

Utilizing review analysis to suggest product advertisement improvements

Takaaki Tsunoda[†]*

Department of Computer Science
University of Tsukuba

tsunoda@mibel.cs.tsukuba.ac.jp[†]

Takashi Inui

Satoshi Sekine

Computer Science Department
New York University

Abstract

On an e-commerce site, product blurbs (short promotional statements) and user reviews give us a lot of information about products. While a blurb should be appealing to encourage more users to click on a product link, sometimes sellers may miss or misunderstand which aspects of the product are important to their users. We therefore propose a novel task: suggesting aspects of products for an advertisement improvement. As reviews have a lot of information about aspects from the perspective of users, review analysis enables us to suggest aspects that could attract more users. To achieve this, we break this task into the following two subtasks: aspect grouping and aspect group ranking. Aspect grouping enables us to treat product aspects at the semantic level rather than expression level. Aspect group ranking allows us to show users only aspects important for them. On the basis of experimental results using travel domain hotel data, we show that our proposed solution accomplishes NDCG@3 score of 0.739, which shows our solution is effective in achieving our goal.

1 Introduction

What are the most crucial parts of an e-commerce website that provide information to encourage users to buy products? For current websites, they are a product blurb and reviews. Blurbs, which are short promotional statements written by a seller and displayed as a short text advertisement, perform an important role in highlighting

^{*} Part of this research was conducted during the first author's internship at Rakuten Institute of Technology New York.

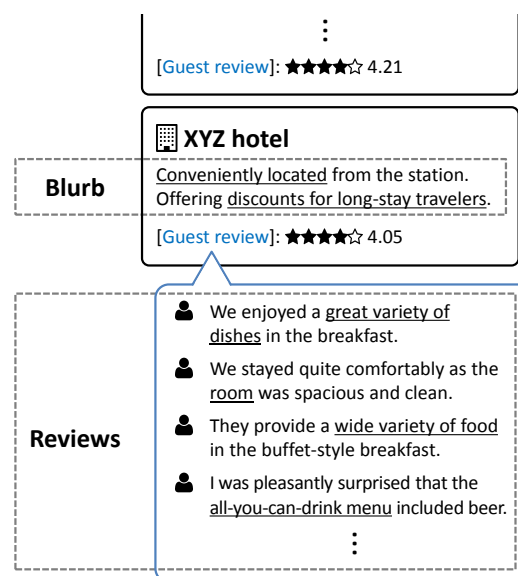


Figure 1: Examples of a blurb and reviews. Aspects of the hotel are underlined.

selling points to users. As it is the among the first things users see, a well written blurb is essential for encouraging users to click on the product link. Reviews, which are opinions or feedbacks on a product written by users who have purchased it, give us direct access to experiences of consumers who have used the product. Unlike blurbs, reviews from a number of users have abundant information about the product from the perspective of users. Figure 1 illustrates examples of a blurb and reviews of a hotel booking website.

Blurbs have to contain descriptions of the most important and appealing aspects of the product because the users will not check the reviews unless they are interested in the product by the blurb. However, due to a blurb writer's misunderstanding, the aspects of the product introduced in the blurb are not always the same as the product aspects that the users consider important or appealing. If these aspects are missing from the blurb, users who are looking for them never discover the

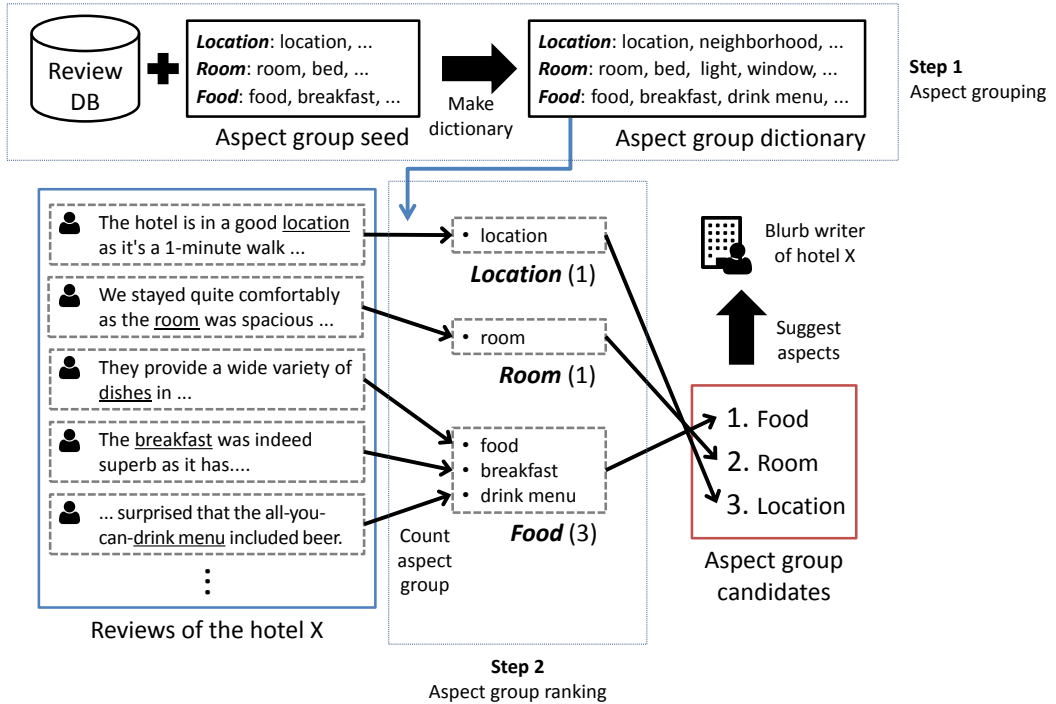


Figure 2: Task overview

existence of this product. For example in Figure 1, if the many users are looking for a hotel that provides great dishes or spacious rooms rather than discounts for long-stay travelers, they might not check the reviews, so the hotel may end up losing many potential customers. According to our observation in the hotel booking website, 81.0% of hotel blurbs lack one or more important aspects in the 3-best setting.

To suggest product blurb improvements, we propose the following novel task: finding aspects of a product that are important to the users and should be included in the blurb. For our initial approach to the task, we concentrate on the user review data. With a sufficient volume of reviews, we are able to determine which aspects of the product are important to users even if these aspects are not present in the blurb.

Figure 2 is an overview of the task. The goal of our task is to show aspect candidates that could be incorporated into the blurb ordered by their importance to users for a given product. To determine which aspects should be incorporated, we divide the task into two steps: aspect grouping and aspect group ranking. First, to treat aspects at the semantic level, we assign aspect expressions to aspect groups. Aspect grouping is essential to show meaningful suggestions, because enumerating aspects that have the same or similar concepts is re-

dundant. Second, to identify important aspects for users, we rank aspect groups on the basis of importance. Aspect group ranking is required to suggest only aspects that improve the blurb, as showing all aspects mentioned in the reviews regardless of their importance would not be helpful for blurb writers.

In this paper, we utilize the following well-known existing techniques for each step to confirm our proposed framework works well. For the aspect grouping, we employ one of the semi-supervised methods described by Zhai et al. (2010). Their technique allows us to make an aspect group dictionary that assigns each aspect expression to aspect group based on a semi-supervised technique with small manual annotation effort. For the aspect group ranking, we adopt an aspect ranking method proposed by Inui et al. (2013). Their ranking method, which is based on log-likelihood ratio (Dunning, 1993), enables us to leverage aspect group scores to extract the aspects that distinguish a product from its competitors.

Our contributions in this paper are as follows:

- We propose a novel task: finding characteristic aspects for blurb improvements.
- To achieve this goal, we break the task into the two subtasks: aspect grouping and aspect group ranking.

- To confirm our two-step framework, we adopt known and suitable methods in each step and investigated the best parameter combinations.

The paper is organized as follows. In the next section, we discuss related work mainly on aspect extraction and aspect ranking. In Section 3, we introduce the proposed method. In Section 4, we evaluate our method with a travel domain data. In Section 5, we conclude the paper and discuss the future directions.

2 Relevant work

Although the task we propose is new, there is a large body of work on sentiment analysis and aspect extraction that we can employ to build the components of our solution. In this section, we concentrate on research most directly relevant or applicable to our task.

First, identifying product aspects as opinion targets has been extensively studied since it is an essential component of opinion mining and sentiment analysis work (Hu and Liu, 2004; Popescu and Etzioni, 2005; Kobayashi et al., 2007; Qiu et al., 2011; Xu et al., 2013; Liu et al., 2014). This direction of research has been changing from merely enumerating aspects to capturing a more structured organization such as aspect meaning, a task that is also attempted as part of this work.

Existing research that focuses on structuring aspect groups is particularly relevant to our task. Although there exist fully unsupervised solutions based on topic modeling (Titov and McDonald, 2008a; Titov and McDonald, 2008b; Guo et al., 2009; Brody and Elhadad, 2010; Chen et al., 2014), the unsupervised approach still faces the challenge of generating coherent aspect groups that can be easily interpreted by humans. On the other hand, approaches using prior knowledge sources or a small amount of annotation data are also studied to maintain high precision while lowering the manual annotation cost (Carenini et al., 2005; Zhai et al., 2010; Chen et al., 2013a; Chen et al., 2013b; Chen et al., 2013c). Particularly, the method proposed by Zhai et al. (2010) can easily incorporate aspect expressions into predefined aspect groups and requires only a small amount of manually annotated data as aspect seeds. Their work utilizes an extension of Naive Bayes classification defined by Nigam et al. (2000), which allows for a semi-supervised approach to assigning

words to appropriate aspect groups. Their method serves as a component that enables us to treat aspects at the concept level instead of just the word level.

Lastly, another area applicable to our task is that of ranking aspects on the basis of various indicators, as exemplified by Zhang et al. (2010a), Zhang et al. (2010b), Yu et al. (2011), and Inui et al. (2013). While Zhang et al. (2010a), Zhang et al. (2010b), and Yu et al. (2011) propose aspect ranking methods based on aspect importance in a whole given domain, Inui et al. (2013) aim to find distinguishing aspect expressions of a product from other ones. They use a scoring method to rank aspect expressions and their variants, so theirs is the most appropriate technique for our task of discovering important aspects for users. To employ their approach for our task, we extend their method as described in the section 3.2.

3 Proposed method

To determine which aspects can be included in the blurb of the product, we utilize the following review analysis technique: aspect grouping and aspect group ranking. We begin by assigning aspect expressions to aspect groups to manage aspects at the semantic level. We next score aspect groups to suggest only important aspect groups.

3.1 Aspect grouping

The simplest way to suggest aspects for blurb improvement would be to extract important aspect expressions from product reviews and show them to users. However, according to this approach, phrases like "easy access", "easy to get to", or even "multiple transport options" might be suggested at the same time if the hotel's most characteristic aspect is its location. As they express very similar concept, suggesting all the aspect expressions that are important is not convenient.

To treat the aspects that expresses similar concept the same, we make a dictionary that assigns aspect expressions to higher level semantic groups. For example, in the case of a hotel, the aspect groups are *Location*, *Room*, *Food*, etc. When we treat aspects at the aspect group level, phrases containing words such as transport or access will all belong to a single *Location* group.

To build this dictionary at a minimum cost without much manual annotation effort, we employ one of the semi-supervised methods described by

Zhai et al. (2010). This dictionary allows us to assign each aspect expression in a review text to an appropriate aspect group.

The aspect grouping method of Zhai et al. (2010) we employ is based on a semi-supervised document clustering method proposed by Nigam et al. (2000). Although Nigam et al. (2000)’s method was proposed for document clustering, it can be applied to aspect grouping with Zhai et al. (2010)’s modifications.

The semi-supervised document clustering method of Nigam et al. (2000) is an extension of the Naive Bayes classifier. To use the Naive Bayes classifier in a semi-supervised approach, they applied the Expectation-Maximization (EM) algorithm (Dempster et al., 1977) to estimate labels for unlabeled documents. In this paper, we show only their calculation steps, not the complete derivation. First, learn the simple classifier using only labeled data (Equations (1) and (2)). Next, apply the classifier to unlabeled data to calculate the probabilities of clusters (Equation (3)). Then iterate the learning and application steps using both labeled and unlabeled data until the parameters converge (Equations (1) to (3)). In the iteration step, Equation (3) corresponds to M-step, and Equations (1) and (2) correspond to E-step. The concrete calculation steps are specified below, where w_i is a word, d_i is a document, and c_i is a cluster. $\{w_1, \dots, w_{|V|}\} = V$, $\{d_1, \dots, d_{|D|}\} = D$, and $\{c_1, \dots, c_{|C|}\} = C$ are a vocabulary, a document set, and clusters, respectively. $N_{w,d}$ is the frequency of word w in document d .

$$P(w_i|c_j) = \frac{1 + \sum_{d_i \in D} N_{w_i, d_i} P(c_j|d_i)}{|V| + \sum_{w_m \in V} \sum_{d_i \in D} N_{w_m, d_i} P(c_j|d_i)} \quad (1)$$

$$P(c_j) = \frac{1 + \sum_{d_i \in D} P(c_j|d_i)}{|C| + |D|} \quad (2)$$

$$P(c_j|d_i) = \frac{P(c_j) \prod_{w_k \in d_i} P(w_k|c_j)}{\sum_{c_r \in C} P(c_r) \prod_{w_k \in d_i} P(w_k|c_r)} \quad (3)$$

Zhai et al. (2010) applied to the semi-supervised aspect grouping in the following manner. They construct a bag-of-words, which is pseudo document for clustering, for each aspect expression using its context. The method is as follows: first, for a target aspect expression e , collect all occurrences of e from all reviews. Next, for all occurrences of e , pick words from a context window (t left words, t right words, and the e itself) except for stop-words¹. We used a window size of $t = 3$,

¹As we used Japanese reviews for our experiment, we re-

which is the same as that of Zhai et al. (2010). Finally, form bag-of-words d_e for e by summing picked words.

For example, if an aspect expression e is ”price” and we find that the following two sentences include this expression, a bag-of-words for e is $d_e = \langle \text{lowest, price, city, competitive, price, product} \rangle$.

- It was the lowest price in this city.
- A competitive price for this product.

3.2 Aspect group ranking

The next step is ranking aspect groups by their importance to display only those with a higher ranking.

To rank aspect groups, we regard aspect groups that are distinguishing as important ones, and base our approach on an aspect ranking method proposed by Inui et al. (2013). Their ranking method is based on log-likelihood ratio (LLR) (Dunning, 1993), which compares the probabilities of observing the entire data under the hypothesis that a given product and aspect are dependent and a hypothesis that they are independent. In this way, the LLR score takes into account the entire review data including other products’ reviews. As it has a higher value for aspects that differentiate a product from the others, it is a great fit for our goal of finding aspects that distinguishes a product from its competitors.

We extend the method proposed by Inui et al. (2013) because their goal differs from ours in two ways. First, as they are interested in ranking aspect expressions regardless of their polarity, expressions that appear many times in negative contexts might obtain high rankings. In contrast, in our task, such aspects are not appropriate for blurb improvements. Second, while they focus on ranking aspect expressions and their variants, we are interested in ranking aspect groups.

For the first point, we select sentences that have positive sentiment before performing the subsequent procedures. More specifically, we make a binary classifier, which classifies a given review sentence into either positive or not positive sentiment. We use the classifier to extract only positive sentiment sentences.

For the second point, we use frequencies of aspect expressions that belong to an aspect group instead of frequencies of a word and its variants to moved particles and auxiliary verbs as stop-words.

calculate LLR. With this approach, the concrete calculation steps of LLR for a product p and aspect group g are as follows. First, we calculate the following four parameters: a, b, c, d .

$$\begin{aligned}
a &= \text{Frequency of words in } g \text{ in } S_p \\
b &= \text{Frequency of words in } g \text{ in } S \\
c &= \text{Frequency of words in all aspect groups} \\
&\quad \text{except } g \text{ in } S_p \\
d &= \text{Frequency of words in all aspect groups} \\
&\quad \text{except } g \text{ in } S
\end{aligned} \tag{4}$$

where $S_p, S = \bigcup_p S_p$ are a set of positive review sentences in a product p 's reviews and in all products' reviews, respectively. Then, when we let $n = a + b + c + d$, the LLR value can be obtained in the following manner.

$$\begin{aligned}
\text{LLR}_0 &= a \log \frac{an}{(a+b)(a+c)} + b \log \frac{bn}{(a+b)(b+d)} \\
&\quad + c \log \frac{cn}{(c+d)(a+c)} + d \log \frac{dn}{(c+d)(b+d)} \tag{5}
\end{aligned}$$

Finally, we correct the LLR_0 value as the LLR_0 cannot distinguish between "an aspect group g is characteristic in p " and "an aspect group g is more characteristic in other products than in p ". We want to obtain the former one, so we employ the following correction:

$$\text{LLR} = \begin{cases} +\text{LLR}_0 & \text{if } \frac{a}{c} > \frac{b}{d} \\ -\text{LLR}_0 & \text{otherwise} \end{cases} \tag{6}$$

A higher LLR value means the aspect group g is more characteristic for product p .

In addition to this, we also tried using sentence or review frequencies instead of word frequencies to calculate LLR. For example, in the sentence and review levels, parameter a is calculated as follows.

$$\begin{aligned}
a_{S\text{-LLR}} &= \text{Frequency of } \underline{\text{sentences}} \\
&\quad \text{that have a word in } g \text{ in } S_p \\
a_{R\text{-LLR}} &= \text{Frequency of } \underline{\text{reviews}} \\
&\quad \text{that have a word in } g \text{ in } S_p
\end{aligned}$$

We can calculate b, c, d for the sentence or review level similarly to the above. We did this to attempt to avoid introducing bias from reviews that elaborate on a certain aspect and influence its frequency for a given product. We expect preventing over-estimation the occurrence of an aspect group from this approach if it is driven by a high frequency within a review from a single user.

4 Experiments

4.1 Experiment conditions

To conduct our experiments, we used hotel blurb and review data from a Japanese website, Rakuten Travel. We chose to focus on this domain as hotels are characterized by numerous aspects, thus presenting a fair challenge for our task. The aggregate review data comes from the publicly available *Rakuten Data Release*². We selected hotels that had between 10 and 1000 user reviews, rendering a total of 13,664 hotels and 2,254,307 user reviews. For data preprocessing, we employed MeCab 0.996 (Kudo et al., 2004) as a word tokenizer and applied a simple rule-based system for sentence segmentation.

To build an aspect dictionary for the travel domain, we predefined the following 12 aspect groups: *Service, Location, View and Landscape, Building, Room, Room facilities, Hotel facilities, Amenities, Bath, Food, Price, and Website*. We selected frequent nouns and noun phrases that occurred in at least 1% of all reviews as aspect expressions. For noun phrase, we considered two following types: 1) complex nouns and 2) "A of B", where both of A and B are nouns. After filtering out proper nouns, we obtained 9,844 aspect phrases. For seeds for the semi-supervised learning, we manually labeled the 281 most frequent ones, which are around 3% of the candidates.

To evaluate the dictionary, we examined the performance of sentence labeling. We compared the golden standard and the automated labeling based on the dictionary, which is obtained by regarding aspect groups in which aspect phrases in the dictionary appeared in a sentence as belonging to the sentence's aspect group. Note that we allowed multiple aspect groups in a sentence. For the golden standard dataset, we annotated random sampled 100 reviews, which consist of 450 sentences. We used precision and recall as an evaluation metrics.

To select positive sentiment sentences, we employed a SVM classifier. For training data, we used the TSUKUBA corpus 1.0, which is a sentence level sentiment-tagged corpus included in the Rakuten Data Release. We constructed a bi-

²Rakuten Data Release: <http://rit.rakuten.co.jp/opendata.html>. In this paper we use a snapshot of the data released in 2012. Hotel blurbs correspond to the information shown on hotel detail pages.

Table 1: Performance of aspect grouping

Aspect group	Precision	Recall
Service	71.67%	70.49%
Location	64.86%	70.59%
View and Landscape	55.56%	38.46%
Building	85.71%	30.77%
Room	78.18%	64.18%
Room Facilities	52.17%	36.36%
Hotel Facilities	28.57%	8.70%
Amenities	83.33%	31.25%
Bath	88.64%	88.64%
Food	71.62%	79.10%
Price	76.92%	44.44%
Webpage	36.36%	12.90%

nary classifier, which classifies a given review sentence into either positive or non-positive sentiment, from 4,309 review sentences by using scikit-learn 0.15.2 (Pedregosa et al., 2011). By 5-fold cross validation, we confirmed that the classifier achieved 83.08% precision and 79.79% recall for finding positive sentences.

The ranking method we use is based on log-likelihood ratio scoring (**LLR**) as described above. To compare the effectiveness of LLR for aspect group ranking, we compared this method with two baseline methods: aspect group frequency (**TF**), which is the same as parameter a in Equation (4), and T-scores of the relative aspect group frequency in the reviews of each hotel (**TScore**). In addition to this, we examined which level performs better for measuring frequency: word (**W-LLR**), sentence (**S-LLR**), or review (**R-LLR**). Likewise, we compared the effect of the frequency unit on TF and T-scores so that we can also compare between the word (**W-TF**, **W-TScore**), sentence (**S-TF**, **S-TScore**), and review level (**R-TF**, **R-TScore**).

To evaluate the aspect grouping methods, we annotated randomly selected 126 hotels for a gold standard dataset. For each hotel, we ranked appropriate aspect groups for a blurb. To judge rank and appropriateness, we referred the following sources: a current blurb, a introduction page of the hotel, and most recent 50 reviews. According to our annotation, the average number of aspect groups that are appropriate for a blurb is 3.09.

We use the Normalized Discounted Cumulative Gain (NDCG) (Järvelin and Kekäläinen, 2002) as evaluation measures. NDCG measures the performance of a ranking system based on the similarity between the system output and the gold standard,

Table 2: Performance of aspect group ranking

Method	NDCG				
	@1	@2	@3	@4	@5
R-LLR	0.743	0.736	0.737	0.752	0.770
S-LLR	0.751	0.741	0.739	0.745	0.764
W-LLR	0.777	0.735	0.736	0.737	0.760
R-TScore	0.567	0.597	0.639	0.663	0.693
S-TScore	0.599	0.607	0.647	0.673	0.701
W-TScore	0.589	0.614	0.643	0.657	0.690
R-TF	0.674	0.663	0.676	0.707	0.753
S-TF	0.695	0.678	0.682	0.718	0.760
W-TF	0.648	0.649	0.654	0.705	0.751

defined as:

$$\text{NDCG}@n = \frac{\text{DCG}@n}{\text{IDCG}@n} \quad (7)$$

$$\text{DCG}@n = rel_1 + \sum_{i=2}^n \frac{rel_i}{\log_2 i} \quad (8)$$

where n is the rank position to measure, rel_i is the relevance grade for a i th suggested aspect group, $\text{IDCG}@n$ is the normalizer to make $\text{NDCG}@n$ varying from 0.0 to 1.0. For n , we show $n = 1, \dots, 5$ (1 to 5-best output) results because the average number of appropriate aspect groups is around 3 according to our annotation, and suggesting a large number of aspects would go against the goal of the task. For rel_i , we used the logarithmic discount $rel_i = 1/\log(1+r)$ where r is the rank of the i th aspect group in the gold standard, which is a feasible discount function for NDCG (Wang et al., 2013).

4.2 Results

4.2.1 Evaluation of aspect grouping

First, we present the performance of aspect grouping. Table 1 shows the dictionary performance for each aspect group.

The result shows the aspect grouping component has reasonable performance except for low recall aspect groups including *Hotel facilities* and *Webpage*. The reason for this is because these aspect groups appear less frequently and difficult to assign aspect expressions to these aspect groups accurately. According to our observation, in the Rakuten Travel, reviewers do not mention about these aspect groups unless they find something special, as these aspect groups are not fundamental aspects of a hotel as opposed to *Room* or *Food*. We examine how this result affects the aspect group ranking in the evaluation of aspect group ranking.

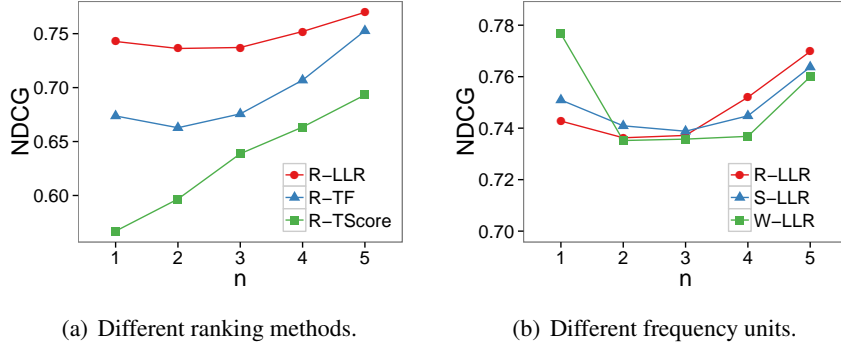


Figure 3: Performance comparison between ranking methods and frequency units

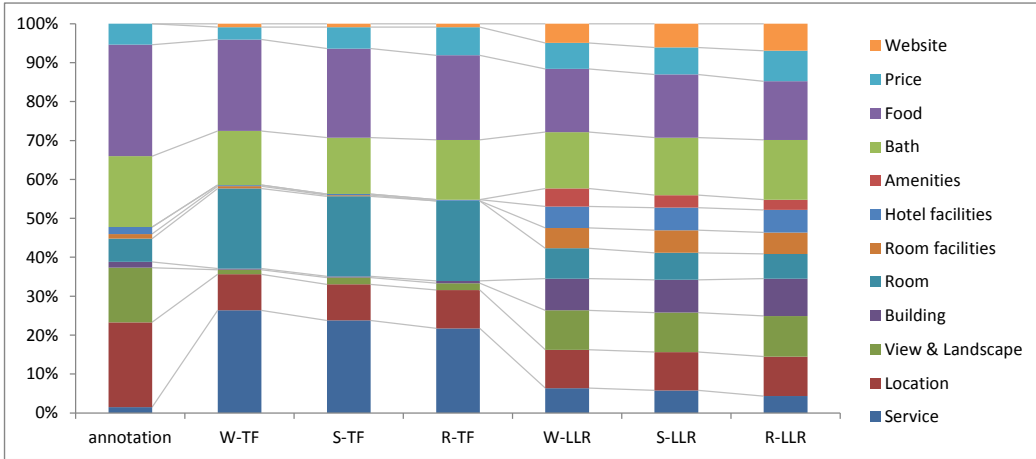


Figure 4: Aspect group distribution at 3-best.

4.2.2 Evaluation of aspect group ranking

Next, we compare the performance between ranking methods (**R-LLR**, **R-TScore**, **R-TF**). Figure 3(a) and top rows of each block of Table 2 show the results obtained by different methods when the aspect group frequency unit is fixed as a single review. According to these results, the log-likelihood ratio score (**R-LLR**) shows a higher NDCG score than the other methods at NDCG@1 to NGCG@5. This establishes that LLR is the most effective method for our task.

Lastly, we compare the performance of the LLR ranking depending on the frequency count unit (**R-LLR**, **S-LLR**, **W-LLR**), as illustrated in Figure 3(b) and the first block of Table 2. The results show that sentence-based **S-LLR** and review-based **R-LLR** have almost the same NDCG score compared to word-based **W-LLR**. Furthermore, we can observe from Table 2 that the same tendency exists for **TF** and **TScore** baselines. This shows that for the purpose of our task, the frequency unit is a less important parameter com-

pared to the ranking method.

For a detailed investigation, we calculated aspect group distribution for the gold standard and outputs for each method at 3-best. Figure 4 illustrates the aspect group distribution. Distribution of TFs, which correspond to how many aspect phrases appeared in reviews, is not very similar to that of the gold standard, especially for aspect groups *Service*, *Room*, *View and Landscape* and *Hotel facilities*. For *Service* and *Room*, we think this is also brought on by the tendency of reviews, that is, many reviewers mentioned them even if reviewers did not find anything special about them as these are fundamental aspects. In contrast to, *View and Landscape* and *Hotel facilities* are not fundamental aspects, so reviewers mention about them only if they find something special. In addition to this, the aspect group dictionary not captures *Hotel facilities* occurrences well as the Recall column of Table 1 shows. On the other hand, distributions of LLR and the gold standard are more similar. The reason for this is the LLR can leverage scores by comparing other

Table 3: Output example. Aspect group column shows system outputs (**S-LLR**) for reviews.

Hotel		Aspect group	Text (for suggestion, an example review sentence)
#1	Blurb	Location	Located in the port town.
	Suggestion	Bath	We enjoyed many kinds of baths including an open air bath, a cascading bath, and a sleeping bath.
		Food	I'm very satisfied with the cuisine that they prepare from the ocean and mountain fare.
		Room	The room was impressively clean like just after refurbishment.
Annotation	#1:Bath, #2:Food, #3:Room		
#2	Blurb	Building	Reopened on Sep. 10th after complete refurbishment!!
	Suggestion	Bath	There were many private hot-spring bath facilities for families, and the hot-spring water was great!
		Food	I was pleasantly surprised that the all-you-can-drink menu included beer and coffee, and the meals tasted great.
		Amenities	I was impressed with the unlimited towel policy in the hot spring, and skin lotion and other beauty products were provided.
Annotation	#1:Food, #2:Bath, #3:View and Landscape, #4:Amenities		

hotels' reviews. More specifically, aspect groups mentioned in many hotel reviews like *Service* or *Room* have low scores, and those mentioned in few reviews like *View* or *Hotel facilities* have high scores. Besides, even the dictionary misses mentions of aspect groups like *Hotel facilities*, LLR can mitigate this problem. Meanwhile, the aspect group *Location* is underestimated and *Website* is overestimated. To deal with this problem, employing prior knowledge about which aspects are preferred for blurbs in a given domain might give us better aspect group suggestions.

For the best performing **S-LLR** method at 3-best, NDCG@3 score is 0.739, which allows us to make reasonable suggestions for enhancing some product blurbs. Table 3 shows output examples of our system (**S-LLR**). In the first example, the blurb describes one aspect group: *Location*. This blurb might lose customers who prioritize other aspects of the hotels. To improve this blurb, our system suggests other aspect groups that could be mentioned in the blurb on the basis of reviews of the hotel: *Bath*, *Food*, and *Room* at 3-best. In view of the annotation and the example of review sentences, we can observe that these aspect groups are characteristic and including them in the blurb would improve it. Meanwhile, in the second example, the blurb mentions about *Building*. The suggestions of our system, *Bath*, *Food* and *Amenities* might help blurb improvement as example review sentences show. However, according to the annotation, the aspect group *Amenities* is the fourth while the *View and Landscape* is the third. We think it is from two causes. First, the system suggests without consideration of the aspect group preference in blurbs and results over-

estimation of *Amenities*. Second, the dictionary captures insufficient reviews which mention about *View and Landscape* as Recall of the aspect grouping is lower as Table 1 shows. We think refining aspect grouping and improving aspect group ranking are both effective to achieve better performance.

5 Conclusions

In this paper, we proposed a novel task of suggesting product blurb improvement and offered a solution to extract important aspect groups from reviews. To achieve our goal, we divided the task into two subtasks, which are aspect grouping and aspect group ranking.

The future directions of our work are as follows. First, instead of using whole given domain reviews to calculate LLR, we could use only real competitors for a target product. For example, in the travel domain, we could use reviews of hotels near to a target hotel to enable the system to suggest the target hotels more unique aspects compared with its competitors. Next, in Table 3, we also showed representative review sentences that illustrate each aspect group. If we could show such sentences along with suggested aspect groups, the system would make writing blurbs much easier. Like in the case of aspect group selection, we could use a scoring method such as LLR to select characteristic sentences.

Acknowledgements

We thank Zofia Stankiewicz for supporting this work as a mentor of the internship program. We would also like to thank to Rakuten, Inc. for releasing the resources of Rakuten Travel.

References

- Samuel Brody and Noemie Elhadad. 2010. An unsupervised aspect-sentiment model for online reviews. In *NAACL-HLT-2010*, pages 804–812, June.
- Giuseppe Carenini, Raymond T. Ng, and Ed Zwart. 2005. Extracting knowledge from evaluative text. In *K-CAP-2005*, pages 11–18, October.
- Zhiyuan Chen, Arjun Mukherjee, Bing Liu, Meichun Hsu, Malu Castellanos, and Riddhiman Ghosh. 2013a. Discovering coherent topics using general knowledge. In *CIKM-2013*, pages 209–218, October.
- Zhiyuan Chen, Arjun Mukherjee, Bing Liu, Meichun Hsu, Malu Castellanos, and Riddhiman Ghosh. 2013b. Exploiting Domain Knowledge in Aspect Extraction. In *EMNLP-2013*, pages 1655–1667, October.
- Zhiyuan Chen, Arjun Mukherjee, Bing Liu, Meichun Hsu, Malu Castellanos, and Riddhiman Ghosh. 2013c. Leveraging multi-domain prior knowledge in topic models. In *IJCAI-2013*, pages 2071–2077, August.
- Zhiyuan Chen, Arjun Mukherjee, and Bing Liu. 2014. Aspect Extraction with Automated Prior Knowledge Learning. In *ACL-2014*, pages 347–358, June.
- A. P. Dempster, N. M. Laird, and D. B. Rubin. 1977. Maximum likelihood from incomplete data via the em algorithm. *JOURNAL OF THE ROYAL STATISTICAL SOCIETY*, B 39(1):1–38.
- Ted Dunning. 1993. Accurate methods for the statistics of surprise and coincidence. *Computational Linguistics*, 19(1):61–74.
- Honglei Guo, Huijia Zhu, Zhili Guo, XiaoXun Zhang, and Zhong Su. 2009. Product feature categorization with multilevel latent semantic association. In *CIKM-2009*, page 1087, November.
- Minqing Hu and Bing Liu. 2004. Mining opinion features in customer reviews. In *AAAI-2004*, pages 755–760, July.
- Takashi Inui, Yuto Itaya, Mikio Yamamoto, Keiji Shinzato, Yu Hirate, and Kaoru Yamada. 2013. Structuring opinion features by relative characteristics on user-opinion aggregation. *Journal of Natural Language Processing*, 20(1):3–25 (in Japanese).
- Kalervo Järvelin and Jaana Kekäläinen. 2002. Cumulated gain-based evaluation of IR techniques. *ACM Transactions on Information Systems (TOIS)*, 20(4):422–446.
- Nozomi Kobayashi, Kentaro Inui, and Yuji Matsumoto. 2007. Extracting aspect-evaluation and aspect-of relations in opinion mining. In *EMNLP-CoNLL-2007*, pages 1065–1074, June.
- Taku Kudo, Kaoru Yamamoto, and Yuji Matsumoto. 2004. Applying conditional random fields to Japanese morphological analysis. In *EMNLP-2004*, pages 230–237, July.
- Kang Liu, Liheng Xu, and Jun Zhao. 2014. Extracting Opinion Targets and Opinion Words from Online Reviews with Graph Co-ranking. In *ACL-2014*, pages 314–324, June.
- Kamal Nigam, Andrew Kachites McCallum, Sebastian Thrun, and Tom Mitchell. 2000. Text classification from labeled and unlabeled documents using em. *Machine Learning*, 39(2-3):103–134.
- F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. 2011. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830.
- Ana-Maria Popescu and Oren Etzioni. 2005. Extracting product features and opinions from reviews. In *EMNLP-2005*, pages 339–346, October.
- Guang Qiu, Bing Liu, Jiajun Bu, and Chun Chen. 2011. Opinion Word Expansion and Target Extraction through Double Propagation. *Computational Linguistics*, 37(1):9–27.
- Ivan Titov and Ryan McDonald. 2008a. A Joint Model of Text and Aspect Ratings for Sentiment Summarization. In *ACL-HLT-2008*, pages 308–316, June.
- Ivan Titov and Ryan McDonald. 2008b. Modeling online reviews with multi-grain topic models. In *WWW-2008*, page 111, April.
- Yining Wang, Liwei Wang, Yuanzhi Li, Di He, and Tie-Yan Liu. 2013. A theoretical analysis of NDCG type ranking measures. In *COLT-2013*, pages 25–54, June.
- Liheng Xu, Kang Liu, Siwei Lai, Yubo Chen, and Jun Zhao. 2013. Mining opinion words and opinion targets in a two-stage framework. In *ACL-2013*, pages 1764–1773, August.
- Jianxing Yu, Zheng-Jun Zha, Meng Wang, and Tat-Seng Chua. 2011. Aspect ranking: Identifying important product aspects from online consumer reviews. In *ACL-HLT-2011*, pages 1496–1505, June.
- Zhongwu Zhai, Bing Liu, Hua Xu, and Peifa Jia. 2010. Grouping product features using semi-supervised learning with soft-constraints. In *COLING-2010*, pages 1272–1280, August.
- Kunpeng Zhang, Ramanathan Narayanan, and Alok Choudhary. 2010a. Voice of the customers: mining online customer reviews for product feature-based ranking. In *WOSN-2010*, June.

Lei Zhang, Bing Liu, Suk Hwan Lim, and Eamonn O'Brien-Strain. 2010b. Extracting and ranking product features in opinion documents. In *COLING-2010*, pages 1462–1470, August.