A Method for Automatically Generating Questions about a User's Political Interest Using Minutes of Municipal Councils

Yasutomo Kimura [1], Hideyuki Shibuki [2], Keiichi Takamaru [3],
Tetsuro Kobayashi [4],Tatsunori Mori [2]
[1] Department of Information and Management Science
Otaru Unversity of Commerce
Japan
kimura@res.otaru-uc.ac.jp

[2] Graduate School of Environment and Information Science
Yokohama National University. Japan

[3] Department of City Life Studies, tsunomiya Kyowa University. Japan

[4] National Institute of Informatics, Japan

ABSTRACT: *This paper proposes a political question generation method that includes Japanese noun phrases of the form N1 no N2. We focus on generating a yes-no question whether a user is interested in the utterance of a councilor. An example of a political question is as follows: nisankatanso(N1) no haisyutu(N2) nituite kuwasiku sir- itai desu ka? (Do you want to know more about emission levels of carbon dioxide?) In this paper, N1 and N2, mean a noun or a com- pound noun. The form N1 no N2 which includes the Japanese post- positional no has a much broader usage. These expressions of the form N1 it no N2 cannot often use a question expression for asking the interesting such as simin no minasan (All of citizens). Therefore, our method checks if an expression is whether a user can understand a question expression. Moreover, our method expands the expression N1 when a question expression of the form N1 no N2 is ambiguous. The experiment results yielded good.*

Key words: Japanese language processing, Phrase analysis, N Gram techniques

Received: 2 June 2010, Revised 5 July 2010, Accepted 10 July 2010

1. Introduction

In recent years, it is possible to find a target document by accessing the enormous resources available on the Internet through information retrieval techniques. However, a user may find it difficult to find technical documents on, 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) and *zeikin* (tax). Using these politi- cal terms, a user cannot retrieve target documents because there are too many target documents that include these political terms such as *yosan no sakugen* (budgetary cutback) , *yosan no gendogaku* (budgeted allowance) and *yosan no saisyutu* (expenditure of the budget). Moreover, it is difficult for a user to think about an appropriate phrase query that would fit those existing in documents. To resolve these problems, we consider an interactive question answering system as a better approach. In the case of interactive question answering, a system asks a user yes-no questions using a phrase 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 verbose because the system fails to recognize an optimal phrase. In consideration of these problems, we propose a method of generating a yes-no question that is both con- crete and clear. 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. created the Vote Match system, where each question is manually created[7]. However, in the case of local politics in Japan, even though the number of municipalities is about 1,760, their cities, towns, and villages have their distinct problems. Therefore, the manual approach is difficult to adopt. In addition to that, each municipality has many councilors.

Therefore, we have created political categories for local politics in order to extract characteristics of each local councilor[6]. We investigated suitable expressions for a user to extract descriptions of minutes corresponding to a political category. As a result, we confirmed that questions having the structure of N_1 *no* N_2 were more efficient than those using a common compound noun.

2.1 Question of the form *N*1 *no N*2

We collected N_1 *no* N_2 expressions from the council minutes of Otaru City, Japan, in 2007. As a 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 that are not suitable as questions. Therefore, we consider a method that filters out incorrect expressions.

2.2 Related work on N_1 *no* N_2

There are many studies on the Japanese noun phrase N_1 *no* N_2 . Kurohashi et al. have conducted semantic analysis of Japanese noun phrases[4]. Japanese noun phrases of the form 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* (card trick) instrument

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

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

In these usages of *no*, we focus on the relationship between N_1 and N_2 being a N 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-suru[6] 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.

N_1 *no* N_2	Frequency
rijisya no touben (The statement *of* the director)	97
simin no minasan (All *of* citizen)	63
teian no riyuu (The explanation *of* the suggestion reason)	25
sitsumon no gaiyou (The summary *of* the question)	24
isi no kakuho (The securement *of* the doctor)	13

Table 1. Examples of N_1 *no* N_1

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 that get more than 20 hits on the Internet.We extracted 805 expressions from 14,336 expressions included in the 2007 minutes of the 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 the Otaru City council. Some of the 805 candidates that are checked by Google N-gram have an ambiguous meaning, for instance, ”keikaku no sakutei” (development of plan). Therefore, we increased the number of words without being restricted by N_1 to find the optimal expression that is both concrete and clear.

3.1 Making experiment data

In our research, we set 96 political categories in order to classify city council minutes, and we annotated the 2007 minutes of the Otaru City council. In this Section, we use 1,667 annotated utterances of the minutes, and their utterances were annotated by eight students. In order to expand the form N_1 *no* N_2, we use the CaboCha as a syntactic analysis tool[3]. We form expanded candidate expressions that lengthen the expression in a stepwise manner toward the front of the utterance. Table 3 shows an example of expanded candidate expressions. In table 3, the expression *hiyo no sisan* (a test calculation of expenses) can be paraphrased as follows:

hiyo wo sisan suru (I calculate expenses 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 created 20 questionnaires by following the method in Table 3. These 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 : ”ambiguous $\Leftrightarrow$ clear,” ”redundant $\Leftrightarrow$ concise,” and ”unreadable $\Leftrightarrow$ readable.” The subject performs five phases of evaluation from the three points of view. Here, we explain how to decide the optimal expressions of expanded candidates in the 2007 minutes of the Otaru City council. First, we calculate the arithmetic mean of the evaluations by six subjects. As shown in equation (1), an arithmetic mean is calculated using three points of view, namely ”ambiguous $\Leftrightarrow$ clear,” ”redundant $\Leftrightarrow$ concise,” and ”unreadable $\Leftrightarrow$ readable.” Here, N denotes the number of subjects, and c is a candidate expression.

[1] *Japanese case particle*

[2] *Verb*

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

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

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

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

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

	An example of the utterance of a councilor						
	· · · -omitted- · · ·						
	, *taisyou kuiki wo hirogeta baai no puru ni kakaru hiyou no sisan* (A test calculation of the expense to be spent for a pool in the case of expanding target area for redevelopment)						
	· · · -omitted- · · ·						
	The candidates of the question expression						
	<———— Expanding from right to left.						
Q1	*hiyou no sisan*						
Q2	*kakaru hiyou no sisan*						
Q3	*puuru ni kakaru hiyou no sisan*						
Q4	*baai no puuru ni kakaru hiyou no sisan*						
Q5	*hirogeta baai no puuru ni kakaru hiyou no sisan*						
Q6	*taisyou kuiki wo hirogeta baai no puuru ni kakaru hiyou no sisan* (A test calculation of the expense to be spent for a pool in the case of expanding target area for redevelopment)						
	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 selected comprehensible, answer the following question.						
		< ——— >					
No.2	ambiguous	1	2	3	4	5	clear
No.3	redundant	1	2	3	4	5	concise
No.4	unreadable	1	2	3	4	5	readable

Table 3. An example of the questionnaire

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

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

Method	Accuracy		Ratio of HM	
Baseline	7/20	0.35	61.84 / 74.63	0.8286
Method 1	15/20	0.75	66.62 / 74.63	0.8927
Method 2	16/20	0.807	0.24 / 74.63	0.9412

Table 5. Experimental result of selection

3.3 Results of the expanded expression of the form 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 with the highest harmonic mean. In this example, we scored an ungrammatical expression as ”0.”

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 section 3.4. Method 1 is limited by the specific particles. Method 1 selects the longest expression that does not include the 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 the selection of the optimal expression of candidates by our proposed method. We perform the evaluation from two points of view. The point using the harmonic mean is shown in Table 4. We calculate the ratio between the sum of the harmonic mean of the selected expressions and the sum of the harmonic means of 20 expression having the 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 result of the experiment that confirms 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. In the future, we are going to implement the system based on our proposed method.

References

[1] Kataoka, Akira., Masuyama, Shigeru., Yamamoto, Kazuhide (1999). Summarization by shortening a japanese noun modifier into expression a no b’, 409–414.

[2] Kudo,Taku., Kazawa, Hideto (2007). Web japanese n-gram version 1.

[3] Kudo,Taku., Matsumoto, Yuji (2002). Japanese dependency analysis using cascaded chunking’, 63–69.

[4] Kurohashi,Sadao., Sakai, Yasuyuki (1999). Semantic analysis of japanese noun phrases : A new approach to dictionary-based understanding, 481–488.

[5] Rzepka, Rafal., Shibuki, Hideyuki., Kimura, Yasutomo.,Takamaru, Keiichi., Matsuhara, Masafumi., Murakami, Koji (2008). Toward automatic support for Japanese lay judge system - processing precedent factors for sentencing trends discovery, 33–41.

[6] Takamaru, Keiichi., Shibuki, Hideyuki., Kimura, Yasutomo., Hasegawa, Dai., Ototake, Hokuto, Araki, Kenji (2009).Extraction of political activity of assemblyman from minutes of municipal assemblies using the political category’, B11.

[7] Takayoshi, Uekami., Sato, Tetsuya (2009).Estimating the policy positions of political actors: an application of computerized coding to the Japanese policy documents(in japanese), 61–73.

[8] VOTEMATCH, ‘http://www.votematch.co.uk/europe/’.