Classifying Hotel Reviews into Criteria for Review Summarization

Yoshimi SUZUKI
University of Yamanashi, 4-3-11 Takeda, Kofu, Yamanashi, JAPAN
ysuzuki@yamanashi.ac.jp

ABSTRACT
Recently, we can refer to user reviews in the shopping or hotel reservation sites. However, with the exponential growth of information of the Internet, it is becoming increasingly difficult for a user to read and understand all the materials from a large-scale reviews. In this paper, we propose a method for classifying hotel reviews written in Japanese into criteria, *e.g.*, location and facilities. Our system firstly extracts words which represent criteria from hotel reviews. The extracted words are classified into 12 criteria classes. Then, for each hotel, each sentence of the guest reviews is classified into criterion classes by using two different types of Naive Bayes classifiers. We performed experiments for estimating accuracy of classifying hotel review into 12 criteria. The results showed the effectiveness of our method and indicated that it can be used for review summarization by guest's criteria.

KEYWORDS: hotel reviews, text segmentation, guest's criteria.

Proceedings of the 2nd Workshop on Sentiment Analysis where AI meets Psychology (SAAIP 2012), pages 65–72,
COLING 2012, Mumbai, December 2012.

1 Introduction

Recently, we can refer to user reviews in the shopping or hotel reservation sites. Since the user's criteria are included in the user review compared with the information offering by a contractor, there is a possibility that many information which is not included in a contractor's explanation but included in the reviews. These customer/guest reviews often include various information about products/hotels which are different from commercial information provided by sellers/hotel owners, as customers/guests have pointed out with their own criteria, *e.g.*, service may be very important to one guest such as business traveler whereas another guest is more interested in good value for selecting a hotel for his/her vacation. Using Consumer Generated Media (CGM) such as hotel reviews, we can obtain different perspective from commercial information. However, there are at least six problems to deal with user reviews:

1. There are a large amount of reviews for each product/hotel.
2. Each review is short.
3. Each review includes overlapping contents.
4. Some reviews include wrong information.
5. The terms are not unified.
6. There are various sentiment expressions.

Moreover, there are many compound sentences in hotel reviews. Similarly, there are two or three criteria in a compound sentence. In order to deal with six problems mentioned in the above, we propose a method for classifying hotel reviews into criteria, such as service, location and facilities. We extracted criterion words and classified sentences of reviews into criteria. We can detect important sentences for review summarization by using the results of criteria extraction.

2 Related work

Our study is to extract list of reviewers' criteria and their sentiment expression. The approach is classified into sentiment analysis and text segmentation. Sentiment analysis is one of the challenging tasks of Natural Language Processing. It has been widely studied and many techniques (Beineke et al., 2004; Yi and Niblack, 2005; Hu and Liu, 2004), have been proposed. Wei et al. proposed HL-SOT (Hierarchical Learning process with a defined Sentiment Ontology Tree) approach (Wei and Gulla, 2010) to label a product's attributes and their associated sentiments in product reviews. Text segmentation has also been well studied. Utiyama and Isahara proposed a statistical method for domain-independent text segmentation (Utiyama and Isahara, 2001). Hirao et al. attempted the use of lexical cohesion and word importance (Hirao et al., 2000). They employed two different methods for text segmentation. One is based on lexical cohesion considering co-occurrences of words, and another is base on the changes of the importance of each sentence in a document.

3 System overview

Figure 1 illustrates an overview of our system. The system consists of two modules, namely "Classification of criterion words" and "Classification of review sentences into criteria". Hotel reviews written in Japanese are classified into criteria by the system.

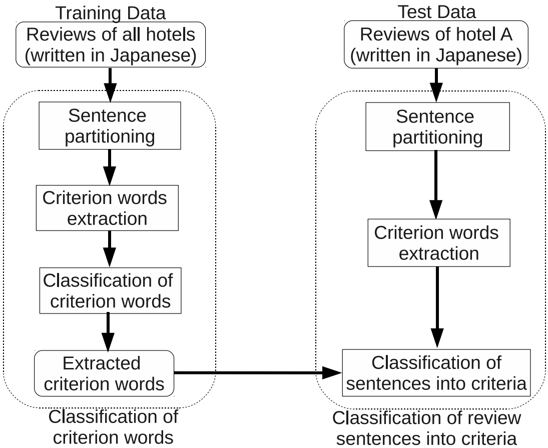

Figure 1: System overview.

4 Sentence partitioning

Compound sentences frequently appear in the reviews. Moreover, two or more criteria may be included within a compound sentence. For example, "The buffet-style breakfast is delicious, the room is also large and the scent of the shampoo and rinse in the bathroom are quite good": "(*chooshoku no baikingu mo oishiidesushi, heyamo hiroishi, ichiban kiniitteiruno ga heya ni oitearu shampuu to rinsu no kaori ga totemo iito omoimasu*)".

It is necessary to divide one sentence into some criteria. Fukushima proposed a method of sentence division for text summarization for TV news (Fukushima et al., 1999). They used rule based method for sentence partitioning. In this paper, each compound sentence was divided into some criteria by using compound sentence markers and "CaboCha" (Kudo and Matsumoto, 2002) which is a Japanese dependency structure Analyzer.

5 Criterion words extraction

Firstly, we defined criterion words as words that the reviewers notice in the reviews. Criterion words were frequently followed by postpositional particle: "*wa*" and adjective in the reviews written in Japanese. For extracting criterion words in reviews, we first extracted the pattern: "noun A + *wa* + adjective" from whole reviews. Next, we extracted "noun A", and finally, we collected words which are extracted as similar words of "noun A" by using the method mentioned in Section 6 and hypernym/hyponym of "noun A" in Japanese WordNet (Bond et al., 2009). Table 1 shows the adjectives which frequently appeared in the pattern: "noun A + *wa* + adjective".

Table 2 shows the extracted criterion words and their frequencies. These words in the table corresponds to criteria of the hotel.

Table 1: Adjectives which frequently appeared in "noun A + *wa* + adjective".

No	Adjective	Frequency	No	Adjective	Frequency
1	good (*yoi*)	142,719	6	delicious (*oishii*)	33,318
2	lack (*nai**)	73,186	7	inexpensive (*yasui*)	28,463
3	good (*yoi**)	67,643	8	delicious (*oishii**)	27,310
4	large(*hiroi*)	55,524	9	much (*ooi*)	23,122
5	near (*chikai*)	52,423	10	narrow (*semai*)	20,345

"*" indicates the word is written in hiragana.

Table 2: Candidate words of criteria (top 10).

No	Words	Frequency	No	Words	Frequency
1	room	56,888	6	service	11,270
2	breakfast	25,068	7	bath room	9,864
3	meal	17,107	8	noise	8,695
4	support	16,677	9	dish	8,252
5	location	14,866	10	hot spring	7,774

6 Similar word pair extraction

Reviews are written by many different people. People may express the same thing by using different expression. For example, "*heya*", "*oheya*" and "room" are the same sense, *i.e.*, room. Moreover, two words such as "*kyakushitsu*":(guest room) and "*heya*":(room) are often used in the same sense in the hotel review domain while those are different senses. Table 3 shows frequency of words which mean 'room' in a hotel review corpus.

Table 3: Extracted similar words of 'room'.

Words	Frequency
heya	171,796
oheya	38,547
room	17,203
kyakushitu	4,446

We thus collected similar words from hotel reviews by using Lin's method (Lin, 1998). Firstly, we extracted similar word pairs using dependency relationships. Dependency relationship between two words is used for extracting semantically similar word pairs. Lin proposed "dependency triple" (Lin, 1998). A dependency triple consists of two words: w, w' and the grammatical relationship between them: r in the input sentence. $||w, r, w'||$ denotes the frequency count of the dependency triple (w, r, w'). $||w, r, *||$ denotes the total occurrences of (w, r) relationships in the corpus, where "$*$" indicates a wild card.

We used three sets of Japanese case particles as r. Set A consists of two case particles: "*ga*" and "*wo*". They correspond to a subject and an object, respectively. Set B consists of six case particles. Set C consists of seventeen case particles. We selected word pairs which are extracted by using two or three sets.

For calculating similarity between w and w' with relation r, we used Formula (1).

$$I(w,r,w') = \log \frac{||w,r,w'|| \times ||*,r,*||}{||w,r,*|| \times ||*,r,w'||} \tag{1}$$

Let $T(w)$ be the set of pairs (r,w') such that $\log \frac{||w,r,w'|| \times ||*,r,*||}{||w,r,*|| \times ||*,r,w'||}$ is positive. The similarity $Sim(w_1,w_2)$ between two words: w_1 and w_2 are defined by Formula (2).

$$Sim(w_1,w_2) = \frac{\sum_{(r,w)\in T(w_1)\cap T(w_2)} (I(w_1,r,w) + I(w_2,r,w))}{\sum_{(r,w)\in T(w_1)} I(w_1,r,w) + \sum_{(r,w)\in T(w_2)} I(w_2,r,w)} \tag{2}$$

Table 4 shows the extracted similar word pairs.

Table 4: Results of extracting similar pairs using particle set A, B, C.

No.	Word1	Word2
1	favorable (*koukan*)	very favorable (*taihen koukan*)
2	route (*michizyun*)	route (*ikikata*)
3	stomach (*onaka*)	stomach (*onaka**)
4	dust (*hokori*)	dust (*hokori**)
5	net (net)	Internet (Internet)
6	renovation (*kaishu*)	renewal (renewal)
7	drain outlet (*haisuiguchi*)	drain (*haisuikou*)
8	word of mouth communication (*kuchikomi*)	word of mouth communication (*kuchikomi+*)
9	morning newspaper (*choukan*)	newspaper (*shinbun*)
10	a breakfast voucher (*choushokuken*)	ticket (ticket)

"*" indicates the word is written in hiragana.
"+" indicates the word is written in katakana.

In Table 4, there are some notational variants. In general, the pair of "morning newspaper" and "newspaper" and the pair of "breakfast voucher" and "ticket" are not the same meaning, while the two pairs are mostly the same sense in hotel reviews.

7 Classification of review sentences into criteria

We classified them into criteria by using lexical information of Japanese WordNet and similarity of words. We selected 12 criteria from the results shown in Table 2. Firstly, we classified each sentence into 12 criteria and miscellaneous as teaching data by hand. Next, we classified each sentence using two kind of Naive Bayes: multinomial Naive Bayes (MNB) and compliment Naive Bayes (CNB)(Rennie et al., 2003). Naive Bayes classifier is often used as a text classification because it is fast, easy to implement and relatively effective even if the training data is small. In the Naive Bayes classifier, we need a lot of training data per class. However, in this task, it is hard to collect many training data for some classes. We thus used CNB. CNB uses the compliment sets of each class for training, and it can be used more amount of data for each class. For expanding training data, we use sentences selected as same criterion by MNB and CNB. Table 5 shows classification results using MNB and CNB.

Table 5: Classification results using MNB and CNB.

Method	Precision	Recall	F-score
MNB	0.72	0.63	0.67
CNB	0.75	0.64	0.69
MNB&CNB	0.81	0.61	0.70

As we can see from Table 5 that when a sentence is classified into the same criterion by MNB and CNB, in most cases classified criterion is correct. Therefore, we used the sentences as additional training data.

Multinomial Naive Bayes classifier is obtained by using Formula (3).

$$MNB(d) = \arg\max_{c}\{\log \hat{p}(\theta_c) + \sum_{i} f_i \log \frac{N_{ci} + \alpha_i}{N_c + \alpha}\}, \tag{3}$$

where $\hat{p}(\theta_c)$ is the class prior estimate. j_i is the frequency count of word i in the reviews d. N_{ci} is number of times the word i appears in the training documents of class c. N_c is the number of words that appear in the training documents in class c. For α_i and α , we used 1 and the size of vocabulary, respectively. Similarly, CNB classifier is defined by Formula (4).

$$CNB(d) = \arg\max_{c}\{\log p(\vec{\theta}_c) - \sum_{i} f_i \log \frac{N_{\tilde{c}i} + \alpha_i}{N_{\tilde{c}} + \alpha}\}, \tag{4}$$

where $N_{\tilde{c}i}$ is the number of times word i occurred in documents in classes other than c and $N_{\tilde{c}}$ is the total number of word occurrences in classes other than c, and α_i and α are smoothing parameters. $\vec{\theta}_c = \{\theta_{c1}, \theta_{c2}, ..., \theta_{cn}\}$.

8 Experiments and discussion

For the experiment, we used hotel review of Rakuten Travel [1]. Table 6 shows Review data of the Rakuten Travel.

Table 7 shows 12 criteria which we used in the experiments.

We classified each sentence into these 12 criteria and a miscellaneous cluster.

We used Japanese WordNet Version 1.1 (Bond et al., 2009) as Japanese Thesaurus dictionary. We employed Lin's method (Lin, 1998) for extracting similar word pairs in hotel reviews.

We conducted experiments for dividing reviews into every criterion. We used reviews of 5 budget hotels. The average number of review per hotel was 51.2. Table 8 shows the results of text segmentation.

[1] url= http://travel.rakuten.co.jp/ We used Rakuten travel review data provided by Rakuten Institute of Technology

Table 6: Reviews of Rakuten Travel.

amount of data	250MB
# of reviews	350,000
# of hotel	15437
# of words for each review	375
# of reviews for each hotel	23

Table 7: 12 Criteria and their criterion words.

No	Criteria	Criterion words	No	Criteria	Criterion words
1	location	location, access	7	bath	bath room, bathtub
2	facilities	swimming pool, massage chair	8	amenity	razor, toothbrush
3	service	support, service	9	network	Wi-Fi, broad band
4	meal	breakfast, meal	10	beverage	beer, coke
5	room	room, noise	11	bed	bed, pillow
6	lobby	lobby, lounge	12	parking lot	parking lot, car

As can be seen clearly from the Table 8, the results obtained by CNB are better than those obtained by MNB.

Table 8: Results of Clustering.

Method	Precision	Recall	F-score
MNB	0.74	0.65	0.69
CNB	0.76	0.67	0.71

We used two kinds of Naive Bayes classifiers: multinomial Naive Bayes (MNB) classifier and compliment Naive Bayes (CNB) classifier in the experiments. The results obtained by CNB were better than those obtained by MNB. One reason why the results obtained by the CNB method were better than those obtained by the MNB is that the difference number of words in the training data used in these methods, and the balance of the data within each class. The number of words in the training data used in the MNB was smaller than that of the CNB. Because we used the data which consists of the limited number of words corresponding to each criterion class. Therefore the number of the training data for each criterion class is different from each other. In contrast, the training data we used in the CNB consist of the complement words in each class. Thus, the number of words in the training data becomes larger than that of the MNB, and the training data itself becomes a well-balanced data with each class.

Conclusion

In this paper, we proposed a method for extracting criteria and their sentiment expression from hotel reviews. The results showed the effectiveness of our method. Future work will include: (i) extracting criterion words with high accuracy, (ii) applying the method to a large number of guests reviews for quantitative evaluation, (iii) applying the method to other data such as grocery stores: LeShop[2], TaFeng[3] and movie data: MovieLens[4] to evaluate the robustness of

[2]www.beshop.ch

[3]aiia.iis.sinica.edu.tw/index.php?option=com_docman&task=cat_view&gid=34&Itemid=41

[4]http://www.grouplens.org/node/73

the method.

Acknowledgements

The authors would like to thank the referees for their comments on the earlier version of this paper. This work was partially supported by The Telecommunications Advancement Foundation.

References

Beineke, P., Hastie, T., and Vaithyanathan, S. (2004). The sentimental factor : Improving review classification via human-provided information. In *the 42nd Annual Meeting of the Association for Computational Linguistics*.

Bond, F., Isahara, H., Uchimoto, K., Kuribayashi, T., and Kanzaki, K. (2009). Enhancing the japanese wordnet. In *The 7th Workshop on Asian Language Resources, in conjunction with ACL-IJCNLP*.

Fukushima, T., Ehara, T., and Shirai, K. (1999). Partitioning long sentences for text summarization. *Journal of Natural Language Processing (in Japanese)*, 6(6):131–147.

Hirao, T., Kitauchi, A., and Kitani, T. (2000). Text segmentation based on lexical cohesion and word importance. *Information Processing Society of Japan*, 41(SIG3(TOD6)):24–36.

Hu, M. and Liu, B. (2004). Mining opinion features in customer reviews. In *Proceedings of Nineteenth National Conference on Artifical Intelligence*.

Kudo, T. and Matsumoto, Y. (2002). Japanese dependency analysis using cascaded chunking. In *CoNLL 2002:Proceedings of the 6th Conference on Natural Language Learning 2002*, pages 63–69.

Lin, D. (1998). Automatic retrieval and clustering of similar words. In *Proceedings of 36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics Proceedings of the Conference*, pages 768–774.

Rennie, J. D. M., Shih, L., Teevan, J., and Karger, D. R. (2003). Tackling the poor assumptions of naive bayes text classifiers. In *Twentieth International Conference on Machine Learning*, pages 616–623.

Utiyama, M. and Isahara, H. (2001). A statistical model for domain-independent text segmentation. In *Proceedings of the 39th Annual Meeting on Association for Computational Linguistics*, pages 499–506.

Wei, W. and Gulla, J. A. (2010). Sentiment learning on product reviews via sentiment ontology tree. In *Annual Meeting of the Association for Computational Linguistics*, pages 404–413.

Yi, J. and Niblack, W. (2005). Sentiment mining in webfountain. In *Proceedings of the 21st International Conference on Data Engineering*.