On Implementing an Automatic Headline Generation for Discussion BBS Systems —Cases of Citizens' Deliberations for Communities—

Katsuhide FUJITA^{†a)}, *Member* and Ryosuke WATANABE[†], *Nonmember*

SUMMARY Recently, the opportunity to discuss topics on a variety of online discussion bulletin boards has been increasing. However, it can be difficult to understand the contents of each discussion as the number of posts increases. Therefore, it is important to generate headlines that can automatically summarize each post in order to understand the contents of each discussion at a glance. In this paper, we propose a method to extract and generate post headlines for online discussion bulletin boards, automatically. We propose templates with multiple patterns to extract important sentences from the posts. In addition, we propose a method to generate headlines by matching the templates with the patterns. Then, we evaluate the effectiveness of our proposed method using questionnaires.

key words: *automated headline generation, electronic discussion bulletin boards, natural language processing*

1. Introduction

In recent years, the general public has taken to exchanging opinions using various online platforms, such as online discussion bulletin board systems (BBSs), microblogs, and other promising collective intelligence platforms ([1] etc.). On web-based platforms, users can freely post their opinions without constraints of time or place. In the near future, large-scale discussion aggregations, discussions, and negotiations will become possible on supporting web-based platforms, and further developments in social crowd engineering are expected ([2]–[5]). With an increase in the amount of information and posts on online discussion platforms, it would be difficult to read and understand the contents of all posts. Therefore, platforms that automatically construct a structured visualization of discussions have been proposed [1]. In particular, an open web-based forum system called COLLAGREE ([6], [7]), which has facilitator support functions, has been developed and deployed for an internet-based town meeting in Japan. In COLLAGREE, a discussion tree is employed and a method that generates headlines for each post is required [8].

In addition to social and academic environments, linguistic resources have become increasingly rich via web developments and the technology of natural language processing (NLP) has been developed. However, it is still difficult to obtain accurate information from large-scale discussion platforms. This is because the amount of unnecessary infor-

mation has also increased with the development of linguistic resources. In addition, argumentation mining which aims at automatically extracting structured arguments from unstructured textual documents has become important for internet-based discussions ([9]–[11]). This topic has recently become of broad interest due to its potential to process information originated from the Web, in particular, from social media. Automatic structuring of discussions and argumentations is enabling breakthrough applications in social and economic sciences, policy-making, and crowd engineering. Argumentation models, methods, and applications have been proposed to combine representational needs with user-related cognitive models and computational models for automated reasoning.

Automatic headline generation is necessary for visualizing user-related cognitive models and computational models. Although a large number of studies have focused on automatic headline generation ([12] etc.), previous approaches have not taken into account the heuristics of the discussion BBS such as the structure of the discussions because their targets have been general documents (newspapers or books etc.), not large-scale online discussion platforms.

In this paper, we propose a method to automatically extract important sentences and generate headlines from posts for discussion BBSs. In our proposed method, templates with multiple patterns are used to extract important sentences from the posts, and headlines are generated by matching the templates with the patterns. For improving the quality of generating the headline, we also propose the hybrid method of automatic headline generation. In this hybrid method, we combined the proposed method using templates with a heuristic method of selecting the first sentence as the headline. The method of selecting the first sentence as the headline is effective for extracting headlines in newspapers, where important contents tend to be placed at the beginning of the article.

Then, we evaluate the effectiveness of our proposed method using questionnaires. The datasets for the evaluations are log data from an internet-based town meeting in Nagoya, which was a city project for an actual town meeting of the Nagoya Next Generation Total City Planning for 2014–2018 [6]. In these evaluations, we compare our proposed approach and hybrid approach with various existing methods of automatically generating headlines using a questionnaire.

Considering the above backgrounds and overviews of this paper, the main contributions of this paper are summa-

Manuscript received March 14, 2017.

Manuscript revised August 7, 2017.

Manuscript publicized January 19, 2018.

[†]The authors are with Institute of Engineering, Tokyo University of Agriculture and Technology, Koganei-shi, 184–8588 Japan.

a) E-mail: katfujii@cc.tuat.ac.jp

DOI: 10.1587/transinf.2016IIP0017

rized as follows:

- This paper proposes a heuristic approach to generate headlines of a set of articles on a BBS system. The proposed approach combines heuristic pattern-based extraction of candidate texts and then applies a multi-sentence compression method.
- This paper investigates the possibility of applying the proposed pattern-based heuristic extractions to the cases of citizens' deliberations for communities.
- This paper evaluates our proposed approach on a real dataset about the cases of citizens' deliberations for communities. In the presented case, the proposed approaches have worked well compared to other baseline approaches. Although the presented approach is currently using some language-specific patterns, once we could improve the language-specific parts in our proposed templates for non-Japanese languages, it could be applied to other languages.

These contributions give some directions of implementation of building a practical Knowledge Management (KM) application for complex and unstructured discussion datasets such as the citizens' deliberations for communities. The approach proposed in this paper is based on the heuristics. Therefore, it can allow application developers or operators of the BBS system to manage and control how these summarized headlines for the discussions could be formed and generated. In addition, the analysis of the proposed approach in the real dataset of the cases of citizens' deliberations for communities gives valuable knowledge in a certain new area.

The remainder of this paper is organized as follows. In Sect. 2, related studies concerning automatic headline generation methods are shown. In Sects. 3 and 4, the approach for automatic headline extraction using templates and the hybrid approach combining the method using templates with heuristic method are proposed. In Sect. 5, we describe the results of the evaluation experiments and discussions. Finally, we present our conclusions.

2. Related Works

Multiple studies concerning headline generation have focused on many languages and document types ([13]–[16] etc.). According to Gattani, three primary techniques can be identified [15].

2.1 Statistics-Based Approaches

These methods apply statistical models to learn correlations between words in headlines and documents. The majority of models are used in supervised learning environments; therefore, large amounts of training datasets with labels are necessary. For example, a method viewing summarization as a problem analogous to statistical machine translation based on Naive Bayes has been proposed [17]. The use of statistical models for learning pruning-rules for parse trees has also

been studied ([18] etc.).

2.2 Summarization-Based Approaches

Headlines can be thought of as very short summaries. Therefore, traditional summarization methods can be used to generate one-line summaries ([19] etc.). The primary difficulty with these approaches is that they use techniques that were not initially devised for generating compressions of less than 10% of the original content, which directly affects the quality of the resulting summary [17]. The generated headline has risks of losing or changing the contextual meaning of the reused words because these approaches generate a headline by reusing words in the article, [20].

2.3 Template-Based Approaches

These methods apply handcrafted linguistic rules to extract or compress important sentences in a document [21]. They are simple and lightweight; however, they fail to understand complex relationships in the text, and it is not easy to prepare templates for them. For example, Alfonseca et al. proposed an approach of extracting the syntactic patterns that a Noisy-OR model generalizes into event descriptions [22]. At the inference time, the method searches the model with the observed patterns in an unseen news collection and identifies the event that best captures the gist of the collection to retrieve the most appropriate pattern to generate a headline. However, original templates for each language are necessary in this approach.

Automatic headline generation focusing on the Japanese language and online BBSs are few. Therefore, we employ this approach because it is lightweight and does not require training datasets to generate templates.

In the work of Gupta et. al. [23], after annotating the parts to be extracted from the source texts, it will semi-automatically generate the sufficient patterns to extract the title. If this kind of information extraction framework could be applied to this evaluation dataset, or some machine learning based approach would be applied, the performance of the proposed approach could produce better performance than the presented patterns in this paper. However, these tools are mainly intended to make direct information extraction (i.e., extracting terms and relations among them directly), rather than matched and higher-ranked sentences from texts.

2.4 Automatic Headline Generation Approaches for Multiple Sentence Compression

The above three approaches used in the previous studies focused on a single article or document. However, this paper focuses on texts in online discussion BBSs; in other words, this paper can consider the task of summarizing a cluster of related sentences with a short sentence, which we call multi-sentence compression, to present a simple headline. The advantage and novelty of our proposed method is that we can

use the characteristics of the discussions and Japanese.

One popular approach to multi-sentence compression was described by Filippova, who presented a simple and robust word graph-based method to generate succinct compressions that require just a part of speech tagger and a list of stop words [12]. The advantage and the novelty of the proposed method is that it is syntax lean and requires little more than a tokenizer and a tagger. However, the approach of Filippova is evaluated under English and Spanish datasets. After that they have been proposed LSTM-based approaches for the similar purpose of their previous paper [24]. The proposed LSTM-based model outperforms the baseline in readability and informativeness. The approach of Filippova is based on learning approach, which approach is totally different from our proposed approach.

3. Automatic Headline Generation Using Templates

BBSs contain several threads by posting their comments to the thread. A thread refers to a set of posts related to a particular topic or issue. In COLLAGREE interface, the first post of the user who creates the thread becomes the parent post, and other users reply to it as child posts. There are also cases where some users reply to a child post as grandchild posts.

Our goal is to propose a method to automatically generate a headline for any post in the thread in the discussion BBS. Figure 1 shows the flowchart of the proposed automatic headline generation using templates. The process of the proposed method consists of five parts: Pre-processing, Extraction of Headline, Generation of Headline, Complement of Headline, and Post-processing. The sentence which is extracted by the method proposed is used as the headline

Fig. 1 Flowchart of automatic headline generation using templates

when the length of the sentence is less than twenty letters. On the other hand, When the length of the sentence is more than twenty letters, the headline is generated automatically based on the following template.

The patterns in the proposed templates were prepared by analyzing the online discussions datasets. In addition, we used the heuristics of NLP and online discussions based on our experiences. Therefore, our proposed method is generic enough to be applied to various conditions because our proposed method has a enough generalization to the discussions datasets without happening the specializations of the proposed models. In addition, our proposed patterns and templates don't use the technique for specializing of improving under the specific datasets at all.

3.1 Pre-Processing

To arrange the text of the BBS for the discussion, the following preprocessing step is conducted. The parentheses and URL are removed. These unimportant information can cause the generated summary to be redundant and should therefore be deleted before summarization. It is also essential to clarify anaphors in the original post. In reference to the result of an anaphora analysis using CaboCha [25], only anaphors that refer to a named entity are retained. Furthermore, only nouns and verbs are extracted from each post (e.g., an opinion or a sentence) based on the result of a morphological analysis.

3.2 Extraction of Headline

A sentence to be used as a headline is extracted from the posts consisted of several sentences. In previous studies focusing on conference notes, headlines were extracted by focusing on expressions such as “about” [26]. However, a clear structure does not exist for online discussion bulletin boards. Therefore, we propose a novel template with several patterns of sentences that seem to express the intent of each BBS participant who wrote the post. Table 1 shows a proposed template with three patterns to extract important sentences from the discussion BBS. Table 1 shows the original form of verb only; however, other inflected word forms are applied. We can match original form of verb with a morphological analysis using CaboCha on the proposed templates. The details of each pattern are as follows.

The details of each pattern are as follows:

- **Pattern 1: Sentence including “need” or “important”**

Sentences including “need” or “important” are the most essential contents for the participants. Therefore,

Table 1 Template for extracting the sentences

	Priority	Pattern
pattern 1	High	“need” or “important”
pattern 2	Medium	“think,” “believe” or “regard”
pattern 3	Low	agreements (including positive keywords)
		disagreement (including negative keywords)

the priority of this pattern is the highest in the template.

- **Pattern 2: Sentence including “think” or “regard”**
Because the discussion BBS is a place for participants to post their opinions, sentences including “think,” “believe,” or “regard” are often seen. Therefore, their priority is set to medium in the template.
- **Pattern 3: Agreements and Disagreements**
Sentences including positive or negative words are important as headlines because participants express their opinions such as agreements or disagreements using these words when having a discussion. Therefore, a sentence with the highest number of positive and negative words (in absolute sum) is extracted as Pattern 3. The negative and positive words are judged using the Evaluative Expressions Dictionary ([27], [28]). The score for words including positive words is “+1” and that for words including negative words is “-1.”

The sentence with the highest priority is extracted when multiple patterns are matched in Table 1.

When the same patterns are matched with multiple sentences, the sentence with a higher cosine similarity between it and the parent post is extracted. We assume that the set of N nouns are appeared in the parent post and the matching sentence. In the N -dimensions, the set of nouns in the parent post is $P = \{p_1, p_2, \dots, p_N\}$, the set of nouns in the matching sentence is $Q = \{q_1, q_2, \dots, q_N\}$, and each element of P and Q means the frequency of occurrence of the corresponding to each noun. The cosine similarity is calculated by the following expression.

$$\cos(P, Q) = \frac{P \cdot Q}{|P||Q|} \quad (1)$$

The sentence with the highest score is then extracted.

The procedure of extracting the sentence when the same patterns are matched with multiple sentences is described as follows:

1. The sentences vectors of nouns appeared in the parent post and the matching sentence are calculated.
2. The cosine similarity between the parent post and the matching sentence are calculated.
3. The sentence with the highest cosine similarity between it and the parent post are decided.

In other words, sentences with similar words to those in the parent post are extracted.

When the sentences do not match the patterns in Table 1, our method outputs “no pattern.”

3.3 Automatic Headline Generation

The sentence which is extracted by the method proposed in the previous section is used as the headline when the length of the sentence is less than twenty letters. The headline is generated automatically based on the following template when the length of the sentence is more than twenty letters. The template is proposed based on the overall analysis of

actual posts in discussions BBSs.

The proposed template is as follows[†]:

Pattern 1: Sentences including “need” or “important”

“*Noun* + ga(が), (((*Noun* or *Verb*) + (need(必要) or important(重要))) or (*Verb*, *Noun* + ga(が)))”
 “*Noun* + de(で), *Noun* + ga(が)”
 “*Noun* + to-shi-te(として), *Noun* + ga(が)”
 “*Noun* + no(の), *Noun* + ga(が) or (ni(に), *Noun* + ga(が)) or no(の))”
 “*Verb* + yo-u-na(ような), *Noun* + mo(も)”
 “*Noun* + wo(を), (*Noun* or *Verb*) + (ga(が) or be-ki(べき) or (de-a-re-ba(であれは), *Noun* + mo(も)))”

Pattern 2: Sentences including “think” or “regard”

“*Noun* + ga(が), *Noun* + (ni(に) or mo(も) or to(と))”
 “*Noun* + to-i-u(という), *Noun* + ha(は)”
 “*Noun* + na(な), (*Noun* or *Verb*) + da-to(だと)”
 “*Noun* + na-no-ha(なの), *Noun*”
 “*Noun* + ni(に), (*Noun* or *Verb*) + (ga(が) or to(と) or to-ha(とは))”
 “*Noun* + no(の), *Noun* + (ga(が) or ni(に) or (ni(に), *Noun* + ga(が)) or wo(を))”
 “*Noun* + ha(は), (*Noun* or *Verb*) + (ga(が) or no(の) or to(と))”
 “*Noun* + wo(を), (*Noun* or *Verb*)”

Pattern 3: Sentence with positive or negative words

Clauses including the two best BM25 [29] in the positive or negative words.

When multiple patterns match the sentence, the pattern with the highest average weights of BM25 [29] of the noun is used. The expression of BM25 of a word (w) in n posts of set ($D = \{d_1, d_2, \dots, d_n\}$) is defined as follows:

$$bm25_score(w, D) = \sum_{i=1}^n idf(w) \cdot \frac{f(w, d_i) \cdot (k_1 + 1)}{f(w, d_i) + k_1 \cdot (1 - b + b \cdot \frac{|d_i|}{avgdl})} \quad (2)$$

$$idf(w) = \log \frac{N - df(w) + 0.5}{df(w) + 0.5} \quad (3)$$

In the above equations, $f(w, d_i)$ is the frequency of the word(w) in the post(d_i), $|d_i|$ is the number of words in d_i . $avgdl$ is the average number of words in the post set(D), k_1 and b are previously decided parameters. In this study, we set, $k_1 = 2.0, b = 0.75$ are used because they are used in a general way. N is the number of posts and $df(w)$ is the number of posts including the word w in all posts. In our proposed method, we used BM25 because the BM25 of characteristic words in the entire BBS tends to be higher.

[†]The part of the template is written in Japanese because it used the original Japanese grammars.

3.4 Complement of Headline

The headline generated using the automatic headline generation method in the previous section is insufficient because there is no clause modified by the extracted clause to explain it. Therefore, modified clauses are added using the modification parsing of CaboCha [25]. Concretely speaking, a clause to modify the clause extracted by the method in the previous section is complemented when there is an extracted clause. In addition, the clause with the highest BM25 is complemented when there are many extracted clauses. This complement is repeated by a constant number of characters. In this study, the constant number of characters is twenty.

3.5 Post-Processing

This step examines the generated headline to improve it. A particle remains at the end of a sentence when a headline is extracted by matching each clause.

Therefore, unnecessary clauses, such as particles at the end of a sentence, are replaced and removed based on the following list[†]. Removal and matching are repeated for the word lists for the end of the sentence until the headline no longer changes.

Removal Words List

de-sho-u(でしょう), de-su(です), de-shi-ta(でした)

Replace Words List

a-ri-ma-su(あります)→a-ru(ある), should(しなければ
ならない)→need, o-ri-ma-su(おります)→i-ru(いる),
shi-ma-shi-ta(しました)→shi-ta(した)

Removal Words List for the end of sentence

may(かもしれない, しません), think(考えられます,
思います, 考える, 思う), is(なっています), think(思わ
れます), yo-u-ni(ように), na-no(なの), ka-mo(かも),
nn-da(んだ), i-u(いう), ka-to(かと), wo(を), mo(も),
to(と), no(の), ni(に), ha(は), da(だ), ne(ね), yo(よ),
ga(が)

4. Hybrid Method of Automatic Headline Generation

We propose the hybrid method of automatic headline generation. In this method, we combine the proposed method in the previous section with a heuristic method of selecting the first sentence as the headline. This method of selecting the first sentence as the headline is effective for extracting headlines in newspapers, where important contents tend to be placed at the beginning of the article [30]. However, this method is not always effective for online BBSs because the

lengths of posts in BBSs are shorter than those in articles.

The details of the hybrid method of automatic headline generation are as follows:

1. The proposed method using the templates generates the headline automatically
2. The heuristic method of selecting the first sentence as the headline generates the headline automatically
3. The headlines generated in step1 and step 2 are evaluated by summing up the BM25 or tf-idf scores of all clauses including a noun or a verb.
4. The headline with the highest scores is selected from the one generated by the proposed method or the heuristic method

5. Experimental Results

5.1 Datasets for the Experiments

We conducted comparative studies to evaluate the proposed method by extracting headlines using questionnaires. One dataset for the evaluations is log data from an internet-based town meeting in Nagoya, which was a city project for an actual town meeting of the Nagoya Next Generation Total City Planning for 2014-2018 [6]. In addition, another dataset is log data from an internet-based town meeting in Aichi, which was a prefecture project for an actual town meeting of the Aichi Design League [31].

In the Nagoya Next Generation Total City Planning for 2014-2018, there were 266 participants and 9 facilitators. The topics of discussion were “human rights,” “the environment,” “attractive,” and “disasters,” and there were a total of 1151 posts. In the Aichi Design League, there were 75 participants and no facilitators. The topics of discussion were “designing the town” for Aichi Prefecture. There were a total of 355 posts. These experiments did not address the posts from the facilitators because their models are completely different from those of the participants.

5.2 Evaluations for Extracting the Headline

The correct datasets for extracting the sentence suitable for the headline were generated by questionnaires. The most effective sentence as a headline was selected from all posts in each thread by ten undergraduate and graduate students of the Tokyo University of Agriculture and Technology. If the selected optimal headline was different among the evaluators, the sentence selected by the largest number of evaluators was considered as appropriate for the correct answer.

The comparative methods for the experiment are as follows:

- **Proposed Method:** The headline extraction method proposed in this paper
- **First Sentence:** The method of selecting the first sentence as the headline
- **Random:** The headline extraction method of selecting a sentence randomly

[†]The part of the template is written in Japanese because it used the original Japanese grammars.

Table 2 Accuracy rates of each method

Method	Accuracy rates
Proposed Method	0.64
First Sentence	0.37
Random	0.32

Table 3 Accuracy rates in each pattern of our proposed method

Pattern	1	2	3	N/A	Total
Number of applicable sentences	22	48	30	0	100
Accuracy	19	29	16	0	64
Accuracy rates	0.86	0.60	0.53	-	0.64

“First Sentence” is effective for extracting headlines in newspapers, where important contents tend to be placed at the beginning of the article [30]. However, this method is not always effective for online BBSs because the lengths of posts in BBSs are shorter than those in articles. “Random” is a baseline method for the comparative study.

Table 2 shows the accuracy rates of the three methods based on the correct dataset. Our proposed method has the best accuracy rate because the tendency of placing important content at the beginning of a post, as in newspapers, is not applicable in discussion BBSs. Instead, the beginning of a post instead tends to be a simple approval or greeting (e.g., “I agree with his idea” or “Hello”). In other words, there is a tendency that the first sentence of a post does not contain the core idea of comments or questions. In fact, simple replies such as “yes” and “of course” are commonly found. Our proposed method is also better than Random (the baseline); therefore, our proposed method is effective for discussion BBSs.

Table 3 shows the accuracy rates for each pattern of our proposed method. For Pattern 1, the number of applicable cases is small and the accuracy rate is high. Posts including “necessary” and “important” generally reflect the participant’s intentions. Conversely, the number of applicable cases is high and the accuracy rate is low for Pattern 2. This is because “think” and “believe” are commonly used expressions and appear in multiple sentences. For Pattern 3, the sentences including positive words are extracted as the important phrases in the post. All posts match our three proposed patterns in this experiment; therefore, our proposed method has effective templates to extract the important sentences.

5.3 Evaluations for Generating the Headline

We evaluated the headlines generated by our proposed method using questionnaires. The evaluation criteria for the headlines were its “readability” and “completeness” because the purpose of our proposed method was to generate headlines with high readability and core content. Each evaluator graded a headline on a five-grade evaluation scale following Tables 4 - 7.

The comparative methods for these experiments are as follows:

Table 4 Evaluations of readability

Score	Description
5	The readability of the headline is extremely well.
4	The headline is mostly readable.
3	The readability of the headline is tolerable but not great.
2	The readability of the headline isn’t well.
1	The readability of the headline is extremely unnatural.

Table 5 Evaluations of comprehension

Score	Description
5	The meaning of the headline can understand, easily.
4	The meaning of the headline can understand without any difficulty.
3	The meaning of the headline can understand, somehow.
2	It is difficult to understand the meaning of the headline.
1	The meaning of the headline can’t understand.

Table 6 Evaluations of completeness

Score	Description
5	The headline covers all of the post.
4	The headline covers most of the post.
3	The headline covers some of the post.
2	The headline covers parts of the post.
1	The headline covers little of the post.

Table 7 Evaluations of unnecessary

Score	Description
5	The headline doesn’t contain the unnecessary words.
4	The headline contains the less unnecessary words.
3	The headline contains the small unnecessary words.
2	The headline contains the many unnecessary words.
1	Most words in the headline are unnecessary.

- **Template Only:** Our proposed headline generation method using the templates only
- **First Twenty Letters:** Extracting the first twenty letters as the headline from the posts
- **Hybrid (BM25):** Our proposed hybrid headline generation method. The evaluation criteria in selecting the headline is BM25
- **Hybrid (tf-idf):** Our proposed hybrid headline generation method. The evaluation criteria in selecting the headline is tf-idf.

In the proposed method, the clauses were extracted based on the template, and clauses complemented with high BM25 were generated using dependency parsing. In the First Twenty Letters method, the first 20 letters were simply extracted as the headline from the posts. Therefore, the number of letters in the headline was equal to or less than twenty. The headlines of our proposed method and the baseline method were evaluated by 11 undergraduate and graduate students at the Tokyo University of Agriculture and Technology.

Table 8 shows the averages of the five-grade evaluation scores of the four evaluation criteria from the questionnaires. Table 9 shows the p-value obtained from t-test in each pair of the automatic headline extracting methods. Totally, the average scores of each proposed method in this

Table 8 Average of five-grade evaluation scores by the questionnaires

	Readability	Comprehension	Completeness	Unnecessity	Total Average
Template only	3.23	2.67	2.93	3.59	3.10
First Twenty Letters	3.31	2.77	2.93	3.15	3.04
Hybrid (BM25)	3.23	2.81	3.12	3.42	3.14
Hybrid (tf-idf)	3.33	2.84	3.07	3.32	3.14

Table 9 P-value obtained from t-test in each pair of methods[†]

		Readability	Comprehension	Completeness	Unnecessity	Total Average
Template only	First Twenty Letters	0.5037	0.3630	1	2.254E-05	0.4836
Template only	Hybrid (BM25)	1	0.08001	0.01517	0.02967	0.5503
Template only	Hybrid (tf-idf)	0.3666	0.1096	0.1544	0.004702	0.6911
First Twenty Letters	Hybrid (BM25)	0.3159	0.5940	0.005407	0.0002178	0.08463
First Twenty Letters	Hybrid (tf-idf)	0.7877	0.2702	0.004877	0.001158	0.03726
Hybrid (tf-idf)	Hybrid (BM25)	0.2662	0.7703	0.4944	0.2099	0.9110

paper (“Template only,” “Hybrid (BM25)” and “Hybrid (tf-idf)”) are higher than that of the baseline (“First Twenty Letters”) and each of our proposed methods can generate the effective headline automatically in some evaluation criteria, significantly. In addition, the total average scores of two hybrid methods (“Hybrid (BM25)” and “Hybrid (tf-idf)”) are the highest compared with other methods. “Hybrid (tf-idf)” is the highest in the readability and comprehension, and “Hybrid (BM-25)” is the highest in the completeness. If a headline with better readability and comprehension is considered as the better headline, “Hybrid (tf-idf)” is the best method on these experimental results.

Focusing on the differences between “Hybrid(BM25)” and “Hybrid(tf-idf),” “Hybrid(tf-idf)” can generate headline with higher readability and comprehension than “Hybrid (BM25).” However, “Hybrid(BM25)” may be effective in generating the shorter headline (less than twenty letters) because BM25 considers the number of letters in the sentence and the score of “unnecessity” for “Hybrid (BM25)” is the second best and higher than that of “Hybrid (tf-idf).”

6. Conclusion

It is important to automatically generate headlines so that readers can understand the core idea of each post at a glance and easily understand the contents of each discussion. In this paper, we proposed a method of automatically extracting and generating headlines from posts by focusing on the online discussion bulletin board systems. We proposed templates with multiple patterns to extract important sentences from posts. In addition, we proposed a method to generate a headline by matching the templates with the patterns. We evaluated the effectiveness of our proposed method using questionnaires. In the task of extracting the headline from the posts, our proposed method had a higher effect than the baseline method in total.

The possible future work includes the automatic headline generation method based on machine-learning. For

achieving it, larger amounts of discussion data on the electronic discussion bulletin boards are necessary. Another possible future work is to improve our templates for the multilingualization. This paper focused on Japanese texts since the proposed patterns are only applicable to Japanese texts and evaluations were only given on Japanese text dataset. However, overviews of the proposed algorithm itself can be language-neutral. Therefore, patterns and templates themselves proposed in this paper have the possibility of applying to other natural languages (e.g. English), and evaluating them under the other language datasets.

Acknowledgments

This work was supported by CREST, JST.

References

- [1] A. Gürkan, L. Iandoli, M. Klein, and G. Zollo, “Mediating debate through on-line large-scale argumentation: Evidence from the field,” *Inf. Sci.*, vol.180, no.19, pp.3686–3702, 2010.
- [2] M. Klein, “Enabling large-scale deliberation using attention-mediation metrics,” *Computer Supported Cooperative Work*, vol.21, no.4-5, pp.449–473, 2012.
- [3] M. Klein, “Achieving collective intelligence via large-scale on-line argumentation,” *Proceedings of the Second International Conference on Internet and Web Applications and Services (ICIW-2007)*, pp.58–58, 2007.
- [4] M. Klein and L. Iandoli, “Supporting collaborative deliberation using a large-scale argumentation system: The mit collaboratorium,” *MIT Sloan Research Paper*, no.4691-08, 2008.
- [5] J. Introne, R. Laubacher, G.M. Olson, and T.W. Malone, “The climate colab: Large scale model-based collaborative planning,” *2011 International Conference on Collaboration Technologies and Systems, CTS-2011*, pp.40–47, 2011.
- [6] T. Ito, Y. Imi, T. Ito, and E. Hideshima, “Multi-sentence compression: finding shortest paths in word graphs,” *Proceedings of the International Conference on Collective Intelligence 2014 (CI-2014)*, pp.322–330, 2014.
- [7] T. Ito, Y. Imi, M. Sato, T. Ito, and E. Hideshima, “Incentive mechanism for managing large-scale internet-based discussions on collagree,” *Proceedings of the International Conference on Collective Intelligence 2015 (CI-2015)*, 2015.
- [8] A. Sengoku, T. Ito, K. Takahashi, S. Shiramatsu, T. Ito, E. Hideshima, and K. Fujita, “Discussion tree for managing large-scale

[†]The red score means the t-score is less than 5%. In other words, the scores written in the red color mean no significant differences between the methods.

- internet-based discussions,” Proceedings of the International Conference on Collective Intelligence 2016 (CI-2016), 2016.
- [9] M. Lippi and P. Torroni, “Argumentation mining: State of the art and emerging trends,” *ACM Transactions on Internet Technology*, vol.16, no.2, pp.10:1–10:25, 2016.
- [10] B.K. Bal and P.S. Dizier, “Towards building annotated resources for analyzing opinions and argumentation in news editorials,” Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC-2010), pp.1152–1158, 2010.
- [11] P. Baroni, M. Romano, F. Toni, M. Aurisicchio, and G. Bertanza, “Automatic evaluation of design alternatives with quantitative argumentation,” *Argument and Computation*, vol.6, no.1, pp.24–49, 2015.
- [12] K. Filippova, “Multi-sentence compression: finding shortest paths in word graphs,” Proceedings of the 23rd International Conference on Computational Linguistics (COLING-2010), pp.322–330, 2010.
- [13] M. Gambhir and V. Gupta, “Recent automatic text summarization techniques: a survey,” *Artificial Intelligence Review*, vol.47, no.1, pp.1–66, 2017.
- [14] V. Gupta and G.S. Lehal, “A survey of text summarization extractive techniques,” *Journal of Emerging Technologies in Web Intelligence*, vol.2, no.3, 2010.
- [15] A. Gattani, Automated Natural Language Headline Generation Using Discriminative Machine Learning Models, Thesis (School of Computing Science), Simon Fraser University, 2007.
- [16] C.A. Colmenares, M. Litvak, A. Mantrach, and F. Silvestri, “Heads: Headline generation as sequence prediction using an abstract feature-rich space,” *Human Language Technologies: The 2015 Annual Conference of the North American Chapter of the ACL (NAACL-2015)*, pp.133–142, 2015.
- [17] M. Banko, V.O. Mittal, and M.J. Witbrock, “Headline generation based on statistical translation,” Proceedings of the 38th Annual Meeting on Association for Computational Linguistics (ACL-2000), pp.318–325, Association for Computational Linguistics, 2000.
- [18] Y. Unno, T. Ninomiya, Y. Miyao, and J. Tsujii, “Trimming cfg parse trees for sentence compression using machine learning approaches,” Proceedings of the COLING/ACL on Main Conference Poster Sessions (COLING-2006), pp.850–857, Association for Computational Linguistics, 2006.
- [19] H. Morita, R. Sasano, H. Takamura, and M. Okumura, “Subtree extractive summarization via submodular maximization,” Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (ACL-2013) - Volume 1, pp.1023–1032, Association for Computational Linguistics, 2013.
- [20] B.K. Boguraev and M.S. Neff, “Discourse segmentation in aid of document summarization,” Proceedings of the 33rd Hawaii International Conference on System Sciences (HICSS-2000) - Volume 3, 2000.
- [21] B. Dorr, D. Zajic, and R. Schwartz, “Hedge trimmer: A parse-and-trim approach to headline generation,” Proceedings of the HLT-NAACL 03 on Text Summarization Workshop - Volume 5, pp.1–8, Association for Computational Linguistics, 2003.
- [22] E. Alfonseca, D. Pighin, and G. Garrido, “Headly: News headline abstraction through event pattern clustering,” Proceedings of 51th Annual Meeting on Association for Computational Linguistics (ACL-2013), pp.1243–1253, 2013.
- [23] S. Gupta and C.D. Manning, “Improved pattern learning for bootstrapped entity extraction,” *Computational Natural Language Learning (CoNLL)*, pp.98–108, 2014.
- [24] K. Filippova, E. Alfonseca, C.A. Colmenares, L. Kaiser, and O. Vinyals, “Sentence compression by deletion with lstms,” Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing (EMNLP’15), pp.360–368, 2015.
- [25] T. Kudo and Y. Matsumoto, “Japanese dependency analysis using cascaded chunking,” Proceedings of the 6th Conference on Natural Language Learning 2002 (COLING 2002 Post-Conference Workshops), pp.63–69, 2002.
- [26] M. Koizumi, S. Shiramatsu, T. Ozono, and T. Shintani, “Implementation of minutes browsing support system based on structured question and answers by relationship of discourse,” 2012 Information Processing Society of Japan, vol.2012, no.1, pp.657–659, 2012.
- [27] N. Kobayashi, K. Inui, Y. Matsumoto, K. Tateishi, and T. Fukushima, “Collecting evaluative expressions for opinion extraction,” Proceedings of the First International Joint Conference on Natural Language Processing (IJCNLP-2004), vol.3248, pp.596–605, 2005.
- [28] K. Inui, S. Abe, K. Hara, H. Morita, C. Sao, M. Eguchi, A. Sumida, K. Murakami, and S. Matsuyoshi, “Experience mining: Building a large-scale database of personal experiences and opinions from web documents,” Proceedings of the 2008 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT 2008), pp.314–321, 2008.
- [29] S. Robertson and H. Zaragoza, The probabilistic relevance framework: BM25 and beyond, Now Publishers Inc, 2009.
- [30] R. Brandow, K. Mitze, and L.F. Rau, “Automatic condensation of electronic publications by sentence selection,” *Information Processing & Management*, vol.31, no.5, pp.675–685, 1995.
- [31] K. Tanaka, shunpei Fukamachi, T. Ito, T. Ito, and E. Hideshima, “A study on the supporting system of consensus building by internet usage: A case study on aichi city development design league (in japanese),” Summaries of technical papers of annual meeting Architectural Institute of Japan, pp.899–900, 2015.

Appendix A: Examples of Extracted Headlines from Original Posts

[Example 1]

Original texts: 返信が遅れまして申し訳ございません。多様性とは人がそれぞれ持つ個性や特徴を尊重し合って助け合うことをさします。今の名古屋は「出る杭を打つ」というか、異質な人を排除する傾向があると思うのです。日本全体的に多様性に対して抵抗(違いを拒否)する段階は超えたものの、依然として違いを一緒にしよとしている気がします。技術・人材と並び必要とされる寛容性・開放性という側面で、違いを認め合い、その次の段階である「違いを生かして競争優位性につなげる」努力が必要だと思います。

Proposed Method: 技術・人材と並び必要とされる寛容性・開放性という側面で、違いを認め合い、その次の段階である「違いを生かして競争優位性につなげる」努力が必要だと思います。

First Sentence: 返信が遅れまして申し訳ございません。

[Example 2]

Original Texts: 定光寺の緑は大自然と言ったほうが的確かもしれませんね。でもあの定光寺、川を挟んで春日井市と瀬戸市という管轄になっていて、今ではすっかり心霊スポットみたいになっています。母から聞いた話では昔は尾張名古屋の奥座敷として賑わっていたようですが。この場所の他にも名古屋市内ではないものの、名古屋圏域にはたくさんの魅力的な自然があります。だからこそ行政区分を超えた議論の場が必要だと思うのですがいかがでしょうか？

Proposed Method: だからこそ行政区分を超えた議論の場が必要だと思うのですがいかがでしょうか?

First Sentence: 定光寺の緑は大自然と言ったほうが的確かもしれませんね。

Appendix B: Examples of Extracting the Sentence for the Headline from the Original Post

In this example, our proposed approach extracted the following sentence as the headline. Different four methods generated the following headline from the extracted sentence.

[Example 1]

Extracted sentence: 技術・人材と並び必要とされる寛容性・開放性という側面で、違いを認め合い、その次の段階である「違いを生かして競争優位性につなげる」努力が必要だと思います。

Template Only: 認め合い次の段階である努力が必要

First Twenty Letters: 技術人材と並び必要とされる

Hybrid (BM-25): 認め合い次の段階である努力が必要

Hybrid (tf-idf): 認め合い次の段階である努力が必要

[Example 2]

Extracted sentence] だからこそ行政区分を超えた議論の場が必要だと思うのですがいかがでしょうか?

Template Only: 行政区分を超えた議論の場

First Twenty Letters: だからこそ行政区分を超えた議論

Hybrid (BM-25): 行政区分を超えた議論の場

Hybrid (tf-idf): 行政区分を超えた議論の場

Ryota Kitagawa was undergraduate student of faculty of Engineering, Tokyo University of Agriculture and Technology. He received the B.S. degree from Departments of Computer and Information Sciences, Faculty of Engineering, Tokyo University of Agriculture and Technology, Japan, in 2016. His research interests are in text mining and automatic summarization.

Katsuhide Fujita is an Associate Professor of Faculty of Engineering, Tokyo University of Agriculture and Technology. He received the B.E., M.E, and Doctor of Engineering from the Nagoya Institute of Technology in 2008, 2010, and 2011, respectively. From 2010 to 2011, he was a research fellow of the Japan Society for the Promotion of Science (JSPS). From 2010 to 2011, he was a visiting researcher at MIT Sloan School of Management. From 2011 to 2012, he was a Project Researcher of School of Engineering, the University of Tokyo.

He is an Associate Professor of Faculty of Engineering, Tokyo University of Agriculture and Technology since 2012. His main research interests include multi-agent systems, natural language processing and decision support systems.