

Political Question Generation Method Using Minutes of Municipal Councils

Yasutomo Kimura¹, Hideyuki Shibuki², Keiichi Takamaru³
Tetsuro Kobayashi⁴ and Tatsunori Mori²

Abstract. This paper proposes a generation method of political questions which include Japanese noun phrases of the form N_1 no N_2 . We focus on generating a yes-no question whether a user is interested in an utterance of a councilor. An example of political questions shows as follows: *nisanakatanso(N₁) no haisyutu(N₂) nituite kuwasiku siritai desu ka?* (Do you want to know emissions of carbon dioxide?). In this paper, " N_1 " and " N_2 " mean a noun or a compound noun. The form N_1 no N_2 which include the Japanese postpositional "no" has a much broader usage. These expressions of the form N_1 no N_2 cannot often use a question expression for asking the interesting such as "*simin no minasan*(All of citizen)". Therefore, our method checks on an expression whether a user can understand the meaning of that as a question expression. Moreover, our method expands the expression of N_1 when a question expression of N_1 no N_2 is ambiguous. We obtained good results in the experimental result.

1 Introduction

In recent years, it is possible to find a target document by accessing the enormous resources available on the www through information retrieval techniques. However, A user is difficult to find technical documents targeting, for example, political information and judicial precedents[5]. One of the reasons is that a user tends to make an ambiguous query for information extraction. In the case of political information, the utterances of many councilors include political terms such as "*yosan*(budget)", "*zeikin*(tax)" and so on. Using these political terms, a user cannot get retrieve target documents because there are too much target documents which include these political terms such as "*yosan no sakugen*(budgetary cutback)", "*yosan no gendogaku*(budgeted allowance)", "*yosan no saisyutu*(expenditure of the budget)" and so on. Moreover, it is difficult for a user to think about an appropriate phrase query that would fit these existing in documents. To resolve these problems, we consider that an interactive question answering system is a better approach. In the case of interactive question answering, a system asks a user yes-no questions using a phrase that includes language from target documents. When the system creates a yes-no question using phrases obtained from target documents, it is possible to create a concrete representation by expanding the candidate phrase. However, these phrases tend to become too many words because the system fails to recognize an

optimal phrase. In consideration of these problems, we propose a method generatizing a yes-no question, which is both concrete and clear. In our previous work, we proposed an interactive question answering method using a decision tree. This system was created for the political information supporting service between local councilor and inhabitants through the WWW. In this paper, we present a yes-no question generation method and its experimental results.

2 Creating a question to clarify user's interest

In political information processing, some research focuses on the relation between a user's opinion and a political party plan[8]. Uekami et al. made the Vote Match system where each question is manually created[7]. However, in case of local politics in Japan, the number of municipalities is about 1,760, and their cities, towns and villages have their specific problems. Therefore the manual approach is difficult to realize. In addition to that, each municipality has many councilors.

Therefore, we have made political categories for local politics in order to extract characteristics of each local councilor[6]. We investigated suitable expressions for a user to extract the description of minutes corresponding to a political category. As a result, we confirmed that the question expression of " N_1 no N_2 " was more efficient than that using an usual compound noun.

2.1 Question expression including " N_1 no N_2 "

We collected " N_1 no N_2 " expressions from the council minutes of Otaru City, Japan in 2007. In result 14,336 " N_1 no N_2 " expressions were extracted. Table 1 shows an example of " N_1 no N_2 " extraction. However, there are expressions not suitable for the question. Therefore, we consider a method that filters out incorrect expressions.

Table 1. Examples of " N_1 no N_2 "

" N_1 <u>no</u> N_2 "	Freq of Minutes
<i>rijiisa no touben</i> (The statement of the director)	97
<i>simin no minasan</i> (All of citizen)	63
<i>teian no riyuu</i> (The explanation of the suggestion reason)	25
<i>sitsumon no gaiyou</i> (The summary of the question)	24
<i>isi no kakuhou</i> (The securement of the doctor)	13

¹ Department of Information and Management Science, Otaru University of Commerce, Japan, email: kimura@res.otaru-uc.ac.jp

² Graduate School of Environment and Information Science, Yokohama National University

³ Department of City Life Studies, Utsunomiya Kyowa University

⁴ National Institute of Informatics

Table 2. Examples of filtering using Japanese Google N-gram

" N_1 no N_2 " in Minutes	Frequency	" N_1 wo N_2 suru" in Japanese Google N-gram	Frequency
<i>teian</i> <u>no</u> <i>riyuu</i> (explanation of the proposed reason)	25	<i>teian</i> <u>wo</u> <i>setsumei</i> <u>suru</u> (to explain the proposed reason)	43
<i>isi</i> <u>no</u> <i>kakuho</i> (securement of doctor)	13	<i>isi</i> <u>wo</u> <i>kakuho</i> <u>suru</u> (to secure doctor)	31
<i>jinkenhi</i> <u>no</u> <i>yokusei</i> (constrain of personal expense)	13	<i>jinkenhi</i> <u>wo</u> <i>yokusei</i> <u>suru</u> (to constrain personal expense)	70
<i>keikaku</i> <u>no</u> <i>sakutei</i> (establishment of plan)	12	<i>keikaku</i> <u>wo</u> <i>sakutei</i> <u>suru</u> (to establish plan)	51

2.2 Related work on " N_1 no N_2 "

There are many studies of the Japanese noun phrase " N_1 no N_2 ". Kurohashi et al. have conducted semantic analysis of Japanese noun phrases[4]. Japanese noun phrases in the form of N_1 no N_2 have a broad usage. For example :

- *watasi* no *kuruma* (my car) possession
- *yakyuu* no *senshu* (player of baseball) category
- *toranpu* no *tejina* (trick of card) instrument

In these usages of "no", we focus on relationship between N_1 and N_2 being AN "object" and A "verbal noun".

Kataoka et al. proposed that if a verb is an abbreviated predicate, a verbal noun modification N_1 ga/wo/...⁵ V-suru⁶ N_2 can be paraphrased into N_1 no N_2 [1]. In this paper, we filter N_1 no N_2 in reference to Kataoka's method.

2.3 Filtering using Japanese Google N-gram

As we mentioned in Section 2, using " N_1 no N_2 " it is often difficult to extract a question expression from the city council minutes. In the section 2.1, we mentioned that by using " N_1 no N_2 " it is often difficult to extract a question expression from the city council minutes. Therefore, we filter expressions that are able to be paraphrased from " N_1 no N_2 " to " N_1 wo N_2 suru". For checking of question expressions, we use Japanese Google N-gram[2]. Japanese Google N-gram includes 7-gram phrases which exist more than 20 times in the WWW. We extracted 805 expressions from 14,336 expressions included in the 2007 minutes of Otaru City council. Table 2 shows the result of filtering by Google N-gram.

3 Investigation of optimal expression

In this section, we find the optimal expressions of expanded candidates in the 2007 minutes of Otaru City council. Some of the 805 candidates that are checked by Google N-gram have ambiguous meaning such as "keikaku no sakutei" (development of plan). Therefore, we increased the number of words without being restricted by N_1 to find the optimal expression, which is both concrete and clear.

⁵ Japanese case particle

⁶ Verb

3.1 Making experiment data

In our research, we set 96 political categories in order to classify a city council minutes, and we annotated the 2007 minutes of Otaru City council. In this section, we use 1,667 utterances of the minutes, and their utterances are annotated in only one (category). In order to expand the " N_1 no N_2 ", we use the CaboCha as a syntactic analysis tool[3]. We make expanded candidate expressions, which lengthen the expression step by step toward the front of the utterance. Table 3 shows an example of expanded candidate expressions. In table 3, the expression of "*hiyo no sisan*" (A test calculation of expenses) is possible to paraphrase as follows:

"*hiyo wo sisan suru*" (I calculate expense as a test). !!

We expand the candidate expressions from right to left until a punctuation mark or the beginning of the sentence is reached. We made 20 questionnaires such as in table 3. 20 questionnaires include 90 expression candidates. In other words, a subject evaluates the optimal expression among about 4 or 5 expression candidates.

3.2 Evaluation point of view

We evaluate candidates by using three points of view which consists of "ambiguous $\iff$ clearness", "redundant $\iff$ concise" and "unreadable $\iff$ readable". The subject performs five phases of evaluation from three points of view. Here, we explain how to decide the optimal expressions of expanded candidates in the 2007 minutes of Otaru City council. First, we calculate an arithmetic mean of the evaluations by 6 subjects. As shown as the equations (1), an arithmetic mean is calculated using three points of view which consists of "ambiguous $\iff$ clearness", "redundant $\iff$ concise" and "unreadable $\iff$ readable". Here, N denotes the number of subjects, and c is a candidate expression.

$$AM_1(c) = \frac{\sum_{i=1}^N \text{ClearnessScore}(i)}{N}$$

$$AM_2(c) = \frac{\sum_{i=1}^N \text{ConciseScore}(i)}{N}$$

$$AM_3(c) = \frac{\sum_{i=1}^N \text{ReadableScore}(i)}{N} \quad (1)$$

$$HM(c) = \frac{3}{\sum_{i=1}^3 \frac{1}{AM_i(c)}} \quad (2)$$

Second, we calculate a harmonic mean from the results of three points of view, as shown the equation (2). Finally, we determine the optimal expression by selecting the highest harmonic mean.

Table 3. An example of the questionnaire

The example of utterance of councilor						
... -omitted- ...						
<i>, taisyou kuiki wo hirogeta baai no puru ni kakaru hiyou no sisan</i> (A test calculation of the expense to affect a pool when I enlarged the object area.)						
... -omitted- ...						
The candidates of the question expression						
<----- Expanding from right to left.						
Q1	<i>hiyou no sisan</i>					
Q2	<i>kakaru hiyou no sisan</i>					
Q3	<i>puuru ni kakaru hiyou no sisan</i>					
Q4	<i>baai no puuru ni kakaru hiyou no sisan</i>					
Q5	<i>hirogeta baai no puuru ni kakaru hiyou no sisan</i>					
Q6	<i>taisyou kuiki wo hirogeta baai no puuru ni kakaru hiyou no sisan</i>					
(A test calculation of the expense to affect a pool when I enlarged the object area.)						
Evaluation item						
In this example, a subject evaluates the following questionnaire from Q1 to Q6.						
No.1	Is this question comprehensible or incomprehensible? ☐ It is a comprehensible question. ☐ It is an incomprehensible question. #If you select comprehensible question, answer the following question.					
<----->						
No.2	ambiguous	1	2	3	4	5
No.3	redundant	1	2	3	4	5
No.4	unreadable	1	2	3	4	5
						clearness concise readable

3.3 Results of expanded expression of " N_1 no N_2 "

We performed an expression evaluation with six subjects, two male and four female university students. Table 4, shows the expressions of the highest harmonic mean. In this example, we calculated an ungrammatical expression as "0".

Table 5. Experimental result of selection

Method	Accuracy		Ratio of HM		
Baseline	7/20	0.35	61.84 / 74.63	0.8286	
Method1	15/20	0.75	66.62 / 74.63	0.8927	
Method2	16/20	<u>0.80</u>	70.24 / 74.63	<u>0.9412</u>	

3.4 Discussion

In table 4, the expressions marked with "#" were not expanded, and these expressions confirmed 7 patterns. In this evaluation, there are 90 candidates, which include the 20 unexpanded expressions. Of the 90 candidates, more than 2 people evaluated 71 candidates as an ungrammatical expression. The characteristics of the 71 candidates are as follows: 1. The expanded block including the specific particle is not suitable for expression. For example, the specific particles include "ga" and "ha" in Japanese. 2. The expanded block including adverbs is not suitable for expression.

4 Proposed methods of generating question expressions

We propose two methods based on the section 3.4. Method 1 is limited by the specific particles. Method 1 selects the longest expression that does not include following Japanese particles: "de", "ga", "mo", "ha", "kara", "niokeru" and "niyori". Method 2 is limited by adverbs. Method 2 selects the longest expression that does not include any adverb.

5 Experiment

We perform an experiment that evaluates to select the optimal expression of candidates by our proposed method. We perform the evaluation from two points of view. The point uses the harmonic mean shown in table 4. We calculate the ratio between the sum of the harmonic mean of the selected expressions and the sum of the 20 highest results of harmonic mean. The second point calculates the accuracy of the correct expressions that define the highest harmonic mean. We compare our methods with a baseline method that does not expand " N_1 no N_2 ".

Table 5 shows the experiment which confirm the effectiveness of Method 2.

6 Conclusion

We proposed two methods for selecting the most appropriate expression from expanded candidates for question generation, and confirmed their effectiveness. As future work, we are going to implement the system using our proposed method.

REFERENCES

- [1] Akira Kataoka, Shigeru Masuyama, and Kazuhide Yamamoto, 'Summarization by Shortening a Japanese Noun Modifier into Expression A no B', 409-414, (1999).
- [2] Taku Kudo and Hideto Kazawa, 'Web Japanese N-gram Version 1', (2007).
- [3] Taku Kudo and Yuji Matsumoto, 'Japanese Dependency Analysis using Cascaded Chunking', 63-69, (2002).
- [4] Sadao Kurohashi and Yasuyuki Sakai, 'Semantic Analysis of Japanese Noun Phrases : A New Approach to Dictionary-Based Understanding', 481-488, (1999).
- [5] Rafal Rzepka, Hideyuki Shibuki, Yasutomo Kimura, Keiichi Takamaru, Masafumi Matsuhara, and Koji Murakami, 'Toward Automatic Support For Japanese Lay Judge System - Processing Precedent Factors For Sentencing Trends Discovery', 33-41, (2008).
- [6] Keiichi Takamaru, Hideyuki Shibuki, Yasutomo Kimura, Dai Hasegawa, Hokuto Ootake, and Kenji Araki, 'Extraction of Political Activity of Assemblyman from Minutes of Municipal Assemblies Using the Political Category', B11, (2009).
- [7] Takayoshi Uekami and Tetsuya Sato, 'Estimating the policy positions of political actors: an application of computerized coding to the Japanese policy documents(in Japanese)', 61-73, (2009).
- [8] VOTEMATCH, 'http://www.votematch.co.uk/europe/'.

Table 4. The result of question expression of the highest harmonic mean.

	Question expression of the highest harmonic mean	Harmonic mean
1	<i>puuru ni kakaru hiyou no sisan</i> (A test calculation of the expense to be spent for a pool.)	3.60
2	<i>sijyo wo meguru kankyou no henka</i> (An environmental change concerning the market.)	3.94
3	<i>sa-bisu no koujyo#</i> (Improvement of a service)#	3.41
4	<i>sidou no tekiseina kinou no hoji</i> (Maintenance of a reasonable function of a municipal road.)	3.80
5	<i>bunka geijyutsu ni fureru kikai no kakujyuu to jinzai no ikusei</i> (Expansion of the opportunity to mention culture art and upbringing of the talented person.)	3.57
6	<i>kosodate oyako no koryuu nado wo sokusin suru jigyou no kakujyuu</i> (Expansion of a business to promote an interchange of the child cares.)	4.00
7	<i>75 sai ijyou no koureisya heno kakokuna hokenryou no futan</i> (A burden of a severe burden to a senior citizen older than 75 years old.)	4.17
8	<i>simin karano youbou noaru basu no sinsetsu</i> (A new establishment of a bus route with the demand from a citizen.)	4.30
9	<i>kigyuu no ikusei#</i> (Upbringing of a company)#	3.62
10	<i>gakkou no annzenn ni kakawaru kankyou no henka</i> An environmental change about security of a school.	4.17
11	<i>miyagehin no kounyuu#</i> (Purchase of a souvenir)#	3.38
12	<i>jinzai no touyou#</i> (Promotion of a talented person)#	4.41
13	<i>byougentai ga tainai ni sinnyuu site kansenn site zousyokusi hassei suru kannjya no sousyou</i> The general term of the disease a pathogen invades the body and is contagious and multiplies, and to develop.	2.24
14	<i>tettai no sidou#</i> (Guidance of a removal)#	3.08
15	<i>kaigoyobou puran no sakusei#</i> (Making of a care prevention plan)#	4.01
16	<i>bu no saihen#</i> (Reorganization of a department)#	3.57
17	<i>hokenshou ga tsukaenai youna sikakusyomeisyo no hakkou</i> (Issue of qualification in case when a health insurance card is not usable.)	3.60
18	<i>akaji no kaisyuu to zaisei no kennzennka</i> (Deficit cancellation and fiscal fitness.)	4.30
19	<i>kigyousai ganri syoukankin no gensyuu ya iji kanri no kouritsuka</i> (Efficiency of redemption cost on revenue bond.)	3.74
20	<i>tiiki no kyoutikyoku wo ikasita yutakana kyoutikyukatsudou no suisin</i> (The promotion of the rich instructional activity that makes use of local education power in.)	3.72
Sum of harmonic mean. (This result is used to evaluate the experiment.)		
The expression attached of ”#” is not expanded.		74.63