

## Opinion sentence and topic relevant sentence extraction by using coherent structure among the sentences

Hironori Mizuguchi Masaaki Tsuchida Kenji Tateishi Dai Kusui  
NEC Common Platform Software Researches Laboratories  
8916-47 Takayamacho, Ikoma, Nara 630-0101 Japan  
{hironori@ab, m-tsuchida@cq, k-tateishi@bq, kusui@ct}.jp.nec.com

### Abstract

*We developed a new sentence extraction framework, the Sliding Window Framework, by using coherent structure among the sentences. Coherent structure means that the sentences that relate to a certain topic in an article are written in clusters to preserve the logical organization. To use the structure, our method makes blocks that consist of sentences in a window of a certain size, then estimates the score of each block, and judges each sentence from the scores. We applied our framework to opinion sentence extraction and topic relevant sentence extraction. In the result of our experiments, our framework achieved a very high recall ratio and a high F-value.*

**Keywords:** *Opinion sentence extraction, Relevant sentence extraction, coherent structure.*

### 1 Introduction

In recent years, people have been able to easily distribute information through the Internet. In that information, there are many opinions about products or news. If we are able to extract and analyze these opinions, we can analyze products' markets and investigate public opinion.

Our research group is studying reputation information extraction[7][8] mainly for information on product review sites. Reputation information is information that contains expression of the evaluation of a product or service and so on. For example, “*Let it be* はすごく良い (*Let it be* is very nice)” includes the expression of the evaluation, “良い (nice)”.

Based on such background, we participated in the two Japanese subtasks at MOAT of NTCIR-7: (1) the opinionated sentence extraction subtask that judges whether each sentence in news articles is opinion or not, and (2) the topic relevant sentence extraction subtask that judges whether each opinion sentence relates to the topic of the article given by the task organizer beforehand. An opinion sentence contains not only

reputation information but also suggestion information. For example, “水口氏は大統領を信任すべきだといった (Mr. Mizuguchi said we should trust in the president)” is an opinion sentence that is a suggestion but does not give reputation information.

In this paper, we propose a new sentence extraction framework, ‘Sliding Window Framework (SWF),’ and apply this framework to the opinion sentence extraction and the topic relevant sentence extraction. SWF is a general framework that can be used for the sentence extraction tasks that have the property of coherent structure. The coherent structure means that the sentences related to a certain topic in an article are written naturally in cluster in order to preserve the logical organization and the readability of the article. This observation indicates that opinions in an article tend to be written in the same paragraph to distinguish opinions from facts, and that sentence related to a certain topic are likely to be written in the same paragraph to avoid confusion with other topics.

Specifically, SWF utilizes the surrounding sentences of the target sentence as the coherent structure, and judges the target sentence whether opinion-related or not, or whether topic-related or not. It processes the following three steps. It first extracts blocks from each article by sliding the first sentence of each block. A block is composed of consecutive  $n$  sentence, and the constant  $n$  is given by the user. It then scores each block with a predefined function (function F) that evaluates the subjectivity or the topicality of the block. Finally, it scores each sentence with another predefined function (function G) that combines the function F scores related to the target sentence. SWF can be applied to a wide range of sentence extraction tasks by changing the function F and G according to the subtask.

The paper is organized as follows. First, we analyze the dataset to ascertain whether or not the opinionated sentences and topic relevant sentences have coherent structure. Then, we describe the details of the Sliding Window Framework and how we apply this framework to opinion extraction and topic relevant sentence

**Table 1. Results of investigating opinions**

Distance	Num. of Sentences	Ratio(%)
1	1486	49.97
2	476	16.01
3	274	9.21
4	169	5.68
5	114	3.83
6	95	3.19
7	67	2.25
8	40	1.34
9	34	1.14
10	21	0.71
Other	169	5.68

**Table 2. Results of investigating topic relevant sentences**

Distance	Num. of Sentences	Ratio(%)
1	5169	80.51
2	387	6.03
3	147	2.29
4	90	1.4
5	39	0.61
6	39	0.61
7	36	0.56
8	30	0.47
9	9	0.14
10	11	0.17
Other	64	1

extraction. Finally, we evaluate two subtasks.

## 2 Coherent structure of opinion sentences and topic relevant sentences

In this section, we reveal whether the opinion sentences and topic relevant sentences have coherent structure.

Empirically, fact and opinion are written in different parts of an article to avoid confusing the two. The sentences that relate to a certain topic are written in the same place. In this paper, coherent structure means that the sentences that relate to a certain theme are written in clusters to preserve logical organization.

We analyze the dataset of the Opinion Analysis Pilot Task in NTCIR6. We investigate whether the opinion sentences and topic relevant sentences are written in clusters by using lenient result data.

Table 1 shows the results for opinion sentences. Each row consists of the distance between the opinion sentence and the next opinion sentence, the number of opinion sentences and the ratio. For example, the second row shows the data for the opinion sentences whose next sentence is not opinion and the next sentence is opinion. The number of these sentences in the whole dataset is 476 and the ratio of 476 of the number of all opinion sentences is 16.01%.

According to the result, there are about 76% of opinion sentences with the distance less than three sentences. So, opinion sentences have coherent structure.

Table 2 shows the results for the topic relevant sentence. Each row consists of the distance between the topic relevant sentence and the next topic relevant sentence, the number of topic relevant sentences and the ratio as same as Table 1.

According to the result, there are about 80% of sentences with the distance less than one. So, topic relevant sentences have coherent structure too.

## 3 Sliding Window Framework

The results in previous section showed that opinion sentences and topic relevant sentences are written in clusters. So, we should consider a block that consists of continuous sentences to extract these sentences. To consider that, we propose a new sentence extraction framework called the 'Sliding Window Framework (SWF).'

SWF estimates the block score of the relationship between a certain topic and each block that has continuous sentences in a window (A window is a frame to make blocks and has a size that is the number of sentences in the block). This framework can judge each sentence by the block scores related to the target sentence. It consists of three steps:

**STEP1** Make blocks by sliding the window from sentence to sentence.

**STEP2** Estimate the score of the relationship of each block by using predefined function F.

**STEP3** Judge whether each sentence should be extracted by using function G with the score of blocks that contain the target sentence.

Figure 1 shows an example when window size is 3. In the figure, S1 through S5 denote sentence each. First, blocks B1, B2 and B3 are made by sliding the window to each sentence. Each block has 3 sentences. Next, block scores BS1, BS2 and BS3 are estimated by using the predefined function F. Function F inputs the information on the block and outputs the score of the relationship of the specified theme and the block. Finally, we can get results R1 through R5 as the results of judgments of sentences S1 through S5, respectively by using the predefined function G. Function G inputs the scores of blocks that contain the target sentence and outputs the results of judgment. For example, result R2 is the result of the judgment of sentence S2 and

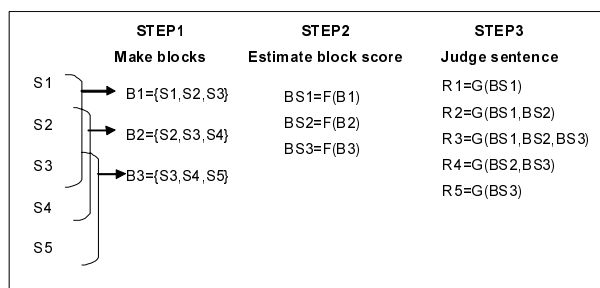


Figure 1. Steps of Sliding Window Framework

is made by using blocks B1 and B2, which contain the sentence S2.

In SWF, we define function F and function G according to each problem. In opinion sentence extraction, function F estimates the number of opinion sentences in the input block. Function G judges the sentence as opinion if the sum of the estimated number of opinion sentences of the blocks is greater than window size. In topic relevant sentence extraction, function F returns the similarity between the topic and the block. Function G judges the sentence as topic relevant if the average of the similarities from the blocks is greater than the predefined threshold.

### 3.1 Opinion sentence extraction

In this section, we describe how we can apply our framework to opinion sentence extraction. To apply our framework, we only define function F and function G. First, we define function F, and then define function G.

Function F in SWF is a regression function that returns the number of opinion sentences in the block. This function is learned from the training dataset. At learning the function, the feature is from the results of natural language analysis, such as morphological features and clause information. Concretely speaking, we use the following features. To learn the function, we make the vector of these features from the sentences in the block. The value of the vector is the frequency of each feature.

**Feature 1** Original form, part of speech and surface string of morpheme

**Feature 2** Semantic attribute of clause

**Feature 3** Pair of semantic attributes of two clauses of dependency relation

**Feature 4** Whether or not character “ 「 ” and “ 」 ” are in the same sentence

**Feature 5** Original form, part of speech and surface string of morpheme between “ 「 ” and “ 」 ”

**Feature 6** Original form, part of speech and surface string of morpheme before “ 「 ”

**Feature 7** Original form, part of speech and surface string of morpheme after “ 」 ”

Features 5, 6 and 7 are added to the vector as the different element from Feature 1. The characters “ 「 ” and “ 」 ” are Japanese characters to enclose something to say, like open quotation mark (“) and closing quotation mark (”) in English. Opinion contains some statements. So we use these characters to detect opinion.

Function G returns that the sentence is opinion if the sum of the scores returned by function F with the blocks that contain the target sentence is greater than the window size. The reason for using the window size is that it equals the number of blocks in which the target sentence appears. So, function G judges the sentence as opinion if all blocks that contain the target sentence have one or more opinion sentences.

### 3.2 Topic relevant sentence extraction

In this section, we describe how we can apply our framework to topic relevant sentence extraction. To adapt our framework to this task we define function F as the cosine similarity function between the topic of the article and the block, and function G as the average of the scores returned by function F.

However, there are not enough words in the topic description to calculate the cosine similarity. To solve this problem, in TREC 2003 Robust Retrieval Track[9], several groups got good results by using web expansion. In the NTCIR6 Opinion Pilot Task, [5] and [3] got good results too by using tf or tf-idf weight at calculating cosine similarity. Therefore, we extend the topic description by web information and we use tf-idf weight. The steps of topic rerouted sentence extraction are the following:

**STEP0** Make extended topic description by using web expansion

**STEP1** Make blocks

**STEP2** Calculate cosine similarity between extended topic description and block

**STEP3** Judge whether the sentence is a topic relevant sentence by using the score of STEP 2

At Step 0, the extended topic description is the word set in the topic description and the snippet of web search results of keywords in the TITLE part of the topic description. The reason for using the snippet is that the words in the snippet strongly relate to the keywords because of using the words near keywords as snippets.

At Step 1, our framework makes blocks in the same way as section 3.

At Step 2, the cosine similarity function is function  $F$  in section 3. This function returns the similarity between the block and the topic of the article.

$$F(B_i) = Sim(B_i, T) = \frac{\sum_j w_{B_i,j} w_{T,j}}{\sqrt{\sum_j w_{B_i,j}^2} \sqrt{\sum_j w_{T,j}^2}}$$

$B_i$  and  $T$  denote the word vector of  $i$ -th block and the word vector of extended topic description  $T$ , respectively.  $w_{B_i,j}$  is  $j$ -th word of vector  $B_i$ .  $w_{T,j}$  is  $j$ -th word of vector  $T$ . The  $N$  most frequent words in the snippet are added to vector  $T$ . The target parts of speech of the words are noun, verb, adjective and adverb. The weight of each word is the tf-idf value:

$$w_i = \frac{tf_i}{\sum_k tf_k} \log \left( \frac{D}{df_i} \right)$$

$tf_i$  is word frequency of  $i$ -th word in the article.  $D$  is the number of whole articles.  $df_i$  is frequency of the articles where the  $i$ -th word appears.

At Step 3, function  $G$  returns that the sentence is topic relevant if the average of the similarities by function  $F$  with the blocks that contain the target sentence is greater than the predefined threshold.

## 4 Evaluation Results

We evaluate our framework by using the NTCIR7 MOAT dataset. We used our Japanese analysis engine[4] for morphological analysis and syntactic analysis. Table 3 and Table 4 show the results of the formal run. We extracted opinion sentence and topic relevant sentence separately and then joined the two results.

In the result, "SWF" is the method of our proposed framework and the window size is 3. In opinion sentence extraction, to learn the function  $F$ , we used Support Vector Regression. In topic relevant sentence extraction, we add 100 words in 20 snippets from Yahoo API<sup>1</sup> to the topic description as the extended topic description. The threshold of function  $G$  is 0.1. These parameters were determined by preliminary experiment.

Baseline is the result of using each sentence without our framework. Baseline of opinion sentence extraction is the method classifying the sentence into opinion by Support Vector Machine with the features written at section 3.1 from each sentence. The baseline of topic relevant sentence extraction is based on our framework, however, window size is one and extended topic information is not used.

According to the results, F-value was improved against the baseline. Precision was down at all results but recall was greatly improved. F-value became

<sup>1</sup><http://www.yahoo.co.jp/>

high because of very high recall. The Sliding Window Framework can judge the sentence by using not only the target sentence but also the sentences near the target. Therefore, our system easily judges the sentence as true if the sentences near the target are opinion or topic relevant sentences.

We got the results of other systems from the organizer of MOAT[6]. Our system has a good F-value and recall of opinion and topic relevant sentence extraction. At the opinion sentence extraction subtask in Japanese, 12 runs from 8 groups were submitted and one group run had the same results in the opinion evaluation, so we had 11 unique runs. Under the lenient standard, our system had the second best performance for F-value, top performance for recall, but 8-th for precision. At topic relevant sentence extraction in Japanese, 6 runs from 4 groups were submitted and one group had the same results in the evaluation, so 5 runs were given. Under the lenient standard, we had top performance for F-value and recall, but 3-rd for precision.

We describe opinion sentence extraction evaluation and topic relevant sentence evaluation in the following sections.

### 4.1 Evaluation of opinion sentence extraction

At this evaluation, to reveal the relationship between the window size and accuracy, we check precision and recall by changing the window size. We consider that the precision is going down and the recall is going up with expanding window size. We observe precision, recall and F-measure while changing the window size from 1 through 5. But, we should be careful about the difference between the baseline introduced in previous and the case of window size 1. Baseline is learned by Support Vector Machine; however, the case of window size 1 is learned by Support Vector Regression(SVR)[2].

We use SVR to learn the function  $F$ , and use LibSVM[1] as the implementation of SVR with linear kernel. The lenient data from sample dataset in NTCIR7 MOAT is used as the training data

Table 5 shows the result of opinion sentence extraction. "(ALL Y)" is the result in the case of all sentences as opinion. According to the result, the precision was going down and the recall was going up while extending window size, as we predicted. While expanding the window size, the results were approaching "(ALL Y)".

### 4.2 Evaluation of topic relevant sentence extraction

We describe the results of the extended topic description and Sliding Window Framework separately.

**Table 3. Results (Lenient)**

	Opinion			Relevant		
	Precision	Recall	F-value	Precision	Recall	F-value
SWF	49.21	<b>73.13</b>	<b>58.83</b>	48.19	<b>63.54</b>	<b>54.81</b>
Baseline	<b>63.89</b>	51.79	57.21	<b>53.88</b>	17.96	26.94

**Table 4. Results (Strict)**

	Opinion			Relevant		
	Precision	Recall	F-value	Precision	Recall	F-value
SWF	37.38	<b>76.27</b>	50.17	28.08	<b>73.21</b>	<b>40.59</b>
Baseline	<b>50.11</b>	55.77	<b>52.79</b>	<b>38.17</b>	25.36	30.47

**Table 5. Results of opinion sentence extraction with window size**

Window Size	Precision	Recall	F-value
1	91.88	8.64	15.79
2	64.29	38.62	48.25
3	49.21	73.13	<b>58.83</b>
4	39.82	86.13	54.46
5	35.14	92.71	50.96
- (ALL Y)	28.9	100	44.84

**Table 6. Results of the extended topic description**

Num. of word	Precision	Recall	F-value
0	51.2	31.84	39.26
5	<b>52</b>	38.91	44.51
10	51.48	40.14	45.11
50	50.16	43.81	46.77
100	48.37	50.48	49.4
ALL	46.38	68.03	55.16

#### 4.2.1 Extension of topic description

In this evaluation, we investigate the effect of extended topic description. So, we use our framework with window size 1. We observe precision, recall and F-measure while changing the number of words added to the topic description to extend the topic. We added N most frequent words to the topic description.

Table 6 shows the results under the lenient standard. According to the results, the precision and recall are increasing until 5 words. Therefore, the web expansion can collect good words. But in the case of greater than 10 words, the precision is going down. So, we need to decide the appropriate number of words.

**Table 7. Result of topic relevant sentence with window size**

N words	Window	Prec	Rec	F
0	1	51.2	31.84	39.26
	2	48.21	49.52	48.86
	3	47.32	56.46	51.49
	4	47	60.68	52.97
	5	46.41	63.27	53.54
5	1	52	38.91	44.51
	2	49.34	61.5	54.75
	3	48.69	70.61	57.64
	4	49	77.01	59.89
	5	48.34	79.18	60.03
100	1	48.37	50.48	49.4
	2	46.92	74.56	57.6
	3	46.32	83.13	59.49
	4	46.69	88.3	61.08
	5	46.21	91.16	<b>61.33</b>
- (ALL Y)		43.21	100	60.34

#### 4.2.2 Window size

We observe the precision, recall and F-value while changing the window size from 1 through 5 and changing the number of words added to the topic description. We evaluate no extended words, 5 words (top performance for precision in previous results), and 100 words (using at formal run).

Table 6 shows the result under the lenient standard. "(ALL Y)" is the result in the case of all sentences as topic relevant. According to the result, the precision was going down and recall greatly increasing even when the number of extended word was changed. In particular, when the window size was changed from 1 to 2, the average of the precision was down about 2.5% but the average of the recall was up about 20%.

## 5 Discussion

According to Table 5, our framework contributes to getting high recall. It serves our purpose, which is to judge the sentence as an opinion sentence when the target sentence is in the cluster of opinion. However, the precision is down. During expanding the window size, the precision is decreasing because our framework uses not only tight clusters but also the coarse clusters having only a few opinion sentences.

The reason for low precision is that our framework mistakes the non-opinion sentence between the opinion sentences as an opinion sentence. Therefore, to preserve high recall and minimize the decrease of precision, we have to correctly judge the non-opinion sentence between the opinion sentences as a non-opinion sentence. To make that correct judgment, in the future we will create a method using not only block information but also each sentence information.

According to Table 6, topic relevant sentence extraction also has the trend that our framework gets high recall and low precision. To minimize the decrease of precision, we can apply the same approach described above.

By comparing Table 6 to Table 7, our framework is more effective with extending topic description for F-value because of very high recall. However, our framework gets poor precision. Extension of topic description can increase precision and recall when we use the applicable number of extended words. Therefore, we can improve the accuracy if we decide the applicable number of extended words and develop a method making the correct judgment of a non-topic relevant sentence.

## 6 Conclusions

We proposed and evaluated our new sentence extraction framework, the "Sliding Window Framework." Our framework can use around sentence information while judging the target sentence. We applied our framework to the opinion sentence extraction subtask and the topic relevant sentence extraction subtask at NTCIR7 MOAT. As a result, we achieved high F-value because of very high recall. In comparing with other systems in NTCIR7, we got second best performance of opinion sentence extraction and best performance topic relevant sentence extraction for F-value under the lenient standard. However, our framework got poor precision with expanding the window size. To improve this, we will develop a method making the correct judgment of non-related sentences in the feature.

## References

- [1] C.-C. Chang and C.-J. Lin. LIBSVM: a library for support vector machines, 2001. (Software available at <http://www.csie.ntu.edu.tw/~cjlin/libsvm>).
- [2] H. Drucker, C. J. C. Burges, L. Kaufman, A. Smola, and V. Vapnik. Support vector regression machines. *Proceedings of Advances in Neural Information Processing Systems 9*, pages 155–161, 1996.
- [3] D. K. Evans. A low-resources approach to opinion analysis: Machine learning and simple approaches. *Proceedings of the 6th NTCIR Workshop Meeting on Evaluation of Information Access Technologies: Information Retrieval, Question Answering and Cross-Lingual Information Access (NTCIR 6)*, pages 290–295, May 2007.
- [4] Y. Sakao, T. Ikeda, K. Satoh, and S. Akamine. Japanese language analysis for syntactic tree mining to extract characteristic contents. *Proceedings of The Tenth Machine Translation Summit (MT Summit X)*, pages 339–345, 2005.
- [5] Y. Seki. Crosslingual opinion extraction from author and authority viewpoints at NTCIR-6. *Proceedings of the 6th NTCIR Workshop Meeting on Evaluation of Information Access Technologies: Information Retrieval, Question Answering and Cross-Lingual Information Access (NTCIR 6)*, pages 336–343, May 2007.
- [6] Y. Seki, D. K. Evans, L.-W. Ku, L. Sun, H.-H. Chen, and N. Kando. Overview of multilingual opinion analysis task at NTCIR-7. *Proceedings of the 7th NTCIR Workshop Meeting on Evaluation of Information Access Technologies: Information Retrieval, Question Answering and Cross-Lingual Information Access (NTCIR 7)*, December 2008.
- [7] K. Tateishi, Y. Ishiguro, and T. Fukushima. A reputation search engine that collects people's opinions by information extraction technology. *IPSJ Transactions on Databases*, 45(No. SIG 7 (TOD 22)):115–123, 2004.
- [8] M. Tsuchida, H. Mizuguchi, and D. Kusui. Opinion extraction by identifying object-attribute-evaluate relations. *Proceedings of 13th Annual Meeting of the Association for Natural Language Processing*, pages 412–415, March 2007. (in Japanese).
- [9] E. M. Voorhees. Overview of the trec 2003 robust retrieval track. *Proceedings of the Twelfth Text REtrieval Conference (TREC 2003)*, pages 69–77, 2003.