

Social Summarization of Text Feedback for Online Auctions and Interactive Presentation of the Summary

Yoshinori Hijikata Hanako Ohno Yukitaka Kusumura Shogo Nishida
Graduate School of Engineering Science, Osaka University
1-3 Machikaneyama, Toyonaka
Osaka 560-8531, Japan

{hijikata, ohno, kusumura, nishida}@nishilab.sys.es.osaka-u.ac.jp

ABSTRACT

Buyers in online auctions write feedback comments to the sellers from whom the buyers have bought the items. Other bidders read them to determine which item to bid for. In this research, we aim at helping bidders by summarizing the feedback comments. Firstly, we examine feedback comments in online auctions. From the results of the examination, we propose a method called social summarization method, which uses social relationships in online auctions for summarizing feedback comments. We implement a system based on our method and evaluate its effectiveness. Finally, we propose an interactive presentation method of the summaries based on the result of the evaluation.

Categories and Subject Descriptors

I.7.m [Document and Text Processing]: Miscellaneous;
H.3.3 [Information Storage and Retrieval]: Information Search and Retrieval

General Terms

Documentation, Experimentation, Human Factors

Keywords

text summarization, social relationship, online auction, feedback comments, social summarization method

1. INTRODUCTION

Online auctions have grown in popularity. In general, auction sites let a winning bidder to rate a seller by three levels, "good", "bad" or "neither". Users can see the number of ratings in each level. Also, auction sites let the bidder to write a review comment, hereinafter which is referred as "a feedback comment", on the seller who exhibited and sold the item to the bidder. Feedback comments on a certain seller inputted by past buyers are listed in one page. When the

user wants to buy an item, the user reads its item descriptions, which is an explanation of the item written by the seller. When the user decides which item to bid for, especially when there are multiple sellers who exhibit the same kind of item, the user checks the credibility of the sellers.

Ratings seem to be useful to check. However, most buyers tend to rate a seller as "good". Consequently, every seller has received a "good" rating on almost all his/her transactions. (If the seller has received several "bad" ratings, he usually closes his auction account.) The user can get more detailed information from feedback comments, such as the response of the seller and the packing of the item, which the user cannot acquire from ratings. Therefore, the user can compare sellers more deeply by using feedback comments. (If the auction site gives a feedback sheet for rating each attribute such as shipping, response and item, users will not need feedback comments. Currently auction sites do not provide this kind of feedback sheet. This is because we think that direct evaluation for the detail will be avoided by many people who want to keep good human relationships on the Web.) However, there are huge number of feedback comments even in one seller. The user needs considerable time and effort for looking through all feedback comments. Moreover, feedback comments include not only comments presenting real opinions but also many stereotyped sentences, clauses or phrases such as expressions for thanks or expressions of courtesy, hereinafter all of which are referred as "descriptions of courtesy". We also mean that a "description" in our paper is a semantic unit in comments, that is, a sentence, clause or phrase.

In this research, we aim at summarizing feedback comments to solve these problems. General summarizing methods give weight to each sentence based on some information and extract the sentences with higher weight from the document. Paice[10] classifies those information into seven types: (1) keyword frequency, (2) location, (3) title, (4) conjunctive sentence structure, (5) key description, (6) semantic relationship between words, (7) distance in vector space model. When we try to summarize fractionary sentences, which means that there is little context in sentences because each sentence is independent, like feedback comments, the only information we can use out of these seven types are (1) frequency and (5) key description. A general method using a frequency regards that descriptions which have a higher frequency are important. However, the summary made by this method includes many descriptions of courtesy because many people write them. A method using key descriptions

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

IUT'06, January 29–February 1, 2006, Sydney, Australia.
Copyright 2006 ACM 1-59593-287-9/06/0001 ...\$5.00.

can use a rule that eliminates predefined text patterns which are considered as descriptions of courtesy. However, even if there is a description that the buyer wrote from his/her real feelings and is important for the users, it will be eliminated if the rule matches the description. This results in that these useful descriptions will not be contained in the summary.

We propose a method called "social summarization method", for summarizing feedback comments. This method uses social relationships in an online auction for summarizing one seller's feedback comments. The main idea of this method is that for summarizing feedback comments on a seller, this method does not focus on the seller but focuses on a buyer who bought an item from the seller. This method compares the feedback comment on the target seller written by a certain buyer to the feedback comments on the sellers other than the target seller written by the buyer. Then it produces a summary by extracting two types of descriptions. One is a description that appears only in the feedback comment on the target seller and the other is a description that appears in the feedback comments on the sellers other than the target seller but not appear in the feedback comment on the target seller. By this method, we can eliminate descriptions of courtesy without deleting descriptions which are seemed that the buyers wrote from their real feelings.

In this paper, firstly, we investigate feedback comments in online auctions for discovering how to make sophisticated summaries. After that, we invent a method for summarizing feedback comments based on the results of the investigation. We implement a system based on our method and evaluate its effectiveness. Finally we propose an interactive presentation method of the summaries from the result of this evaluation.

2. RELATED WORKS

Feedback comments on online auctions can be regarded as a kind of reputation. As our related works, this section introduces some researches whose target are reputations. These researches are categorized into reputation collection, categorization and analysis[16]. Researches for collecting reputations find and extract text segments, in which reputations are written[7, 18]. Kobayashi et al.[7] assumed that a reputation mainly consists of a subject, feature and evaluation. Then, they proposed a method for extracting features and their evaluations from messages on BBS. Wiebe et al.[18] proposed a method for extracting several words (n-gram) which represent subjective opinions. Both researches extract subjective segments based on a dictionary which the researchers has created in advance by registering words with the values how much subjective they are.

There are many researches for categorizing reputations[17, 11, 3, 14, 2]. They categorized reputations as the positive or the negative. Turney[17], Pang et al.[11] and Dave et al.[3] focused on reputations about products and movies, Spertus[14] focused on opinions on politics, and Das et al.[2] focused on opinions on companies. Turney's method calculates the average semantic orientation of the phrases in the reputation. The mutual information between the given phrase and the word "excellent" or "poor" is used for this calculation. The semantic orientation represents the degree of positive or negative meaning and has been studied by many researchers[5, 6]. Pang et al. used a set of tuples (the frequencies of characteristic words in a reputation and ratings of movies) as training data. They categorized repu-

tations by using Naive Bayes, maximum entropy classification and support vector machine. Dave et al. used n-grams as features and a label (positive or negative) to a text inputted by the authors as training data. They categorized a text by calculating the frequencies of its n-grams in positive documents and those in negative documents. Spertus used characteristic words as features and used a decision-tree for the categorization. Das et al.'s method judges the class of a message by using five different categorization methods. It selects a class which more than three methods support.

Researches on reputation analysis analyze the reputation and visualize the results [19, 15]. Yamanishi et al.'s method [19] extracts characteristic words from the FAQs about some products and displays the result of correspondence analysis of those words in a two-dimensional map. Pero et al.[15] focused on 83 types of emotion. Their method calculates how many emotional expressions are written in a document by using the emotional intensities of words that were defined by human in advance. They displayed the results in a graph.

There are some researches which integrate reputation collection, categorization and analysis [16, 20, 9, 4, 8]. Tateishi et al.[16], Yi et al.[20] and Nasukawa et al.[9] proposed systems which collect reputations about products or companies on the Web and categorize them as the positive or the negative. Fujimura et al.[4] not only categorized reputations but also applied the created categorizing rules to the support of decision-making. Morinaga et al.[8] proposed a framework for collecting, categorizing and analyzing reputations.

Above-mentioned researches did not take into account whether or not reputations were written with real feeling. They only considered whether the reputation is positive or negative in one document. Also, reputations they focused on were reviews about products or services which anyone can use without seeing their producers or opinions on public things like politics and companies. People can freely remark their reviews or opinions without taking cares of the people who are related to the target. These reputations are not influenced by social relationships. Therefore, it is insufficient to use the results of the investigations or the methods in the above related works when summarizing feedback comments in online auctions.

3. INVESTIGATION

We investigated real feedback comments in an online auction in Japan (Feedback comments are written in Japanese). Firstly, we describe the investigations into contents and grammatical features of feedback comments. The purpose is to know what kind of content is in feedback comments and to know whether or not we can extract the above content by grammatical clues. Second, we describe the investigation into the frequencies of content patterns that appear in feedback comments. The purpose is to know whether or not we can use the most general method in natural language processing. Third, we describe the investigation into the frequencies of content patterns by every buyer.

3.1 Contents and Grammatical Features

We selected 100 sellers randomly and picked up the newest 10 feedback comments from each seller. We examined total of 1000 comments in terms of what contents were written in these comments (There are totally 3522 descriptions.). The result is that we can classify contents into 13 topics such as "item" and "response" shown in the first column of Table

Feature	Feature value	Example	Number	
thanks		thank you	869	869
greetings		will deal again	371	371
apologies		sorry for the delay of reply	68	68
arrivals	item	item arrives	193	655
	safely	I got safely	160	
	now	now I received	151	
	today	arrived today	48	
	quickly	item arrived quickly	32	
	others	received rightly	71	
response	prompt	prompt response	269	515
	courteous	response was courteous	106	
	kindly	responded kindly	42	
	quickly	responded quickly	19	
	smooth	response was very smooth	14	
	others	rapid response	65	
transactions	smooth	smooth transaction	101	286
	reliable	reliable transaction	63	
	comfortably	did business comfortably	49	
	prompt	prompt transaction	31	
	good	good transaction	20	
	others	satisfied with transaction	22	
shipping	prompt	prompt shipping	29	83
	others	speedy delivery	54	
communication	others	prompt communication	75	75
packing	courteous	courteous packing	46	67
	others	good packing	21	
sellers	dependable	the seller is dependable	24	37
	others	great seller	13	
payments	others	gave a credit yesterday	18	18
requests	others	great service	31	31
items	satisfied	very satisfied with item	64	345
	excellent	it was in excellent condition	51	
	beautiful	beautiful item	26	
	like	I like it	20	
	treasure	I'll keep it as treasure	13	
	others	quality product	171	
others			102	102
total			3522	3522

Table 1: The number of pairs of feature and feature value. (Feature values and examples are translated from Japanese to English)

1. We found that descriptions in the feedback comments represent how one of the above topics is like. When we examined what kind of keywords is used to represent the content about each topic, we found that there are various kinds of keywords. For example, there are keywords such as "prompt", "courteous" and "quickly" to the topic about "response", and "beautiful" and "like" to the topic about "item". From this result, we defined these topics as features and a keyword that represents the content about a topic as a feature value. By pairs of feature and feature value, we can represent most descriptions in feedback comments and also we can deal with feedback comments which are represented in slight different ways each other as the same one.

We examined grammatical features about the pairs of feature and feature value of the above 3522 descriptions. We found typical 14 types of pairs in word classes. Table 2 shows these pairs with number and frequency. In Table 2, pairs whose frequency is less than 1% are combined as "others". "-" in Table 2 shows the type of description which has no feature or feature value. Examples of description whose feature does not exist are "Very cute." and "I like it.". Although these sentences are about the item, the keyword for the item does not exist. Examples of description whose feature value does not exist are "Arrived." and "Received.". Although these sentences are about the item's arrival, the

Word class		Number	Frequency	
Feature	Feature value			
thanks	-	869	24.7	95.6
noun	adjectival verb	585	16.6	
verb	noun	505	14.3	
greetings	-	371	10.5	
noun	noun	209	5.9	
verb	adjectival verb	196	5.6	
noun	adjective	161	4.6	
noun	verb	124	3.5	
-	verb	91	2.6	
apologies	-	68	1.9	
-	adjectival verb	54	1.5	
-	adjective	53	1.5	
verb	adverb	42	1.2	
-	noun	41	1.2	
others		155	4.4	4.4
total		3522	100	100

Table 2: The number and frequency of pairs of feature and feature value in word classes.

keyword for the feature value like "today" and "just now" does not exist. This kind of omission is very popular especially in Japanese. The frequency of total of these typical 14 pairs is 95.6%. We can say that these grammatical patterns cover most descriptions. When we register words to be extracted as features to a dictionary, we can find a feature's position in a feedback comment by searching the registered words. We can extract its feature value by using the word classes of the above 14 types of pairs.

3.2 Frequencies of Contents

We counted the number of times that each pair of feature and feature value occurred in the above 1000 feedback comments. The column of 'Number' in Table 1 shows the results. The numbers of pairs such as "thanks", "greetings", "response-prompt" and "arrivals-item" are high. We see these pairs frequently in feedback comments. These descriptions can be seemed as descriptions for thanks, descriptions of courtesy or commonly-used descriptions, and they do not necessarily present the characteristic of the seller or the sold item. Consequently, we should delete descriptions with higher frequencies for making a useful summary.

3.3 Frequencies of Contents in Every Buyer

We selected 100 buyers randomly and picked up the newest 10 feedback comments in every buyer. Figure 1 shows a part of feedback comments written by Buyer 1. In these comments, there are some repeated descriptions such as "Thank you." and "Arrived today.". We counted the number of times that each pair of feature and feature value occurred in every buyer. We show the results of only ten users in Table 3 due to space limitation. We can see that the numbers of particular pairs (the colored cells in Table 3) are high. It seems that buyers write the same descriptions to many sellers. The number of buyers who wrote the same feature and feature value more than 5 times is 73 out of 100 buyers. Therefore, we guess that among many descriptions written by a certain buyer against sellers, descriptions written only against a certain seller may contain real opinion to the seller. Furthermore descriptions written against sellers other than a certain seller but not written against the seller also may contain real opinion to the seller. To extract these descriptions, it is better to compare feedback comments by every buyer than by every seller.

- Arrived today. Product works well. Thank you.
 - I'm very pleased. Thank you.
 - Arrived today. My son is pleased. Thank you.
 - Arrived today. Well packaged. Thank you.

Figure 1: Examples of feedback comment written by Buyer 1. (Translated from Japanese to English)

Feature	Feature value	Buyer									
		1	2	3	4	5	6	7	8	9	10
thanks		10	2	24	11	10	11	19	11	10	8
greetings		0	10	9	10	4	2	0	0	1	5
arrivals		7	2	2	0	10	0	0	0	3	4
	sefely	0	8	7	0	0	0	8	7	0	0
trans- actions	smooth	0	1	2	0	2	5	1	1	0	1
	comfortable	0	0	0	0	2	1	0	0	0	0
	rapid	0	0	0	0	1	1	0	0	0	0
response	prompt	0	0	10	0	1	0	7	2	1	3
	quickly	0	0	0	10	2	1	0	0	0	1
	courteous	0	0	8	0	1	0	1	0	0	3
packing	courteous	0	0	1	0	0	0	0	0	0	2
	beautiful	1	0	0	0	0	0	0	0	0	0
items	like	0	6	0	0	0	1	0	0	0	0
	nice	0	6	0	0	0	0	0	0	0	0
	okay	1	0	0	0	0	0	0	1	0	2
	good	0	1	0	0	0	0	0	0	0	2
	beautiful	1	0	0	0	0	0	0	0	1	1
	pleased	3	0	0	0	0	0	0	0	0	0
	cute	2	0	0	0	0	0	0	0	0	0
sellers	faithful	0	1	0	0	0	0	0	1	0	1
	request	0	1	0	1	0	0	0	0	0	0

Table 3: The frequency of pairs of feature and feature value in each buyer. (Feature values are translated from Japanese to English)

4. SOCIAL SUMMARIZATION METHOD

From the results of the investigation in Section 3, we propose a method for summarizing a target seller’s feedback comments. We call this method “social summarization method (SS method)”. This method summarizes a seller’s feedback comments by focusing on each buyer who is one of the authors of those comments. It calculates the frequencies of descriptions to sellers other than the target seller in the buyer’s feedback comments. It keeps descriptions, which are written only against the target seller, as a summary (after here “presence summary”). Since it seems that descriptions, which are written against sellers other than the target seller but not written against the target seller, are also useful for the users, this method keeps these descriptions as a summary (after here “non-presence summary”). To deal with a slight difference of descriptions, this method computes the frequencies in each pair of feature and feature value. This section explains the procedure of SS method.

4.1 Outline of Procedure of SS Method

The procedure of SS method is as follows (Figure 2 shows the procedure for making a summary of Seller A):

Step 1 Searching for feedback comments

SS method searches for all the feedback comments written by a buyer who wrote a feedback comment on the target seller by using a template of link and page structure for the auction site. This template is made by hand in advance.

Step 2 Finding differences

SS method finds and extracts descriptions that exist only in the feedback comment on the target seller and descrip-

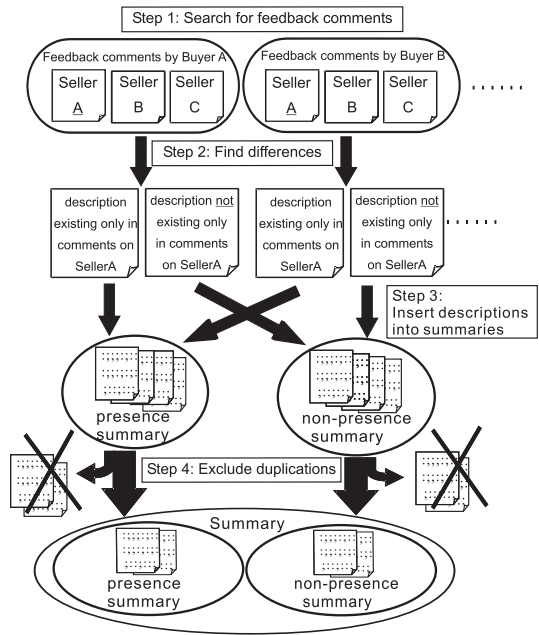


Figure 2: Procedure of SS method.

tions that do not exist only in the feedback comment on the target seller. Section 4.2 describes how to find these descriptions in detail.

Step 3 Inserting descriptions into summaries

SS method inserts the descriptions that are included only in the feedback comment on the target seller into a presence summary and inserts the descriptions that are not included only in the feedback comment on the target seller into a non-presence summary.

Step 4 Excluding duplication from summaries

SS method repeats Step 1 to Step 3 for every buyer and excludes duplicated descriptions from each summary. This means that SS method selects one description from several buyers’ descriptions whose feature and feature value (See Section 4.3) are same.

4.2 Method for Finding Differences

Here is a method for finding descriptions to be kept for the summary.

(1) SS method calculates the frequencies of all descriptions in the feedback comments on the sellers other than the target seller. Actually, descriptions are pairs of feature and feature value.

(2) SS method selects descriptions with higher frequencies (more than threshold α) and puts them into set S .

(3) SS method finds out the following descriptions:

- Descriptions in the feedback comment on the target seller which do not exist in set S .
- Descriptions in set S which do not exist in the feedback comment on the target seller.

4.3 Method for Extracting Feature and Feature Value

Here is a method for extracting feature and feature value. Since we found that there are only 13 types of features in the investigation in Section 3.1, we predefined synonyms to the names of 13 types of features by using a synonym dictionary

[12]. We selected from five to 56 synonyms for every feature. SS method performs keyword matching based on the predefined keywords and extracts a matched word as a feature. We also found in the investigation in Section 3.1 that there are 14 popular patterns of word class in the set of a feature and a feature value. SS method determines the word class of a feature value to a certain feature based on the above patterns. It extracts the word with the determined word class which exists in the nearest position from the feature as a feature value. Since we found that most of the pairs, whose features do not exist, are written about the items, SS method assigns such feature values to the feature of "item".

5. EVALUATION

In Section 5.1, we see the performance of our extracting method of feature and feature value. In Section 5.2, we verify how many descriptions of courtesy are eliminated and how many descriptions which are not descriptions of courtesy remain in the summary made by SS method. The characteristic of SS method is to investigate the frequency of description in every buyer. We verify the effectiveness of this characteristic in Section 5.3. For this purpose, we compare SS method with a method which investigates the frequency of description not in every buyer but in every seller (after here "comparative method"). In Section 5.4, we verify the effectiveness of the summaries obtained only by SS method.

5.1 Evaluation of Extracting Feature and Value

We examined the precision and recall of extracting pairs of feature and feature value from 100 feedback comments. Three evaluators, who were university students, analyzed the dependency in each sentence by hand and gave the correct answers by majority vote. The result is that the precision was 78.1% and the recall was 82.8%. Because we thought that they are high enough for examining the characteristics of SS method, we decided to carry forward the rest of the evaluations.

5.2 Evaluation of Sentences of Summaries

When we evaluate how many descriptions which each buyer wrote with real feeling remain in the summaries and how many descriptions of courtesy are eliminated, we have to ask every buyer whether or not the buyer wrote the description with real feeling. However, it is difficult to ask such a question to every buyer in real online auctions. Therefore, we decided to show some evaluators a buyer's feedback comment on a target seller and feedback comments on the sellers other than the target seller. We made evaluation data by asking evaluators to judge whether or not each description in the feedback comment on the target seller was a description of courtesy. Although evaluators cannot know what the buyer really feels, they can judge to some extent whether or not the description is written repeatedly and is meaningless by comparing both the two types of feedback comments.

To make the judgment easy, we asked evaluators to judge in sentence. We selected 10 sellers randomly as target sellers and used the latest 20 feedback comments in each seller. The evaluators were three university students. We decided whether or not each sentence is a sentence of courtesy by majority vote. In each buyer, we showed the evaluators the latest ten feedback comments on the sellers other than the target seller. There were 567 sentences in 200 feedback comments. Out of 567 sentences, 374 sentences were judged as

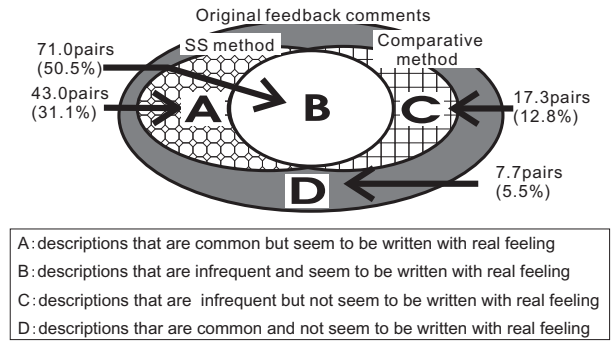


Figure 3: The difference between the result of SS method and that of the comparative method.

sentences of courtesy and 193 sentences were not judged as sentences of courtesy. In SS method, we set threshold α in Section 4.2 at 0.4 in the case of the presence summary. The result is that SS method eliminated 322 sentences and kept 245 sentences in the summary. In terms of eliminating sentences of courtesy, the recall was 80.8% and the precision was 93.8%. In terms of keeping sentences which are not sentences of courtesy in the summary, the recall was 89.5% and the precision was 70.8%. Consequently, we found that SS method can eliminate sentences of courtesy and keep sentences which are not sentences of courtesy in the summary on a certain level of recall and precision.

5.3 Comparison with Comparative Method

We selected six sellers randomly and summarized the feedback comments on each seller by SS method and by the comparative method. We set threshold α in Section 4.2, at 0.4 in the case of the presence summary and at 0.8 in the case of the non-presence summary. The result is that the average compression ratio of SS method was 28.5% and its standard deviation was 3.59. We found that all the descriptions in the original feedback comments can be categorized into four sets from the viewpoint whether or not they are included in the two types of presence summary, Set A, B, C and D as shown in Figure 3. SS method extracts Set A and Set B. The comparative method extracts Set B and Set C. Set B can be extracted by SS method and also extracted by the comparative method. Set D cannot be extracted either by SS method or by the comparative method. Figure 3 also shows the average number of descriptions and its ratio in each set. A summary of SS method includes descriptions which the buyers, who are the authors of the target seller's feedback comments, wrote only for the target seller. That means it includes descriptions that seem to be written with real feeling. A summary of the comparative method includes descriptions which are found in the target seller's feedback comments but are not found in the other target sellers' feedback comments. That means it includes descriptions that are generally infrequent. Therefore Set A, B, C and D have characteristics shown in the bottom of Figure 3.

Since the descriptions in Set B are infrequent and seem to be written with real feeling, we can say that they are the most remarkable descriptions when selecting a seller. The descriptions in Set C are infrequent but seem not to be written with real feeling. When the user sees these descriptions, the user may consider them as important descriptions.

These descriptions account for 12.8% of the original feedback comments. SS method can eliminate these descriptions. On the contrary, the descriptions in Set A are common but seem to be written with real feeling. These descriptions account for 31.1% of the original feedback comments. SS method can keep these descriptions in the presence summary. It becomes important whether Set A includes useful descriptions for selecting a seller.

5.4 Effectiveness of Summary Obtained only by SS method

When we see the descriptions obtained only by SS method (descriptions in Set A in Figure 3) one by one, they do not seem to be important since they are common expressions. Therefore we cannot directly evaluate the effectiveness of these descriptions. However, when we calculate the frequency of description, which is the frequency of each pair of feature and feature value before excluding duplicated descriptions in Step 4 in Section 4.1, we can estimate how many buyers wrote that description with real feeling. We calculate the ratio of each pair to all the pairs in the summary made by SS method. When we compare several sellers about a certain pair of feature and feature value, there is a seller who has high ratio and a seller who has low ratio. We also calculate the ratio of each pair to all the pairs in the original feedback comments. As our hypothesis, the variances of the ratios of the original feedback comments between sellers become small since many descriptions of courtesy exist in them. On the contrary, the variances of the ratios of SS method between sellers become large since descriptions of courtesy are eliminated and only descriptions that seem to be written with real feeling remain in the summaries. In Section 5.4.1, we compare the original feedback comments and the summaries made by SS method by using the average variance of the ratios of all pairs. However, seeing the ratios in pairs in the original feedback comments may be enough when there is no difference in the order of the sellers about the ratio between the original feedback comments and the summaries made by SS method. Section 5.4.2 verifies whether or not these orders are different.

5.4.1 Easiness of Comparing Sellers

We focused on the summaries obtained only by SS method (after here "summaries" in this subsection). We calculated the ratio of each pair of feature and feature value to all the pairs in the summary and that in the original feedback comments (after here "original comments" in this subsection). Next, we calculated the variance of the ratios among the sellers in the summaries and in the original comments in every pair. The simplest variance is calculated by the following formula.

$$\sigma^2 = \frac{1}{k-1} \sum_1^k (\hat{p}_i - \bar{p})^2 \quad (1)$$

Here, σ^2 represents the desired variance and k represents the number of sellers. \hat{p}_i represents the ratio of a pair of feature and feature value on Seller i . \bar{p} represents the average of \hat{p}_i . However, the variance of the ratios for the summary who has less descriptions tends to become large by this method. To calculate the correct variance, we used the following formula.

$$\sigma^2 = \frac{1}{k-1} \sum_1^k (\hat{p}_i - \bar{p})^2 - \frac{1}{k} \sum_1^k \frac{\hat{p}_i(1-\hat{p}_i)}{n_i} \quad (2)$$

Here, n_i represents the total number of pairs of feature and feature value on Seller i . Table 4 shows the ratios and the standard deviations of 20 pairs whose standard deviations in the summaries are large. Table 4(a) shows the ratios and the standard deviations of the original comments and Table 4(b) shows those of the summaries. The standard deviation in the summaries is larger than that in the original comments in every pair. We calculated the average standard deviation of all the pairs, which results in 0.54 in the original feedback comments and 1.28 in the summaries. Thus, our hypothesis has been proved.

When we see each row of Table 4(b), we can compare sellers in each pair of feature and feature value. For example, the ratio of Seller 6 is the largest in the pair of "response-prompt". We can guess the seller whose response is the most prompt among six sellers is Seller 6. When we see the row of Table 4(a), it becomes difficult to know the difference because the difference of values is not big. When we see the column of Table 4(b), we can see what are the seller's characteristics. For example, we can guess Seller 6's response is prompt and his transaction is reliable. We can also guess that the buyers will like the item bought from Seller 3 and gets happy that they can get the item. By seeing the table in rows or in columns, we can compare sellers or see one seller's characteristics.

5.4.2 Difference of Orders of Sellers

When comparing the ratios of the original comments with the ratios of the summaries by every pair in Table 4, the order of the sellers is different in all of the pairs. For example, in the pair "response-prompt", the ratio of Seller 3 is 7.2 and is second in the original comments. On the contrary, the ratio of Seller 3 is 3.4 and is sixth in the summaries. We calculated the Spearman's rank-order correlation coefficient for the relationship between the order of the sellers in the original comments and that in the summaries. The results became that the average for 20 pairs is 0.84. Thus, we can say that the order of the sellers in the feedback comments is different from the order in the summaries.

6. SUMMARY PRESENTATION METHOD

From the results of the evaluation in Section 5, we think that it is necessary to devise a presentation method of the summaries to maximize the effectiveness of SS method. The descriptions which are infrequent and seem to be written with real feeling (Set B in Figure 3) will be useful even if they are shown as they are extracted. On the contrary, the descriptions which are common and seem to be written with real feeling (Set A in Figure 3) will not be useful if they are shown as they are extracted. It is necessary to show the user the statistical values at the level of feature and feature value in a table or in a graph. Therefore, to maximize the effectiveness of SS method, it is necessary to categorize descriptions into the above two types by combining the comparative method that compares descriptions between sellers with SS method.

The order of display is also important when we have different granularity of information. In the area of information retrieval, an overview of target set of information can help users to find areas of potential interest before providing detailed information[1, 13]. When the user wants to find a good seller, we think that the following interaction is useful for him: (1) Firstly see an overview about the feature he

Feature	Feature value	Ratio(%)						Standard deviation
		Seller 1	Seller 2	Seller 3	Seller 4	Seller 5	Seller 6	
response	prompt	5.8	6.2	7.2	4	6	8.7	0.82
items	like	0.9	2.3	3.5	0.3	0.7	0	1.20
items	operable	0	0	0	2.3	0	2.2	1.06
arrivals	item	8.1	11	7.2	7.1	10.6	7.1	0.98
arrivals	safely	4.3	4.2	3.5	2.3	1	1.6	1.06
arrivals	today	3.2	1.4	3.5	0.8	3.7	2.2	0.84
items	okay	0	0	0	1.4	0	1.6	0.68
communication	late	1.2	0.8	0.3	0	0	0	0.39
arrivals	immediately	0	0.6	0.3	2.3	1	0.6	0.65
transactions	good	0.3	1.1	0.3	0.8	0	1.6	0.40
transactions	reliable	1.2	0.8	2.6	2.8	3	2.8	0.50
items	exist	0	0	0	0.8	1	0.9	0.33
response	rapid	0.3	0.3	0.3	1.4	0	0.6	0.31
items	glad	0.6	0.3	1.2	0	0	0	0.36
transactions	comfortable	0.6	0.8	0.3	0.3	0	0	0.09
items	problem	0	0	0	0.8	0.3	0.9	0.27
arrivals		0.6	0.6	1.2	0	0.3	0	0.28
items	use	0.9	0.3	0.6	0	1	0	0.22
response	kind	0	0.6	0	0.3	0	0.6	0.12
response	swift	0	0	0.6	0	0.3	0	0.13
Average								0.54

(a) ratios in original feedback comments

Feature	Feature value	Ratio(%)						Standard deviation
		Seller 1	Seller 2	Seller 3	Seller 4	Seller 5	Seller 6	
response	prompt	11.6	7.1	3.4	3.6	9.3	14.4	3.49
items	like	0	4	7.7	0.7	1	0	2.73
items	operable	0	0	0	5	0	5.9	2.58
arrivals	item	11.6	2	6	4.3	6.2	5.1	2.21
arrivals	safely	7	6.1	6.8	3.6	1	0.8	2.06
arrivals	today	7	3	3.4	1.4	8.2	3.4	1.64
items	okay	0	0	0	2.9	0	4.2	1.62
communication	late	4.7	3	0.9	0	0	0	1.56
arrivals	immediately	0	1	0	2.9	3.1	0	1.07
transactions	good	0	3	0.9	0.7	0	3.4	1.04
transactions	reliable	1.2	1	2.6	4.3	4.1	5.9	1.00
items	exist	0	0	0	2.2	1	2.5	0.75
response	rapid	0	0	0.9	2.2	0	1.7	0.56
items	glad	1.2	1	2.6	0	0	0	0.56
transactions	comfortable	1.2	3	0.9	0.7	0	0	0.54
items	problem	0	0	0	2.2	1	1.7	0.53
arrivals		2.3	2	1.7	0	0	0	0.48
items	use	1.2	1	1.7	0	3.1	0	0.45
response	kind	0	2	0	0.7	0	1.7	0.43
response	swift	0	0	1.7	0	1	0	0.36
Average								1.28

(b) ratios in summaries of SS method

Table 4: The ratio of pairs of feature and feature value.

Original feedback comments

Comment: I received it! Thank you for your careful and beautiful packing. I felt safe to do the transaction. I'm looking forward to buy items from you again.

Comment: Thank you very much! I could meet a wonderful seller and I spent a very good time. That was such a cute clothe that all family gave a shout of joy.

Comment: I received it! I really like it because it's so cute. I sent you postal stamps for the payment. I appreciate your quick response this time. I'm looking forward to buy items from you again.

Comment: I received it! I really like it. My kid is getting well. Thanks for your consideration.

Comment: The item has arrived today in safe. Thank you for your quick response. I'm looking forward to buy items from you again.

Comment: I received it! Sorry for the trouble. I'm satisfied with the long tops as I requested. I'm looking forward to buy items from you again.

Comment: I received it! Thank you for your quick and polite response. The item is so cute and I like it. I'm looking forward to buy items from you again.

Comment: I received it! Thank you very much.

Presence summary by SS method and comparative method

Comment: Thank you for your careful and beautiful packing.

Comment: I could meet a wonderful seller and I spent a very good time. That was such a cute clothe that all family gave a shout of joy.

Comment: I sent you postal stamps for the payment.

Comment: My kid is getting well. Thanks for your consideration.

Comment: (None)

Comment: I'm satisfied with the long tops as I requested.

Comment: (None)

Comment: (None)

Non-presence summary by SS method

Comment: The transaction is speedy.

Comment: Thank you for your smooth transaction.

Figure 5: Summary example produced by the combination of SS method and the comparative method (Translated from Japanese to English).

concerns and narrow down the sellers, (2) Second, see the more detailed feedback comments of only the selected sellers and decide from which seller to buy.

From this concept, we developed a summary presentation method shown in Figure 4. The user compares three sellers in this example. 14 buttons in the top of the window allow the user to select which feature to focus. When the user pushes one of these buttons, the system displays an overview of the descriptions which are common and seem to be written with real feeling. In detail, the system displays the ratios of the above descriptions to all descriptions in each seller in the form of graph. The user sees this graph and selects an interesting seller by the buttons below the graphs. The system displays the selected seller's descriptions which are infrequent and seem to be written with real feeling in the form of original description. The user can compare the details of the sellers by changing sellers by these buttons.

Finally, we show an example of the summary produced by the combination of SS method and the comparative method in Figure 5. We can see that only descriptions which are uncommon and seem to be written with real feeling are included in the presence summary. From the non-presence summary, we can guess that the seller's transaction is not so fast compared with other sellers. We think that the summary helps the user to do the final decision.

7. SUMMARY AND FUTURE WORKS

In this research, we aimed at summarizing feedback comments in online auctions which include many descriptions of courtesy. We proposed social summarization method (SS method) which investigates the feedback comments on a target seller in every buyer and produces a summary by extracting two types of descriptions. One is a description that appears only in the feedback comment on the target seller and the other is a description that appears in the feedback comments on the sellers other than the target seller but does not appear in the feedback comment on the target seller. We verified whether or not descriptions of courtesy are eliminated in the summary by SS method and whether or not the summary by SS method is different from the summary by the comparative method which investigates the feedback comments in each seller. Also, we verified the effectiveness of the summary obtained only by SS method when the user determines a seller. We found that descriptions of courtesy are eliminated and that SS method can extract the descriptions that the comparative method cannot extract. We also found that the summary obtained only by SS method is effective for selecting a seller. Finally, we proposed a summary presentation method which gives an overview produced from the summary obtained only by SS method in the form of graph for narrowing down the sellers, and then displays the actual descriptions on selected sellers for detail.

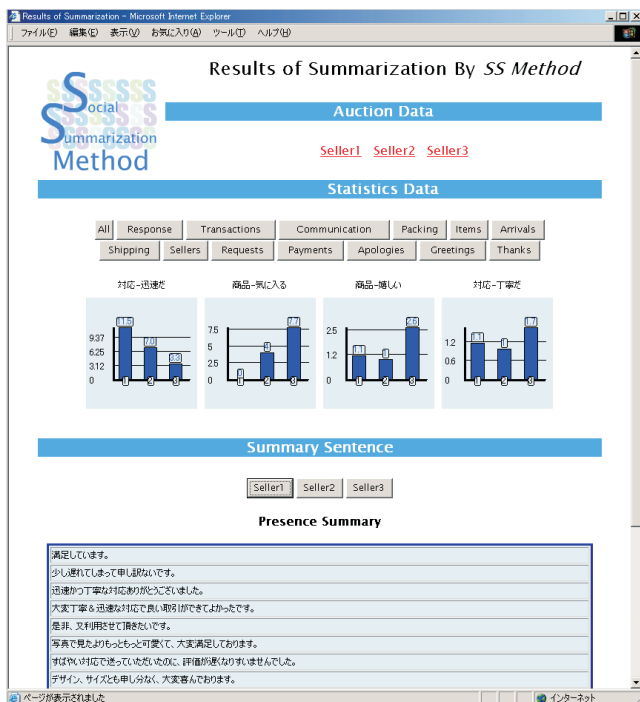


Figure 4: An example of summary presentation.

Although we focused on feedback comments in online auctions in this research, we think that another domain also has the same kinds of descriptions. One example is users' feedback comments in portal sites for searching hotels or restaurants. Working staffs in the hotel or the restaurant check these feedback comments. In some systems, they must reply to users' feedback comments. In such cases, users' feedback comments include more or less descriptions of courtesy. We think that SS method can be applied to another domain by changing the features, dictionaries and thresholds of the frequency. As a future work, we will apply SS method to another domain.

8. REFERENCES

- [1] R. Baeza-Yates and B. Ribeiro-Neto. *Modern Information Retrieval*. ACM Press, New York, 1999.
- [2] S. R. Das and M. Y. Chen. Yahoo! for amazon: Sentiment extraction from small talk on the web. In *Proceedings of 8th Asia Pacific Finance Association Annual Conference*, July 2001.
- [3] K. Dave, S. Lawrence, and D. M. Pennock. Mining the peanut gallery: opinion extraction and semantic classification of product reviews. In *Proceedings of 12th International World Wide Web Conference*, pages 519–528, May 2002.
- [4] S. Fujimura, M. Matsumura, N. Okazaki, and M. Ishizuka. Decision-making support based on reputations on bbs. In *Proceedings of 17th Annual Conference of Japanese Society for Artificial Intelligence*, pages 2B1–05, June 2003.
- [5] V. Hatzivassiloglou and K. R. McKeown. Predicting the semantic orientation of adjectives. In *Proceedings of 35th Annual Meeting of the Association for Computational Linguistics and 8th Conference of the European Chapter of the ACL*, pages 174–181, July 1997.
- [6] V. Hatzivassiloglou and J. M. Wiebe. Effects of adjective orientation and gradability on sentence subjectivity. In *Proceedings of 18th International Conference on Computational Linguistics*, August 2000.
- [7] N. Kobayashi, K. Inui, Y. Matsumoto, K. Tateishi, and T. Fukushima. Collecting evaluative expressions for opinion extraction. In *Proceedings of 1st International Joint Conference on Natural Language Processing*, pages 584–589, March 2004.
- [8] S. Morinaga, K. Yamanishi, K. Tateishi, and T. Fukushima. Mining product reputations on the web. In *Proceedings of 8th ACM SIGKDD International Conference on Knowledge Discover and Data Mining*, pages 341–349, July 2002.
- [9] T. Nasukawa and J. Yi. Sentiment analysis: Capturing favorability using natural language processing. In *Proceedings of 2nd International Conference on Knowledge Capture*, pages 70–77, October 2003.
- [10] C. D. Paice. Constructing literature abstracts by computer: Techniques and prospects. *Information Processing and Management*, 26(1):171–186, 1990.
- [11] B. Pang, L. Lee, and S. Vaithyanathan. Thumbs up? sentiment classification using machine learning techniques. In *Proceedings of Conference on Empirical Methods in Natural Language Processing*, pages 77–86, July 2002.
- [12] T. Shibata and S. Yamada. *Thesaurus*. Kodansha, Tokyo, 2002.
- [13] B. Shneiderman. The eyes have it: A task by data type taxonomy. In *Proceedings of IEEE Symposium on Visual Languages 96*, pages 336–343, September 1996.
- [14] E. Spertus. Smokey: Automatic recognition of hostile message. In *Proceedings of Innovative Applications of Artificial Intelligence*, pages 1058–1065, July 1997.
- [15] P. Subasic and A. Huettner. Affect analysis of text using fuzzy semantic typing. *IEEE Transaction On Fuzzy Systems Special Issue*, 9:483–496, August 2001.
- [16] K. Tateishi, Y. Ishiguro, and T. Fukushima. Texts processing for intelligent access of information on the web. *Journal of Japanese Society for Artificial Intelligence*, 19(3):317–322, May 2004.
- [17] P. Turney. Thumbs up or thumbs down? semantic orientation applied to unsupervised classification of reviews. In *Proceedings of 40th Annual Meeting of the Association for Computational Linguistics*, pages 417–424, July 2002.
- [18] J. Wiebe, T. Wilson, and M. Bell. Identifying collocations for recognizing opinions. In *Proceedings of ACL/EACL Workshop on Collocation*, July 2001.
- [19] K. Yamanishi and H. Li. Mining open answers in questionnaire data. *IEEE Intelligent Systems*, 17(5):58–63, September/October 2002.
- [20] J. Yi, T. Nasukawa, R. Bunescu, and W. Niblack. Sentiment analyzer: Extracting sentiments about a given topic using natural language processing techniques. In *Proceedings of 3rd IEEE International Conference on Data Mining*, pages 427–434, November 2003.