Interactive Aspect Summarization Using Word-Aspect Relations for Review Documents

Kazutaka Shimada, Masashi Yamaumi, Ryosuke Tadano, Masashi Hadano and Tsutomu Endo
Department of Artificial Intelligence, Kyushu Institute of Technology
680-4 Iizuka Fukuoka 820-8502 Japan
{shimada, m_yamaumi, r_tadano, m_hadano, endo}@pluto.ai.kyutech.ac.jp

***Abstract*—As the World Wide Web rapidly grows, a huge number of online documents are easily accessible on the Web. We obtain a huge number of review documents that include user's opinions for products. To summarize the opinions is one of the hottest topics in natural language processing. We focus on aspects of a product in the summarization process. First, we identify a relation between aspects and each word in review documents. Our method employs an unsupervised approach for the identification process. The experimental result shows the effectiveness of the method. Next, we generate a summary by using the relations. In our system, users obtain the summary by an interactive approach.**

I. Introduction

As the World Wide Web rapidly grows, a huge number of online documents are easily accessible on the Web. Finding information relevant to user needs has become increasingly important. The most important information on the Web is usually contained in the text. We obtain a huge number of review documents that include user's opinions of products. When buying a product, users usually survey reviews of the product. More precise and effective methods for evaluating the products are useful for users. To analyze the opinions is one of the hottest topics in natural language processing [1], [9], [12], [18].

In this paper we focus on a summarization task for review documents of a product. In review summarization tasks, systems often treat aspects of the target products. Lu and Zhai [9] have proposed an opinion integration method with aspects using the PLSA model. The purpose of their study is to add supplementary opinions and similar opinions into reviews. Blair-Goldensohn et al [1] have proposed a system that summarizes the sentiment of reviews for a local service such as a restaurant or hotel. They classified the polarity of sentences by using the maximum entropy method with WordNet and ratings in reviews. However, machine learning methods usually need a large amount of training data to generate a high accuracy classifier. They also reported two aspect extractors; dynamic aspect extraction and static aspect extraction They combined them and summarized review texts. However, sentences with high polarity scores are not always suitable for summarized texts because of the variety of users' needs.

In this paper we present an aspect summarization system for review documents. First, we identify aspects of each word in the target review documents. In this process, we employ two characteristics in the review documents; structure information and rating information. Next, we identify the aspect of a word by using features from structure information and rating information. Finally we summarize the review documents on the basis of the word-aspect relations. We apply an interactive approach to the summarization process for the variety of users' needs. In the process, the system displays words with high aspect scores from the word-aspect identification process first. Then it extracts output sentences from the review documents by using the words selected by users.

In Section 2, we explain the aspect identification process for the summarization. In Section 3, we describe the summarization process using word-aspect relations acquired in the previous section and conclude this paper in Section 4.

II. Word-Aspect Identification

For summarization, we need to identify relations between words and aspects in review documents. Figure 1 shows the outline of the word-aspect relation identification process. In this section, we explain a method to extract features for the identification process first. Next, we propose an identification method based on a scoring approach with the extracted features. Finally, we evaluate the identification method with a dataset.

A. Feature extraction

We focus on two characteristics in the feature extraction; structure information and rating information in review documents. In this process, we handle nouns, adjectives verbs[1] and adverbs. We use the morphological analyzer ChaSen [11].

1) Structure information:
Structure information in review documents is useful for sentiment analysis. Kaji and Kitsuregawa [6] have proposed a method for building sentiment lexicon by using layout structures; itemization structure and table structure.

In our dataset, users often itemize opinions in the reviews of a product[2]. In multi aspect review documents, the itemization is usually based on each aspect in the review form. We focus this structural information (SI) for the feature

[1]We use a stopword list for verbs. The stopword list includes high frequency verbs such as "is" and 'think".

[2]As a preliminary experiment, we examined 670 review documents in a dataset. The result said that 130 documents in them contained itemization.

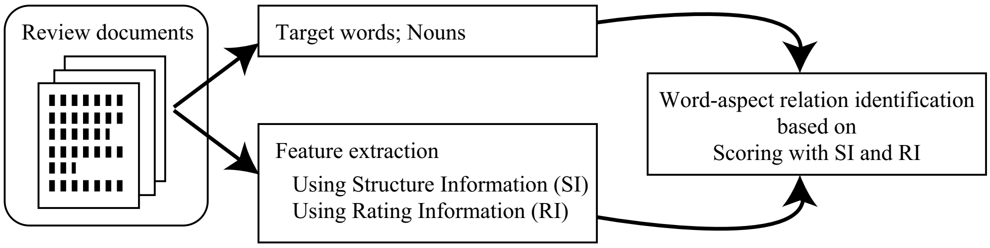

Fig. 1. The outline of word-aspect identification.

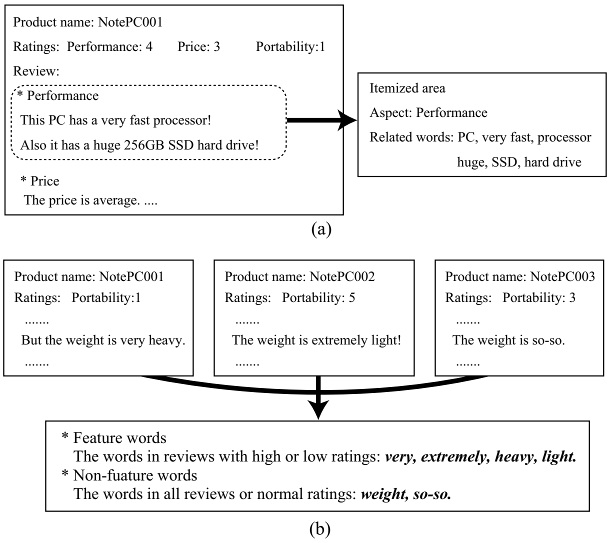

Fig. 2. The characteristics for the feature extraction.

extraction. For the process, we use template patterns such as

{Line head mark} + {Aspect name} +
{Separator or New line} + {Sentence area}

We regard words in the "Sentence area" as related words for the "Aspect name". Figure 2 (a) shows an example of the characteristic. In the figure, "*" is the line head marker. Our system detects "sentence area" on the basis of the next line head marker. Then, it extracts characteristic words such as nouns from the sentence area. We regard the extracted words as the related words of the aspect "Performance".

2) Rating information:
Another characteristic is based on rating information (RI) in review documents. Rating information is one of the most important features for sentiment analysis [14] and sentiment summarization [10], [17]. Reviews with a high or low rating for an aspect contain sentences related to the aspect. The target is a review document of which an aspect is a high or low value and other aspects are normal values. Figure 2 (b) shows an example of the characteristic. In the figure, the three reviews contain 1, 3 and 5 points for the aspect "Portability", respectively. Here we assume that the range of the rating is 1-5 points. Our system extracts words that are included in reviews with high (5 points) or low (1 point) ratings and ignores words that are in reviews with average ratings (3 or 4 points). In the figure, the words such as "heavy" and "light" are associated with the aspect "Portability". On the other hand, the word "weight" is deleted from the characteristic word list because the word appears in many reviews evenly.

B. Scoring

We identify the aspect of a word by using SI and RI features. First, we compute the scores of words of SI and RI features for each aspect.

$$SIScore_{asp}^{w_i} \text{ or } RIScore_{asp}^{w_i} = \log \frac{tf_i \cdot R}{N \cdot R_i} \tag{1}$$

where R and R_i denote the number of review documents and the number of documents containing word w_i, respectively. tf_i and N are the frequency of w_i and the number of words in the sentences about a feature set (SI or RI) and an aspect (asp), respectively. We normalize the scores to [0,1]. The score of a word w_i is computed as follows:

$$SumScore_{asp}^{w_i} = SIScore_{asp}^{w_i} + RIScore_{asp}^{w_i} \tag{2}$$

Next, we identify the aspect of a target word by using the scores. In this process, we use four words in front and backward of the target word in the sentence as features for the identification. The score of a word t_i belonging to an aspect asp is computed as follows:

$$Score_{asp}^{t_i} = \sum_{f_j \in F_i} Freq_{f_j} \times SumScore_{asp}^{f_i} \times Weight_{POS} \tag{3}$$

where F_i is features for the target word t_i. $Freq_{f_j}$ is the frequency of f_j. $Weight_{POS}$ is a weight value of the part-of-speech tag of the f_j. In this paper, the weights of nouns, adjectives verbs and adverbs are 0.5, 1.0, 0.1 and 0.3 respectively. These values are determined experimentally. If f_j does not exist in the feature set explained in Section II-A, the value is 0. We select the aspect that contains the maximum score as the aspect of the word. Figure 3 shows an example of the scoring process. First our system detects sentences which contain the target word "image". Next, it extracts words which surround the target word. Here we assume that there are three aspects and a word list with scores which are computed by Eq (1) and Eq (2). Our system computes $Score_{asp}^{t_i}$ of each aspect by using Eq (3). In this example the aspect of the target word "image" is the aspect y.

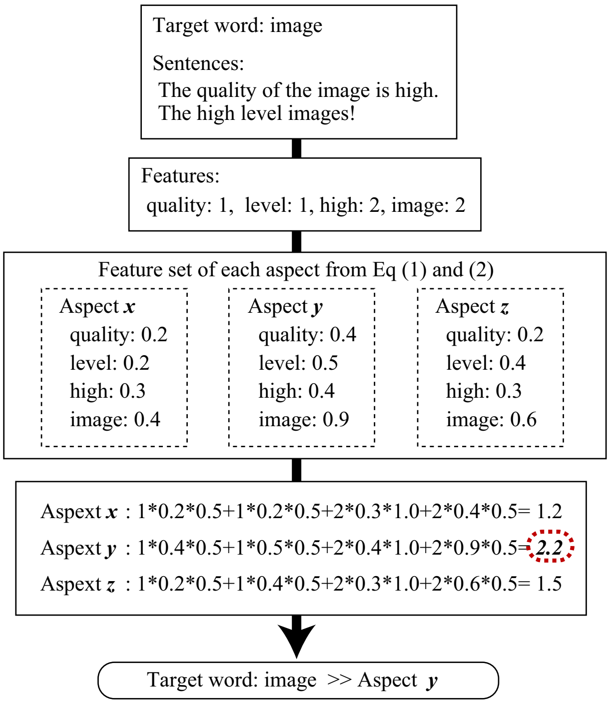

Fig. 3. An example of the scoring process.

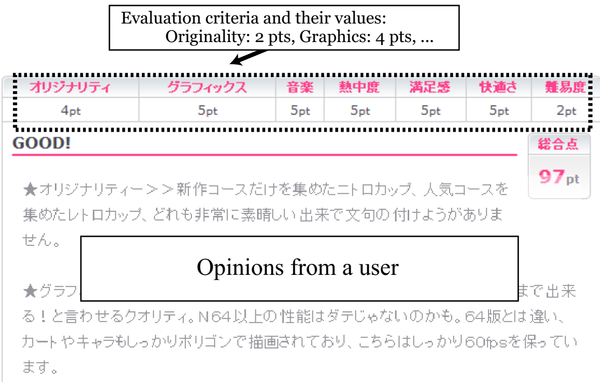

Fig. 4. An example of a review document.

C. Experiment

In this paper we handle review documents about game softwares. We extracted 4174 review documents from a Web site[3]. We constructed a feature set described in Section II-A from them. Figure 4 shows an example of a review document. The review documents consist of evaluation criteria, their ratings, positive opinions, negative opinions and comments for a product. The number of evaluation criteria is 7: "Originality (O)", "Graphics (G)", "Music (M)", "Addiction (A)", "Satisfaction (S)", "Comfort (C)", and "Difficulty (D)". In this paper, we regard the evaluation criteria as aspects of the review documents.

First, we evaluated our method with 130 review documents

[3]http://ndsmk2.net

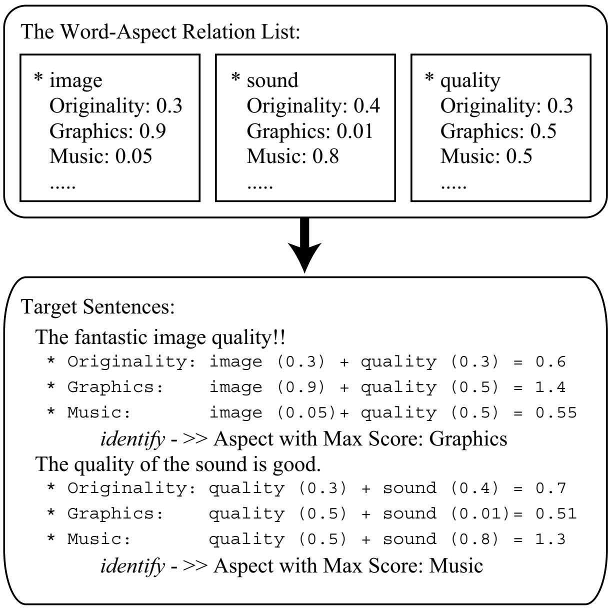

Fig. 5. Examples of the sentence classification task.

about a game software. The number of words in the documents was 1817. We selected 20 words for each aspect, namely 140 words, from them as the test dataset. Table I shows the experimental result. "Words" in the table denotes the number of words that were classified into each aspect. The average of the recall rate was 72.1%. The precision rates of "Originality" and "Addiction" were lower than the others. The reason why the precision rates were low was that "Originality" and "Addiction" are generally related to "Satisfaction". In this experiment, words about "Satisfaction" were often incorrectly classified into "Originality" . Tadano et al. [15] have reported that even human beings tend to confound words about "Originality" with words about "Satisfaction" in a sentiment annotation task.

The tabel I was the result of an evaluation with limited words. Next, we evaluated the word-aspect relations identified by our method on overall viewpoint. We applied the word-aspect relations into a sentence classification task. If the word-aspect relations is applicable, each sentence is classified into the correct aspect. We used 700 test sentences with aspects that were annotated by handwork. Figure 5 shows an example of the sentence classification task. In this task, we computed the sum of the scores of words in each sentence by using the word-aspect relations and scores by our method. We selected the aspect that contains the maximum score as the aspect of the sentence. The accuracy of the sentence aspect identification was approximately 65%. Our method is an unsupervised approach using structure information (SI) and rating information (RI). In other words, our method can identify the aspect of a sentence with a non-tagged corpus.

TABLE I
THE ACCURACY OF WORD-ASPECT RELATIONS.

Aspects	O	G	M	A	S	C	D
Recall	75% (15/20)	90% (18/20)	100% (20/20)	75% (15/20)	45% (9/20)	75% (15/20)	45% (9/20)
Precision	56% (15/27)	90% (18/20)	91% (20/22)	58% (15/26)	82% (9/11)	75% (15/20)	75% (9/12)
Words	421	112	131	386	157	356	225

III. SUMMARIZATION

In this section, we describe a method of review summarization with the word-aspect relations and scores explained in the previous section. We apply an interactive approach to the summarization process for the variety of users' needs. Our summarizer consists of the following three steps;

Step 1 : display words with high aspect scores for each aspect,

Step 2 : display words of the aspect that a user selects in the step 1,

Step 3 : display sentences related to the word that a user selects in the step 2.

In the 1st step and 2nd step, we visualize the data by using the Tree-Map style [5]. The Tree-Map is a mapping approach with rectangular regions in a space. The size of each rectangle is based on the aspect score computed in Section II-B. In other words, the size implies the number of mentions and the importance of the word and aspect in the target review documents.

First, a user selects an aspect in the Tree-Map in the 1st step. Then, the system extracts words that are related to the aspect from the word-aspect relation list. Next, the user selects a word in the Tree-Map in the 2nd step. Our system outputs sentences that contain the selected word in the target review documents.

In the step 3, our system displays clustered sentences for the readability. We use a $k-$means based method that has been proposed by Seki et al [13]. This method determines the optimal k in the $k-$means method. The method computes a rate of change in $k-1, k$, and $k+1$. We use a statistic proposed by Krzanowski and Lai [7] (in this paper, hereinafter referred to as KL statistic). The statistic $KL(k)$ where the number of clusters is k is computed as follows:

$$KL(k) = \left| \frac{DIFF(k)}{DIFF(k+1)} \right| \tag{4}$$

$$DIFF(k) = (k-1)^{2/m} W(k-1) - k^{2/m} W(k) \tag{5}$$

where m is the dimensions of the vector space. $W(k)$ is computed as follows:

$$W(k) = \sum_{j=1}^{k} ss(j) \tag{6}$$

$W(k)$ is the summation of the sum of squares (ss) in a cluster. $W(k)$ becomes small in the case that each cluster contains similar words only. In the (6), the $ss(j)$ of a cluster π_j is computed as follows:

$$ss(j) = \sum_{\mathbf{x} \in \pi_j} \|\mathbf{x} - \mathbf{m}_j\|^2 \tag{7}$$

$ss(j)$ is the sum of squares of the distance between the mean vector ($\mathbf{m}$) and each vector ($\mathbf{x}$) belonging to a cluster π_j. The value of ss is large if the variance of a cluster is large.

$KL(k)$ is a criterion to evaluate a rate of change by considering the dimensions of the vector space. $KL(k)$ is large in the case that $DIFF(k)$ is large and $DIFF(k+1)$ is small. In other words, KL is large in the case that the values of the summation of ss are large on the change from $k-1$ to k and small on the change from k to $k+1$. They concluded that the k is optimal because this denotes that the clustering process improves the output on the k as compared with the $k-1$ and does not improve the output on the $k+1$ as compared with the k.

In this step, we visualize sentences by using the fisheye-like style [2]. The fisheye view is a distorted view method of a data set. First we select a representative sentence that contains the word in the 2nd step from each cluster. The representative sentence is essentially the centroid of the vectors contained in the cluster. Then we compute a similarity between the representative sentence and each sentence in the cluster belonging to the representative sentence. The similarity between two sentences is computed by

$$SIM(S_x, S_y) = \frac{2M_{xy}}{M_x + M_y} \tag{8}$$

where M_{xy} is the number of morphemes matched between a sentence S_x and a sentence S_y. M_x and M_y are the number of morphemes in S_x and S_y. We display sentences with the fisheye style on the basis of the similarity.

Figure 6 shows an example of the summarization process. In this example, the user selected the aspect “comfort” in the 1st step. Therefore, the system displayed 30 words belonging to the selected aspect “comfort” in the 2nd step. Next, the user selected the word “save” in the word list. In the 3rd step, the summarizer classified sentences into several clusters, and extracted sentences that contained the word “save”. For the clustering process, our method tended to overdivide clusters. The improvement of this problem is our future work. If the output is, however, not suitable for user's needs, the user can obtain the summary that he/she wants by using the iteration of the step 2 and the step 3.

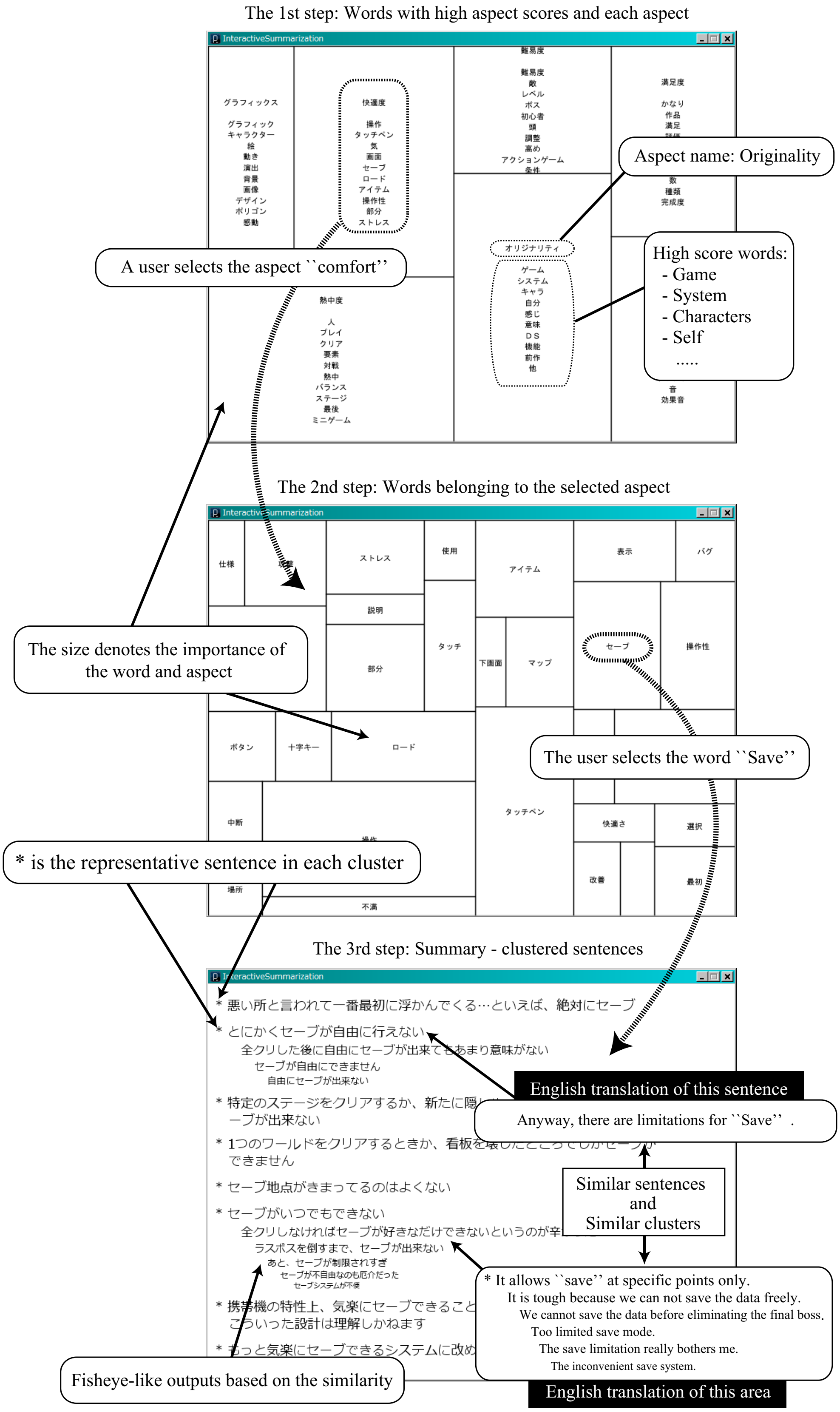

Fig. 6. An example of the summarization process.

IV. Discussion and Conclusions

In this paper we described an aspect summarization system for review documents. We identified aspects of each word in the target review documents on the basis of two characteristics; structure information and rating information. The average of the recall rate of the identification process was 72.1%. We also evaluated the word-aspect relations by our method with a sentence classification task. The accuracy of the sentence aspect identification was approximately 65% without a tagged corpus. However, the accuracy was insufficient. Titov and McDonald [17] have also proposed a unsupervised learning method of aspect identification with Multi-Grain Latent Dirichlet Allocation. Incorporating the related work into our method might lead to the improvement of the accuracy.

Then, we summarized the review documents by using the word-aspect relations identified by a scoring process with the characteristics. In the summarization process, we applied an interactive approach with the tree-map and fisheye-like styles. Gamon et al. [4] have reported a visual summarization system of product features. Carenini et al. [3] also have a visual summarization system with a natural language summary. There studies also applied the Tree-map style to the summarization. Our system realized a readable summary by applying fisheye-like style to the natural language summary. In the output process of related sentences, we used a $k-$means based clustering approach and a similarity measure. The clustering method determined the optimal k automatically. However, it tended to overdivide clusters. Therefore, we need to apply an integration process of the overdivided clusters. The similarity measure was based on the correspondence of morphemes. We need to compare it with other similarity measures. Besides, we need to quantitatively evaluate our method, e.g., the ROGUE-N [8]. We have studied a static summarization task with the review documents [16]. Applying the knowledge from the research into this dynamic and interactive summarization is one future work.

Acknowledgment

This work was supported by Kayamori Foundation of Informational Science Advancement.

References

[1] S. Blair-Goldensohn, K. Hannan, R. McDonald, T. Neylon, G. Reis, and J. Reynar. Building a sentiment summarizer for local service reviews. In *WWW Workshop on NLP in the Information Explosion Era (NLPIX)*, 2008.

[2] G. W. Furnas. Generalized fisheye views. In *In Proceedings of the ACM Conference on Human Factors in Computing Systems*, pages 16–23, 1986.

[3] Carenini G, Ng R, and Pauls A. Interactive multimedia summaries of evaluative text. In *Proceedings of the 10th International Conference on Intelligent User Interfaces (IUI 2006)*, 2006.

[4] Michael Gamon, Anthony Aue, Simon Corston-Oliver, and Eric K. Ringger. Pulse: Mining customer opinions from free text. In *Proceedings of the 2005 Conference on Intelligent Data Analysis (IDA)*, pages 121–132, 2005.

[5] Brian Johnson and Ben Shneiderman. Treemaps: A space-filling approach to the visualization of hierarchical information structures. In *Proc. of the 2nd International IEEE Visualization Conference*, pages 284–291, 1991.

[6] Nobuhiro Kaji and Masaru Kitsuregawa. Building lexicon for sentiment analysis from massive collection of html documents. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP-CoNLL2007)*, 2007.

[7] W. J. Krzanowski and Y. T Lai. A criterion for determining the number of clusters in a data set. *Biometrics*, 44:23–34, 1985.

[8] Chin-Yew Lin. Rouge: A package for automatic evaluation of summaries. In *Proceedings of the Workshop on Text Summarization*, 2004.

[9] Yue Lu and Chengxiang Zhai. Opinion integration through semisupervised topic modeling. In *In Proceedings of the 17th International World Wide Web Conference (WWW'08)*, 2008.

[10] Yue Lu, ChengXiang Zhai, and Neel Sundaresan. Rated aspect summarization of short comments. In *Proceedings of the 18th International World Wide Web Conference (WWW'09)*, pages 131–140, 2009.

[11] Y. Matsumoto, A. Kitauchi, T. Yamashita, Y. Hirano, H. Matsuda, , and M. Asahara. Japanese morphological analysis system chasen version 2.0 manual 2nd edition. Technical report, NAIST, 1999.

[12] B. Pang, L. Lee, and S. Vaithyanathan. Thumbs up? sentiment classification using machine learning techniques. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 79–86, 2002.

[13] Tsunehito Seki, Kazutaka Shimada, and Tsutomu Endo. Acquisition of synonyms using relations in product specification tables. *Systems and Computers in Japan*, 38(12):25–36, 2007.

[14] Kazutaka Shimada and Tsutomu Endo. Seeing several stars: a rating inference task for a document containing several evaluation criteria. In *Proceedings of the 12th Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD 2008)*, pages 1006–1014, 2008.

[15] Ryosuke Tadano, Kazutaka Shimada, and Tsutomu Endo. Effective construction and expansion of a sentiment corpus using an existing corpus and evaluative criteria estimation. In *Proceedings of the 11th Conference of the Pacific Association for Computational Linguistics (PACLING2009)*, pages 211–216, 2009.

[16] Ryosuke Tadano, Kazutaka Shimada, and Tsutomu Endo. Multi-aspects review summarization based on identification of important opinions and their similarity. In *Proceedings of the 24nd Pacific Asia Conference on Language, Information and Computation (PACLIC24)*, 2010.

[17] Ivan Titov and Ryan McDonald. A joint model of text and aspect ratings for sentiment summarization. In *Proceedings of the 46th Meeting of Association for Computational Linguistics (ACL-08)*, pages 308–316, 2008.

[18] P. D. Turney. Thumbs up? or thumbs down? semantic orientation applied to unsupervised classification of reviews. In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*, pages 417–424, 2002.