Query Expansion for Contextual Question Using Genetic Algorithms

Yasutomo Kimuraand[1] and Kenji Araki[2]

[1] Otaru University of Commerce 3-5-21, Midori, Otaru, 047-8501 Japan
kimura@res.otaru-uc.ac.jp
[2] Graduate School of Information Science and Technology, Hokkaido University,
Kita 14 Nishi 9, Kita-ku, Sapporo-shi, 060-0814 Japan
araki@media.eng.hokudai.ac.jp

Abstract. We propose a query expansion method using Genetic Algorithms(GA) in Japanese. Recently, question answering research focuses on contextual questions. Therefore a question answering system has to resolve contextual problems by using both previous questions and previous answers. This problem is largely related to query expansion because of the need to find new keywords. In the contextual processing, a query needs to find other suitable keywords from related resources. Although it is easy for a system to find related words, it is difficult to find a suitable combination of keywords. GA is better suited for a combination problem just like a knapsack problem. Therefore we apply GA to our contextual query expansion method. In the evaluation experiment, MRR was 0.2531 in 360 contextual questions. We confirm the MRR of our method is higher than that of the baseline. We illustrate our method and the experiment.

1 Introduction

Recently, question answering tasks focus on the problem of retrieving the appropriate answer rather than retrieving document lists (Ellen. M. Voorhees. et al. 2000). This task has recently been evaluated in TREC-QA (Ellen. M. Voorhees. 2004), NTCIR-QAC (Tsuneaki Kato et al. 2004) and so on. TREC Question answering was the first large-scale evaluation of domain independent question answering systems. QAC is a challenge encouraging the evaluation of question answering technology in Japanese. Such systems that answer isolated factoid questions are the most basic level of question answering technology. Question answering systems are used interactively to answer a series of related questions, whereas in the conventional setting, systems answer isolated questions one at a time. Question answering systems generally reply about location, organization, date and so on. Recently question-answering systems are able to correctly answer isolated question. Moreover, it is often said that QA tasks appear simple by not including the why-question types and the how-question types. However a series of related questions is more difficult than isolated questions. Question answering systems have to interpret a given question within the context of a specific

dialogue. In the case of a series of related questions, QA systems have to include the context processing abilities of systems such as anaphora resolution and ellipses handling. While, Murata et al. have obtained high accuracy results by adding keywords instead of using anaphora resolution and ellipses handling, these methods sometimes can not find candidate answers using context processing (Masaki Murata et al. 2005). These methods have two problems. First, not every result includes a correct answer even if a system fills the ellipsis correctly. Secondly, a suitable query can not always be generated from the words that composed the question. To resolve these problems, we consider that a question answering system makes a query using related keywords (Xu, Jinxi et al. 1996). We also assume that the keyword exists near the IR results which are from related questions that have been previously retrieved and therefore context processing is not necessary. From here, the extraction of keywords would follow this process.

- To generate a query from the first related question which does not need context processing.
- To find sentences which include noun words which correspond to inputted words.
- To extract noun words from the extracted sentences.
- To make keyword candidates from the extracted words.

This process is similar to pseudo-relevance feedback. Pseudo-relevance feedback is generally used by query expansion (Chen, Hsinchun et al. 1998). However in addition to keywords used in the query there exist many candidates. Moreover, we have to retrieve information to check for the best query. Therefore confirming the optimum keyword combination for a query takes much time. We propose a contextual question answering system using Genetic Algorithms because GA is suitable for finding the optimum combination 1989. GA is applied to make keyword combinations in contextual question answering methods.

2 Basic idea

In natural language processing, GA has not been adapted much because it is difficult to configure a fitness function. For example with knapsack problems in GA 1989, the fitness function is simple. The knapsack problem is a problem in combinatorial optimization. This problem task gives a set of items, each with a cost and a value, then determines the number of each item to include in a collection so that the total cost is less than some given cost and the total value is as large as possible. Chen et al proposed GA method for Information Retrieval (Chen, Hsinchun et al. 1998). In a similar way, GA of our method uses a vocabulary from natural genetics. In our reasearch GA finds not the correct answer but a suitable query. The suitable query is a combination of keywords. In the case of isolated questions, a question answering system usually finds a suitable combination of keywords from the isolated question's words. On the other hand, a contextual question answering system has to add more related

question words, previous answers and a high number of relevant words. This means a suitable query is selected by a combination of many keyword candidates. In addition, GA is better suited for an optimization problem just like a knapsack problem. Therefore we apply GA to our contextual question answering system.

3 Process

3.1 Overview

In the case of isolated questions, a query is usually generated by a combination of question words. In the case of series type questions, the first question does not usually have to handle ellipses. From the second question which handles ellipses, our system generates a suitable query using GA. Figure 1 shows an overview of process.

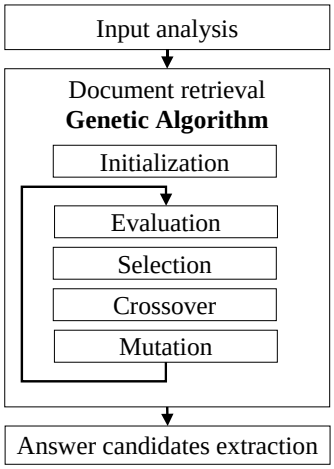

Fig. 1. Overview of GA process.

Input analysis is performed extracting a question type and keywords. We focus on document retrieval and passage retrieval GA is applied. In the following sections, we will explain the details of our method.

4 Query expansion using GA

Related contextual question answering systems focus on anaphora and ellipsis resolution. However, a system sometimes can not extract correct documents by correct ellipsis resolution. In other words, even if a person makes a relevant query to question, search engines sometimes can not find any correct documents with

the correct query. After the first question in a series question type, we consider that it is possible to make a suitable query from many related keywords just like neighbor words near question keywords in retrieved documents. Therefore our system tries combining related keywords using GA. In our GA, an answer format consists of keywords for searching. We consider that most suitable answers consists of words which are extracted from highly similar sentences with the question.

4.1 Initialization

Query candidates are as follows:

- Question's words
- Related words which compose a previous question.
- Related words which compose high similar sentence to the question.

A query randomly composes two or three keywords from keyword candidates because we made a preliminary experiment about number of keywords.

Initialization is as follows:

- To focus on questions without contextual processing.
- To generate a query from the question.
- To extract sentences which includes keywords of the query from retriever documents.
- To make a keyword candidates list from the extracted sentences.

4.2 Evaluation of fitness value

Fitness value means a measure for checking adaptability to the individual. In this paper, GA finds not the correct answer but a suitable query. In the case of finding a suitable query, we need information retrieval results for evaluation. If we could not confirm whether the query is the best query or not, we need to try all query patterns. A system has to check contextual questions many times compared with isolated questions. In our system, a fitness function mainly includes three elements which consist of keywords, related measure and possibility of document retrieval. First, our system randomly generates combinations of a few keywords. A combination of keywords is evaluated by sum of each keyword's IDF. Our system searches these queries in descending order until finding 10 successful queries.

Our system extracts related sentences using the keyword type. Keywords are divided into three types which consist of contextual keywords, question keywords and query keywords. A related evaluation function is shown as follows:

$$Fitness = Con + Inp + 0.5 \cdot Qry \tag{1}$$

$$Con = \frac{Sum\,of\ contextual\ keyword's\ IDF}{Number\,of\ contextual\ keywords} \tag{2}$$

$$Inp = \frac{Sum\ of\ question\ keyword's\ IDF}{Number\ of\ question\ keywords} \quad (3)$$

$$Qry = Sum\ of\ query\ keyword's\ IDF \quad (4)$$

4.3 Selection

A proportion of the existing population is selected to breed a new generation. Selection means to extract the best solutions in the existing population. In our method, the population consists of individual queries. Each query is evaluated by a fitness function. The important assumption of the selection is to obtain document candidates by the query.

4.4 Crossover

Crossover is a genetic operator used to vary the programming of chromosomes from one generation to the next. Two chromosomes which obtain high fitness value randomly are selected as parents. Our system generates new chromosomes(queries) by the crossover of two chromosomes.

4.5 Mutation

The purpose of mutation is to avoid local minima by preventing the population of chromosomes from becoming too similar to each other. Mutation means to generate new chromosome by changing randomly. Our system uses words which have never existed in tried queries in words of population and the random ratio is 2%.

4.6 Generation

A generation consists of selection, crossover and mutation. In each generation, the fitness of the whole population is evaluated. Individuality is evolving with generation. However it takes much time with trying many generations. Therefore, our system tries three generations.

5 Evaluation experiment

In this experiment, we evaluate the contextual query expansion method by GA. the baseline is made up of keywords in a series and we try to two GA methods which consist of First Question GA(FQGA) method and Each Question GA(EQGA) method. FQGA generates query candidates based on retrieved documents from the first question of a series. EQGA generates query candidates based on retrieved documents from previous question.

A Question set consists of 360 questions in NTCIR5-QAC3 (Tsuneaki Kato et al. 2004). This question set consists of 50 series question set which includes 35 gathering series and 15 browsing series. Corpus are 2 years of newspaper articles

from the Yomiuri newspaper and the Mainichi newspaper. The Mean Reciprocal Rank(MRR) is used as the evaluation measure (Ellen. M. Voorhees. et al. 2000). The top 20 answers for each instance were considered by the MRR. The result of MRR is shown in Table 1.

Table 1. Experiment.

Method	MRR
Baseline	0.0316
First Question GA(FQGA)	0.2531
Each Question GA(EQGA)	0.1448

6 Consideration

This question set includes 35 gathering series and 15 browsing series. Gathering type questions are concerning a common global topic. For example, a gathering series shows as follows:

Q1 What genre does the "Harry Potter" series belong to?

Q2 Who is the author?

Q3 Who are the main characters in that series?

On the other hand, browsing type questions is more difficult than gathering type questions because each consecutive question shares a local context. For example, a browsing series shows as follows:

Q1 Where was Universal Studios of Japan constructed?

Q2 Which train station is the nearest?

Q3 Who was the actor that attended the ribbon-cutting ceremony on the opening day?

Q4 What is the name of the movie he features in that was released in the New Year season of 2001?

Q5 What is the name of the movie starring Kevin Costner released in the same season?

In this experiment, the FQGA method had the highest performance of the three methods. We confirmed the effectiveness of the FQGA method. Although the EQGA method was lower than the FQGA method, the EQGA method was better suited for browsing series. We found 11 questions that the EQGA method could correctly answer although other systems could not answer. In these questions, we confirmed that 7 questions were the gathering type.

7 Conclusion

In this paper, we described resolving contextual questions using GA. In contextual questions, we tried to handle ellipses by adding keywords in retrieved documents. Our system found a suitable combination of words in retrieved documents. In the experiment using NTCIR5-QAC3 data, we confirmed a high evaluation.
In future work, we will apply it to other QA.

References

Ellen. M. Voorhees.: Overview of the TREC 2004 Question Answering Track. . (2004) –

Ellen. M. Voorhees and Dawn M. Tice: Building a question answering test collection. In The 23 International ACM SIGIR Conference on Research and Development in Information Retrieval. **78** (2000) 200–207

Tsuneaki Kato, Jun'ichi Fukumoto, Fumito Masui and Noriko Kando.: Handling Information. Access Dialogue through QA Technologies – A novel challenge. HLT-NAACL2004 Workshop on Pragmatics of Question Answering. (2004) 70–77

Masaki Murata, Masao Utiyama and Hitoshi Isahara.: Japanese Question-Answering System Using Decreased Adding with Multiple Answers at NTCIR 5. NTCIR Workshop 5 Proceedings, (2005) 380–385 2005

D.E.Goldberg: Genetic Algorithms in Search, Optimization, and Machine Learning., Addison-Wesley (1989)

Tatsunori Mori and Shinpei Kawaguchi. : Answering Contextual Questions Based on the Cohesion with the Knowledge. In Proceedings of the Fifth NTCIR Workshop Meeting, (2005) 386–393.

Chen, Hsinchun and Shankaranarayanan, Ganesan and She, Linlin and Iyer, Anand A Machine Learning Approach to Inductive Query by Examples: An Experiment Using Relevance Feedback, ID3, Genetic Algorithms, and Simulated Annealing Journal of the American Society for Information Science **49(8)** (1998) 693–705.

Xu, Jinxi and Croft, W. Bruce Query Expansion Using Local and Global Document Analysis, Proceedings of the 19th Annual ACM-SIGIR Conference, (1996) 4–11.