Rits-QA: List answer detection and Context task with ellipses handling

Jun'ichi FUKUMOTO, Tatsuhiro NIWA, Makoto ITOIGAWA and Megumi MATSUDA
Ritsumeikan University
1-1-1 Noji-higashi, Kusatsu-shi, Shiga 525-8577 JAPAN
fukumoto@media.ritsumei.ac.jp, {t_niwa, m_itoigawa, m_matsuda}@nlp.is.ritsumei.ac.jp

Abstract

In this paper, we describe RitsQA: Ritsumeikan question answering system. We have improved our QA system submitted in QAC1 in subcategorization of NE type and question type analysis and in handling numerical expressions using corpus based approach. We also participated subtask 2 and 3 using these core QA systems. We have implemented list answer detection module for subtask 2 and ellipsis handling module for subtask 3.

We had evaluations of subtask 1, 2 and 3. According to Formal Run evaluation, our QA system had some progress in some aspects. Especially in subcategorization of NE types, we have obtained better improvement in performance of core QA system. In subtask 2 and 3, we have a little success in detection of the number of answers from surface pattern of question sentence. Also, in ellipsis handling, almost 37% of follow-up questions was correctly treated in ellipsis. However, it is necessary to expand patterns of ellipsis and improve list answer processing.

1 Introduction

We have participated QAC1 of NTCIR Workshop 3 and had evaluation of our QA system[1] [2] [3] [4]. Our system strategy to extract answer was very simple and standard one. We firstly classified question type into five kinds and prepared about 70 patterns to analyze an input question. Then we extracted answer candidates using Named Entity tagger NExT [6] and calculate score of answer candidates based on word distance between answer candidates and clue words extracted from question sentence. Higher scored answer candidates are selected for final answers.

In the analysis of QAC1 evaluation, we found several points to improve our QA system. The about 70 question patterns and five kinds of question types were not enough to cover all question expressions used for QAC1 test set. Also, our QA system can not handle numeric expressions well because of less patterns for numeric type questions. Moreover, there was some bad points on system interface between question analysis and answer detection modules, for example, different handling strategy of Named Entity.

In QAC2, we will expand question types and their patterns to recognize various patterns of question sentences and classify Named Entity categories in more detail to handle expanded question types and patterns. In order to handle numeric type questions, we have introduced corpus-based approach using co-occurrence data of unit expressions and their object names. Moreover, we use the above improved QA system as a core system for list answer extraction of subtask2 and context task of subtask 3.

For list answer extraction, we use top twenty answer candidates and select correct answers using difference of scores and their position in a document. We also use some specific expressions which can indicate the number of answers. For context task, we complete sentence expressions of a series of questions from the first or an appropriate question. We found some ellipses handling patterns from the analysis of a series of questions.

2 Improvement of core QA system

We have submitted two systems for all subtasks of QAC2. One system used detailed information of Named Entity types for question analysis and answer candidate detection. In the other system, we have improved mechanism to handle numeric expressions of question sentences and target documents.

2.1 Detailed classification of Named Entites

According to the analysis of QAC1 test set and their target documents, we classified information of Named Entities in details. We have classified original Named Entity types used for IREX [5] into several subtypes as follows:

- PERSON
  F_NAME(FAMILY_NAME),
  G_NAME(GIVEN_NAME), POSITION

© 2004 National Institute of Informatics

- LOC(LOCATION)
  COUNTRY, CAP(CAPITAL), STATE, PREFEC(PREFECTURE), CITY, REG(REGION)
- ORG(ORGANIZATION)
  ACAD(ACDEMIC), GOV(GOVERNMENT), CORP(CORPORATION), BANK, COM(COMPANY), INS(INSURANCE), MED(MEDICAL), JUD(JUDICIAL), JOUR(JOURNAL), TEMPLE, HOTEL
- NUM(NUMERICAL)
  AGE, RATE, MONEY, WEIGHT, LENGTH, VEL(VELOCITY), AREA, VOL(VOLUME), TEMP(TEMPARTURE), HUMID(HUMIDITY), RANK, FREQ(FREQUENCY), SERISE, NoP(NUMBER of PERSON), NoT(NUMBER of TIMEs)
- TIME
  HOUR, MIN(MINUTE), SEC(SECOND), DURATE
- DATE
  ERA, CENT(CENCURY), B_YEAR(BUSINESS_YEAR), MONTH, WEEK, DAY, DURATE
- PROPER NOUN
  RAIL, STA(STATION), TRANSPORT, ANI(ANIMAL), INST(INSTRUMENT), NAT(NATURE) {MT(MOUNTAIN), RIV(RIVER), LAKE, STAR }

2.2 Numeric expression handling

In order to handle various kinds of numeric expressions, we used co-location between unit name and numeric type. We classified numeric expression into eighteen types. We will show classified types and their unit names of these types in Table1.

When there exists inquired some numeric unit name in a question sentence, QA system utilize this information to extract answer candidates. For example, in a question sentence of " (how many persons)" these is unit name of person (person) will be appropriate clues to search answer candidates. However, there is a case that there is no unit name in a question sentence. For example, in a question sentence "(How is the height of Giant Baba?)", there is no unit name for the height of Giant Baba. In this case, the topic word "height" will be clue to identify suitable unit name of the height of person name.

We will recognize topic word to identify type of unit in the following procedure.

Table 1. Classification of numeric types

category name	unit name
MONEY	(yen), $, (Euro)
LENGTH	m (meter), km (kilo-meter)
WEIGHT	g (gram), (ton)
AREA	h (hectare), km^2
VOLUME	(liter), cc
CAPACITY	GB (giga-bute), MB (mega-byte)
AGE	(years)
NPERSON	(person), (person)
PERCENT	(1/10), %, (percent)
CENTURY	ł (century)
YEAR	(year), (fiscal year)
MONTH	(month)
DAY	(day)
HOUR	(hour)
MINUTE	(minute)
SECOND	(second)
DURATION	(month) (days)
OPTIONAL	(kilo)
NUMBER	other numeric units such as (hon), (times)

1. the nearest noun or adjective which have Japanese topic marker " (ha)" in a question sentence.
2. If there is no such marker in a question sentence, the nearest noun or adjective to interrogative expression in a question sentnece.

We will select topic word in the above procedure, then unit name which co-occured with the topic word will be clue to identify correct answer for a given question. We have developed co-occurence data, which are co-occurence of unit name and noun, and unit name and ajdective with ten word window size, from five years Mainichi newspaper articles.

Table2 shows co-occured unit name with the word " (height)".

Table 2. Co-occurred unit name with the word "ÈÄ„"

unit name	frequency
(centi)	2013
(years)	529
(kiro)	280

According to the above table, unit name " (centi) " will be selected as the most frequent unit name. Then, numeric type "LENGTH" will be selected and numeric

expressions which are categorized in this type will be extracted from Newspaper articles (target documents).

3 List answer extraction

We will use top twenty answer candidates which are extracted by core QA system for a question sentence. In order to recognize correct answers and incorrect answers in answer candidates we two kinds of scoring method with score of answer candidates which are given by core QA system. The first scoring method is to weight answer candidates in the same sentence. If answer candidates exist in the same sentence, score of the candidates will be 1. The second one is to weight answer candidates which locate in nearer position. Such answer candidates are grouped by comparing word distance between these candidates. Grouped candidates will have score of 1. Final score of an answer candidate will be sum of score given by core QA system and scores calculated by two kinds of methods.

Correct answers in a list task will be recognized by difference of scores in ranked order. If top two scores are very similar but the third score has difference with the second onethen the top two answer candidates will be an answer list. Moreover, the number of answers will be identified by surface expressions of a question sentence. For example, " (three animals)" indicates that the number of answers is three. " (combi)" indicates two and " (who and who)" indicates two.

4 Ellipses handling in Context task

Our approach for context task is to complete a series of questions which have ellipses from the previous sentence. According to the analysis of QAC1 test set, we have classified ellipsis patterns into three kinds. We assumed that main inquiry part of a question sentence consists of topic presentation, adjective expression of topic part and inquiring part. For example, in a question sentence " (Who is the president of America?)", topic presentation part is " (the president)", adjective expression part is " (of America)", and inquiring part is " (who is)".

Three kinds of ellipsis types are as follows:

1. Ellipsis of adjective expression

   In the next question, adjective expression is omitted. For example, in the next question " (Where is the captal city?)", adjective expression " (of America)" is omitted. So, it is necessary to put this part to handle the next question independently.

2. Ellipsis of topic presentation part

   In the next question, topic presentation part is omitted. For example, in the next question " (Who is for France?)", topic presentation part " (the president)" is omitted.

3. Replacement with pronoun

   In the next question, pronoun is used to indicate some element in the main question. For example, the next question " (Where is his birth place?)", pronoun " (his)" indicates an answer of the first question.

For each ellisis type, we analyzed features of main sentence and its subsequent sentence using QAC1 test set. The points of analysis are ellipsis element, question type, existence of pronoun, and ellipsis element.

1. In type 1, ellipsis part is adjective expression and question types are different in both sentences. Also there is no pronoun use for anaphora. In this type we will put adjective expression of the main sentence to its subsequent sentence.

2. In type 2, ellipsis part is adjective expression and question types are the same in both sentences. Also there is no pronoun use for anaphora. In this type we will put topic element of the main sentence to its subsequent sentence.

3. In type 3, there is no ellipsis part in the subsequent sentence but pronoun is used to identify antecedent of the main question. In this case, we can infer what kinds of information is ellipsis element from type of pronouns such as he, its and so on.

Ellipsis part will be identified by the above strategy. However, in QAC2 suntask DryRun, there were several follow-up questions for one main question. One question set in QAC2 consists of 5 to 10 sentences. So, it is necessary to identify which sentence will be target sentence to ellipsis handling. We have analyzed QAC2 Dry Run test set for subtask 3 to investigate which sentence will be anticedent for a followup-up question sentence. We assumed that anticedent sentence will be the first question or previous sentence. In type 1 and 2, almost 80% questions indicate the first question, and in type 3, almost half sentences indicate the previous sentence. Therefore, in our current inplementation, we have followed this analysis results for ellipsis handling.

5 Results

We have submitted two kinds of systems for all subtasks. We developed two kinds of core QA systems and evaluated them for subtask 1. We also used these two systems as core QA system for the other two subtasks, subtask 2 and 3. The results of subtask 1, 2 and

Table 4. Results of Subtask 2

SysID	Output	Correct	Recall	Precision	MF
SYS-2124	539	90	0.140	0.167	0.164
SYS-2224	416	61	0.095	0.147	0.122

Table 5. Results of Subtask 3

Formal Run

SysID	Total	First	Rest	B	G	SG
SYS9	0.094	0.188	0.079	0.085	0.091	0.131
SYS13	0.057	0.125	0.046	0.039	0.071	0.030

Reference Run

SysID	Total	First	Rest	B	G	SG
SYS9	0.150	0.188	0.143	0.175	0.122	0.227
SYS13	0.100	0.125	0.096	0.134	0.087	0.091

Table 3. Results of Subtask 1

SysID	MRR	Rate.1st	Rate.5th
SYS-1124	0.311	0.236	0.436
SYS-1224	0.189	0.128	0.287

3 will be shown in Table 3, Table 4 and Table 5, respectively. (All the results are taken from task results announcement in March 2004 from Task organizers.)

In each Table, the upper system used detailed classification of NE as core QA system and the lower system used numeric expression handling as core QA system. All the results obtained by core QA system using classified NE types got better results than the results of QAC1. So, classified NE types work well in this evaluation. All the results obtained by core QA system using numerical expression handling did not have much progress. It is because the number of questions for numerical values was not so much like the QAC1 test set. However, according to the detailed analysis of the results, there was much progress in questions of numerical expressions. In incorrect questions, most of the reason of incorrectness was in information retrieval. In prepared documents by human, answer extraction worked well and correct answers were extracted correctly.

In list answer extraction (subtask 2), we have evaluated 109 questions in which core QA system returned more than one answers among 200 questions of QAC2. There were 58 questions of multiple answers and average F-measure was 0.510 in these questions. We have not analyzed these results indetail but we can not conclude our approach of list answer extraction work well.

In context task (subtask 3), we have evaluated how ellipsis handling work well for follow-up questions. There were 215 follow-up questions for 36 related question sets.We compared completed questions by our method and human completion. If a completed questionby our approach is semantically acceptable, this completion will be correct one. Among 215 questions, 79 questions (36.7%) were correctly completed. Therefore, in prepared patterns of ellipsis, our approach worked well but it was not sufficient.There were several reasons of incorrectness. Some pronouns in questions sentence referred an event described in the previous sentence. There were some other pattern of ellipsis and some of them were complex. It is necessary to analyze QAC2 test set and other test set in detail to prepare more patterns for ellipsis handling.

6 Conclusion

In this paper, we described improvement of our QA system and system construction for subtask 2 and 3 using our core QA system. In QAC2, we will expand question types and their patterns to recognize various patterns of question sentences and classify Named Entity categories. Moreover, in order to handle numeric type questions, we have introduced corpus-based approach using co-occurrence data of unit expressions and their object names. We use the above improved QA system as a core system for list answer extraction of subtask 2 and context task of subtask 3.

For list answer extraction, we use top twenty answer candidates and select correct answers using difference of scores and their position in a document. We also use some specific expressions which can indicate the number of answers. For context task, we complete sentence expressions of a series of questions from the

first or an appropriate question. We found some ellipses handling patterns from the analysis of a series of questions.

According to Formal Run evaluation, our QA system had some progress in some aspects. Especially in sub-categorization of NE types, we have obtained better improvement in performance of core QA system. In subtask 2 and 3, we have a little success in detection of the number of answers from surface pattern of question sentence. Also, in ellipsis handling, almost 37% of follow-up questions was correctly treated in ellipsis. However, it is necessary to expand patterns of ellipsis and improve list answer processing.

References

[1] J. Fukumoto, T. Kato and F. Masui. Question and Answering Challenge (QAC-1) : Question answering evaluation at ntcir workshop 3. In *Working Notes of the Third NTCIR Workshop Meeting Part IV: Question Answering Challenge (QAC-1)*, pages 1–10, 2002.

[2] J. Fukumoto, T. Kato and F. Masui. Question and answering challenge (QAC-1) : Question Answering Evaluation at NTCIR workshop 3. In *AAAI 2003 Spring Symposium New Direction in Question Answering*, pages 122–133, 2003.

[3] J. Fukumoto, T. Endo and T. Niwa. RitsQA: Ritsumeikan question asnwering system used for QAC-1 In *Working Notes of the Third NTCIR Workshop Meeting Part IV: Question Answering Challenge (QAC-1)*, pages 113–116, 2002.

[4] NTCIR (NII-NACSIS Test Collection for IR Systems) Project http://research.nii.ac.jp/ntcir/index-en.html

[5] Information Retrieval and Extraction Exercise (IREX) http://cs.nyu.edu/cs/projects/proteus/irex/

[6] F. Masui, S. Suzuki and J. Fukumoto. Named Entity Extraction Tool NExT for Text Processing In *Proceeding of The Eighth Annual Meeting of The Association for NLP*, pages 176-179, 2002. (in Japanese) http://www.ai.info.mie-u.ac.jp/ next/next.html