure are passed to next module(NE extractor).

2.3 NE extractor

This module extracts answer candidates. In many cases, answer candidates are Named Entity. Thus, to find answer candidates is extracting named entities from retrieved documents. We use CaboCha as NE extractor[7]. Although CaboCha is a Japanese dependency structure analyzer, CaboCha can also extract NE. When Cabocha extracts NE, it uses 8 categoris of NE defined by IREX. We add a extraction rule for NUMBER category, which is a very simple way that extracts all numerical expression.

2.4 Answer selector

This module scores each answer candidate extracted by the NE extractor and selects top 5 NEs to present as answers. First, this module matches question categories classified by the question analyzer and NE categories extracted by NE extractor. Second, answer candidates which matches question categories scored by scoring method uses similarity score of the Article retrieval engine. The scoring method is (1).

$$\sum_{i=0}^{N} c_i(1 + \log \sum_{j=0}^{M} w_{ij}) \qquad (1)$$

c_i is similarity score of the ith article which retrieved by article regrieval engine, w_{ij} is jth answer candidate, N is the number of articles, and M is the number of answer candidates.

The highest 5 answer candidates are present as answer.

3 Result

This section describes the result of all 200 questions in NTCIR QAC-2.

3.1 Question analyzer

We evaluated the handcrafted correct classification and the output of the question analyzer. A handcrafted result classified 153 questions. 47 questions can't be classified into any categories.

The result of the question analyzer is shown in Table 1. This result is removing not match all patterns.

Table 1. Result of the question analyzer

category	# of question	# of classified article	# of correct doc.	recall	precision	F measure
ORGANIZATION	18	45	11	0.61	0.24	0.34
PERSON	47	44	44	0.94	1.00	0.97
LOCATION	36	51	35	0.97	0.69	0.81
ARTIFACT	20	8	3	0.15	0.38	0.21
DATE	8	10	8	1.00	0.80	0.89
TIME	3	1	1	0.33	1.00	0.50
MONEY	2	3	2	1.00	0.67	0.80
PERCENT	0	0	0	N/A	N/A	N/A
NUMBER	19	15	12	0.63	0.80	0.70
total	153	177	116	0.76	0.66	0.71

3.2 Article retrieval engine

Top 5 articles retrieved by the article retrival engine has a correct answer for 164 questions. The number of articles which has an answer with the higest rank is shown in Table 2.

Table 2. The number of documens which have a correct answer

rank	Mainich 98	Mainichi 99	Yomiuri 98	Yomiuri 99
1	74	61	57	49
2	18	21	13	13
3	7	6	6	8
4	6	7	4	7
5	0	5	6	5

According to this result, When the article retrieval engine could retrieve the article which has a correct answer, 80 % of the articles has the correct answer in top 2 articles.

3.3 NE extractor

The score of NE extractor is evaluated by Yamada et.al. The result of evalution is shown in Table 3.

Table 3. The result of NE extraction[9]

Named Entity	F CM(CaboCha)
ARTIFACT	0.471
DATE	0.922
LOCATION	0.825
MONEY	0.943
ORGANIZATION	0.790
PERCENT	0.942
PERSON	0.863
TIME	0.832
total	0.832

3.4 Answer selector

To evaluate only answer selctor, question which retrieaval engine couldn't find articles have correct answer are eliminated. The result is shown in Table 4.

Table 4. The result of the answer selector

category	# of questions	MRR
ARTIFACT	3	0.7
DATE	8	0.2
LOCATION	35	0.23
MONEY	2	0.5
ORGANIZATION	11	0.20
PERCENT	0	0
PERSON	44	0.41
TIME	1	0.3
NUMBER	12	0.17
total	116	0.32

3.5 Whole system

The result of the whole system are shown Table 5.

Table 5. NTCIR QAC2 result

correct	95
recall	0.242
precision	0.095
F measure	0.137
MRR	0.231

4 Discussion

In this section, we discuss the result of each module.

4.1 Question analyzer

The result of the question analyzer affects answer selection. Thus, this result is very important for system. In the experiment, recall is high, but precision is

so low. These causes are many questions don't match regular expression pattern. Two reasons are considered to this result.

- The pattern for classifying a question is not enough

- The question itself can't be classified into the predifined category

The number of the question which doesn't match pattern in the question analyzer are 74. And 47 questions can't be classified by the manual classification. Therefore, new categories for these questions are needed.

4.2 Article retrieval engine

The article retrieval engine could retrieve articles which have a correct answer for 80 % of questions. And 80 % of them are ranqked top 2.

Some qutestions couldn't be retrieaved correct articles has misleading information. For example, QAC2-10131-01: "If the worlds longest bridge is the Second Lake Pontchartrain Causeway in the United States, what is the worlds longest bridge with a railroad?" has many unneccesary information. These information make the article retrieval engine retrieve wrong articles. To avoid this probrem, a system eliminating misleading information is needed.

The another cause of retreiving wrong articles is that a question is too short to have ehough information for retrieval. QAC2-10080-01: "What does flugels of the Yokohama Flugels mean?" has only two content words; "flugels", and "Yokohama". Questions which have few information tend to make the article retrieval engine find unrelated articles. To find correct articles from few information, it is thought that filtering like syntactical structure matching is effective.

To make better the article retrieval engine, it is thought that use sophisticated retrieval method. For examle, to use Latent Semantic Indexing(LSI) instead of simple vector space method is good idea. LSI is a way of demensionality reduction on vector space retrieaval[2], since vector space retrieval with LSI can be exploited co-occurence.

4.3 NE extractor

This module only uses existing NE extractor, and the result of it is Table 3. The result is good except ARTIFACT. ARTIFACT is a Named Entity which express a title of drama, literary work and prise. In many cases, these expressions are unknwon words in morphorogical analysis. An unknwon word has few information to other morpheme, thus many classification failure is occured.

The number of category is few for QA. This catgories are based on Named Entity task of IREX. Thus, fine grained categories for QA are effective. But to use new categories for CaboCha, learning by corpus with new categories are needed.

4.4 Answer selector

The Scoring method has some probrems. First, the NEs which occur in a question sentence have high scores. In many cases, NEs in a question sentence are not correct answers. The simplest way of avoiding this probrem is eliminate NEs in a question from answer candidates. Some NEs occure in a question can be a correct answer.

Second, This scoring method give high score to an answer candidate which occure in two or more articles. There are many articles in '98 and '99, since QAC2 uses newspaper in '98 and '99 as corpus. There are few articles in other year. Therefore, if a question has not asked news in '98 or '99, the system presents wrong answers frequently.

Third, this task must show an article ID with an answer. Our system get an article ID from top ranked article which has same expression of answer. Thus, although the system could find correct article, it may show a wrong article ID. The scoring method that uses articles are difficult to show one article ID as an evidence.

5 Conclusion

We implemented and evaluated question answering system which selects answers based on similarity computed by article retrieave engine.

To use similarity socre is simple way to select answer. And This system has MRR 0.231.

References

[1] Irex(ne)
. http://www.csl.sony.co.jp/person/sekine/IREX/NE.

[2] H. S. Christpher D. Manning. *FOUNDATIONS OF STATISTICAL NATURAL LANGUAGE PROCESSING.* MIT Press, 1999.

[3] J. Fukumoto, T. endo, and T. Niwa. Ritsqa: Ritsumeikan question answring system used for qac-1. *Working Notes of the Third NTCIR Workshop Meeting. PartIV: Question Answering Challenge*, pages 113–116, October 2002.

[4] J. Fukumoto and T. Kato. Question answering challenge (qac-1) an evaluation of question answering task at ntcir workshop 3. *Working Notes of the Third NTCIR Workshop Meeting. PartI: Over*, pages 77–86, October 2002.

[5] A. Ikeno and H. Ohnuma. Oki qa system for qac-1. *Working Notes of the Third NTCIR Workshop Meeting. PartIV: Question Answering Challenge*, pages 17–20, October 2002.

[6] K. Kawata, H. Sakai, and S. Masuyama. A question and answering system using newspaper corpus as a knowledge base. *Working Notes of the Third NTCIR Workshop Meeting. PartIV: Question Answering Challenge*, pages 25–29, October 2002.

[7] T. Kudo and Y. Matsumoto. Japanese dependency analysis using cascaded chunking. In *CoNLL 2002: Proceedings of the 6th Conference on Natural Language Learning 2002 (COLING 2002 Post-Conference Workshops)*, pages 63–69, 2002.

[8] Y. Matsumoto, A. Kitauchi, T. Yamashita, Y. Hirano, H. Matsuda, K. Takaoka, and M. Asahara. *Japanese Morphological Analysis System ChaSen version2.2.1 Manual*. 2000.

[9] H. Yamada, T. Kudo, and Y. Matsumoto. Japanese named entity extraction using support vector machine. *Transactions of IPSJ(in Japanese)*, 43(1):44–53, 2002.