Analysis of Travel Review Data from Reader's Point of View

Maya Ando   **Shun Ishizaki**

Graduate School of Media and Governance

Keio University

5322 Endo, Fujisawa-shi, Kanagawa 252-0882, Japan

maya@sfc.keio.ac.jp   ishizaki@sfc.keio.ac.jp

Abstract

In the NLP field, there have been a lot of works which focus on the reviewer's point of view conducted on sentiment analyses, which ranges from trying to estimate the reviewer's score. However the reviews are used by the readers. The reviews that give a big influence to the readers should have the highest value, rather than the reviews to which was assigned the highest score by the writer. In this paper, we conducted the analyses using the reader's point of view. We asked 20 subjects to read 500 sentences in the reviews of Rakuten travel and extracted the sentences that gave a big influence to the subjects. We analyze the influential sentences from the following two points of view, 1) targets and evaluations and 2) personal tastes. We found that "room", "service", "meal" and "scenery" are important targets which are items included in the reviews, and that "features" and "human senses" are important evaluations which express sentiment or explain targets. Also we showed personal tastes appeared on "meal" and "service".

1   Introduction

Reviews are indispensable in the current e-commerce business. In the NLP field, there have been a lot of works conducted on sentiment analyses, which ranges from trying to estimate the reviewer's score or analyzing them by the aspects of reviewer's evaluations. However the reviews are used by the customers, not by the reviewers.

So, the business value of the review lies on the customer's point of view, rather than the reviewer's point of view. The reviews which give a great influence to the customers should have the highest value, rather than the reviews to which were assigned the highest score by the writer. We defined customers as readers and reviewers as writers. We found the differences between the writer's view and the reader's one using scores given by reviewers. Especially the negative information is found much more influential to the readers than the positive one (Ando et al., 2012).

We conducted the analyses using the *reader's point of view*. We asked 20 subjects to read 500 review sentences in Rakuten travel reviews[1] and extract the sentences from them that gave a great influence. We analyzed the influential sentences from the following two points of view, 1) *targets* and *evaluations* (Chap. 4) and 2) Personal tastes (Chap. 5).

2   Previous Study

There have been a lot of works on sentiment analysis in the past decade. Some of them were classifying reviews into positive, negative, or neutral (Turney, 2002; Pang et al., 2002; Koppel et al., 2006; Pang, 2005; Okanohara et al., 2006; Thelwall et al., 2010). These works were conducted based on the writer's point of view, i.e. the targets are mainly assigned by the writers. In our research, we will describe reader's point of view.

---

[1] Rakuten Travel Inc.
http://travel.rakuten.co.jp/ (Japanese)

In some reviews, there is information called helpfulness which is given by readers. Ghose et al. (2007) used it as one of the features in order to rank the reviews. Passos (2010) also used it to identify authoritativeness of reviews. They didn't conduct any detailed analysis like what we conducted in this paper. So far, the usage of the helpfulness information is limited, and indeed the information is too obscure to be used in the analyses we are trying to conduct.

3 Data Preparation

We use hotel's reviews of Rakuten travel Inc. We defined influential sentences as those that influence readers to make them book the hotel. In practice, influential sentences are very sparse. So, in order to collect them efficiently, we used a heuristic that it is relatively more likely to find them in the sentences with exclamation marks ("!") located at their ends. We randomly extract 500 sentences which have more than one "!" at the end, and used for the analyses. Note that exclamation mark doesn't change the meaning of the sentence. We conducted a preliminary survey and found that our assumption works well.

We asked 20 subjects to extract influential sentences from the 500 sentences. The task is to extract sentences by which each subject thinks it influential enough to decide he/she wants to book or never to do the hotel. We asked them not to include their personal tastes. There are 84 influential sentences on which more than 4 subjects agreed. In the following sections, these 84 sentences will be called *the influential sentences* and the other sentences are regarded as *the non-influential sentences*.

4 Analysis of Target and Evaluation

We analyze classes of targets and evaluations which are most influential to the readers. Here, the targets are such as meals or locations of the hotels, and the evaluations are the reader's impressions about the targets such as good or convenient. We allow duplication of the classification, i.e. if a sentence contains more than one target or evaluation then we extract all the target or evaluation terms.

We categorized the targets into 11 classes and the evaluations into 7 classes (Table1). The table contains the Chi-square test results for each class. It indicates how significantly each class appears in the influential sentences compared to the non-influential sentences. "Less than 1%" means that the chance having the number of classes in the influential sentences and that in the non-influential sentences is less than 1%, if random distribution is assumed. "None" means there is no significant influence. The results of Chi-square test show that the three classes of target, "room", "meal" and "service" give influence to the readers (less than 1%), and "scenery" is also influential (less than 5%). Two classes of the evaluations, "human senses" and "features" are influential (less than 1%). "Features" are expressions describing the writer's view about particular targets in the hotel.

We found that some particular combinations of a target and an evaluation are influential (Table 2). "-" indicates infrequence (less than 6). We will discuss the combinations of "meal + human senses", "service + feelings" and "room/ meal/ service/ scenery + features".

In the combination of "meal + human senses", "human senses" are all about taste. The number of the influential sentences is 12, and the non-influential sentences are 19. We analyze each set of sentences, and found that the influential sentences include particular name of dish like "sukiyaki" much more often (less than 1%). Non-influential sentences include more abstract expressions, like "breakfast". The readers are influenced by particular food.

The combination of "feeling + service" appeared in influential sentences relatively more often(less than 2.5%). "Service" includes service of the hotel like "welcome fruit" or "staff's service". "Feeling" is influential only when it combines with "service" (ex. 1).

> Ex. 1: …there was happy surprise service at the dinner!!

"Features" is very frequent. Investigating the combination with targets, we found that "room", "meal" and "service" are the ones which made significant difference (less than 1%) by combining with "features". These are the key to make "features" more influential for readers. "Scenery" is a target originally created and has a significant difference less than 5%. It is a bit unexpected, but was useful information for some readers.

Table 1. Target and Evaluation with Chi-square test

Result of Chi-square test	Target	evaluation
Less than 1%	Room, meal, service	Human sense (e.g. delicious, stink), Features (e.g. marvelous, bad)
Less than 5%	Scenery	-
None	Location, staff, facility, hotel, bath, plan, price	recommendation (e.g. This is my recommendation) next visiting (e.g. I'll never use this hotel), feeling (e.g. happy) request (e.g. I want you to…), others (e.g. Thank you)

Table 2. Combination of Target and Evaluation with Chi-square test

	room	meal	bath	service	facility	scenery
features	less than 1%	less than 1%	NO	less than 1%	NO	less than 5%
feelings	NO	-	-	less than 2.5%	-	-
human senses	-	less than 1%	-	-	-	-

5 Personal tastes in the influential sentences

Although we instructed the subjects not to include particular personal tastes, we observed the selections of the influential sentences are different among the subject. 289 sentences are selected as influential sentences by at least one subject, and 94 sentences are selected by only one subject.

The personal tastes often appear on the target, so we analyzed differences of targets among the subject. We clustered the subjects based on their choice of the targets. For each subject, we create a frequency vector whose elements are including the most popular 7 targets, namely "location", "room", "meal", "bath", "service", "facility", and "scenery". Then the cosine metrics is applied to calculate the similarity between any pair of the subjects. Next, we run the hierarchical agglomerative clustering with the farthest neighbor method to form their clusters. Three figures, Figures 1 to 3, show the results of three clusters in Rader charts. Each of three clusters has a typical personal taste, namely groups who are influenced more by "service" very strongly (Fig. 1), by "meal" (Fig. 2) or by both "service" and "meal" (Fig. 3).

We analyze influential sentences by using the number of sentences including "service". Table 3 shows the numbers of sentences that were judged influential by certain numbers of subjects on "service". In this analysis, we categorize the influential sentences into positive and negative ones. For example, there were 2 positively influential sentences that were judged influential by 9 subjects. From Table 3, we can observe that the sentences can clearly be grouped into two;

sentences which 7 or more subjects judged influential (we will call them as a popular group) and sentences less than 7 subjects judged influential (unpopular group).
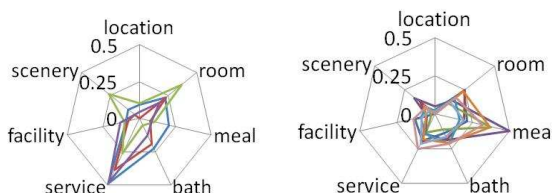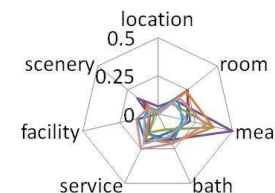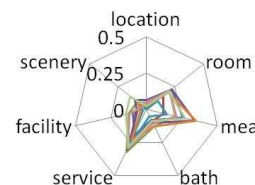

Figure 1. "Service" type　　Figure 2. "Meal" type

Figure 3. "Service & meal" type

Table 3: the number of influential sentences judged by certain number of subjects on "service"

	10 or more	9	8	7	6	5	Less than 5
Positive	3	2	1	0	1	5	33
Negative	3	3	1	0	0	2	4

In the "service" target, 63 sentences are selected as influential by at least one subject. Among them, 45 sentences are positive, 13 sentences are negative and 5 sentences are classified other (i.e. neither positive nor negative). There are four sets of data by combining positive-negative axis and axis. We will analyze them one by one.

[Negative & Popular]
There are 7 sentences in this group and we found that 3 of them include "feeling" evaluation, such as "surprised" or "angry". In contrast, there is no sentence including feeling in the negative & unpopular group. Also, very unpleasant events

like "arrogant attitude of hotel staff," "lost the luggage" and "payment trouble" are found negatively influential by many subjects.

[Negative & Unpopular]
There are sentences about staff's attitude in this group, too, but it is less important compared to the ones in the popular group. For example, staff's attitude is about greetings or conversation by the hotel staff. We believe it is depending on people if they care those issues or not.

[Positive & Popular]
In this group, there are 2 sentences that show unexpected warm service (ex. 2). Also, there are sentences that express high satisfactions not only in service but also in other targets, such as meal.

> Ex. 2: …they kept the electric carpet on because it was cold. We, with my elderly farther, were so glad and impressed!!

[Positive & Unpopular]
All sentences include some positive descriptions about services, such as "carrying the luggage" or "welcome fruit". Some subjects are influenced, but the others aren't. We believe it is because some people think that these are just usual services to be provided.

Now, we describe analyses on the "meal" target. There are 68 influential sentences selected by at least one subject. There are 58 positive sentences, 5 negative sentences and 4 sentences otherwise. We analyze the four groups, just like what we did for "service".

[Negative & Popular]
We find strong negative opinion about meal itself like "Their rice was cooked terrible", which are not found in the unpopular group. Many people are influenced when the meal is described badly.

[Negative & Unpopular]
There are 2 sentences about the situation of the restaurant, such as "crowded" or "existence of a large group of people". We believe that the most important feature of meal is taste, not the situation. Many people might know such situation happens by chance, so only some people cares about this kind of issue.

[Positive & Popular]
The sentences in both popular and unpopular groups include "delicious", but "delicious" with emphasizing adjectives, like "really delicious" were found only in the popular group.

[Positive & Unpopular]
The sentences including "cost performance" and "large portion" only appear in the unpopular group. We believe that the size might be influential to people who like to eat a lot, but people who might not be interested in them.

The analyses show that there is personal taste and we analyzed it in detail by examining the examples. It indicates that personalization is very important for the readers to find the reviews that might satisfy readers.

6 Conclusion

The main focus of our study is on the *reader's point view* to evaluate reviews, compared to the *writer's point of view* that was the major focus in the previous studies. We defined the influential sentences as those that could make the reader's decision. We analyzed the 84 influential sentences, based on the selection by the 20 subjects from the 500 sentences. We conducted the following two analyses.

1) We analyzed targets and evaluations in influential sentences. We found that "room", "service", "meal" and "scenery" are important targets, and "features" and "human senses" are important evaluations. We also analyzed combinations of the targets and evaluations. We find that some combinations make it more influential than each of them.

2) We analyzed the personal tastes. The subjects can be categorized into three clusters, which can be explained intuitively. We found that the most important targets to characterize the clusters are "service" and "meal".

There are many directions in our future work. One of the important topics is to conduct cognitive analysis on the influential sentences. We found that expressions can be very influential by adding a simple modifier ("really delicious"). Furthermore, many metaphorical expressions are found in influential sentences (this topic was not covered in this paper). We would like to conduct the cognitive analyses on these topics to clarify the characteristics of the reader's point of view. We believe it will reveal new types of information in reviews that is also useful for applications.

References

Alexandre Passos and Jacques Wainer, 2010, What do you know? A topic-model approach to authority identification, Proc. Of Computational Social Science and Wisdom of Crowds(NIP2010).

Anindya Ghose and Panagiotis G. Ipeirotis. 2007. Designing novel review ranking systems: Predicting usefulness and impact of reviews. Proc. of the International Conference on Electronic Commerce (ICEC), pp. 303-309.

Bo Pang, Lillian Lee and Shivakumar Vaithyanathan. 2002. "Thumbs up? Sentiment Classification using Machine Learning Techniques". Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP). pp. 79–86.

Bo Pang and Lillian Lee. 2005. "Seeing stars: Exploiting class relationships for sentiment categorization with respect to rating scales". Proceedings of the Association for Computational Linguistics (ACL). pp. 115–124.

Daisuke Okanohara and Jun'ichi Tsujii. 2007.Assigning Polarity Scores to Reviews Using Machine Learning Techniques. Journal of Natural Language Processing. 14(3). pp. 273-295.

Koppel, M. and Schler, J. 2006. "The Importance of Neutral Examples in Learning Sentiment". Computational Intelligence. 22(2). pp.100-109.

Maya Ando and Shun Ishizaki. 2012, Analysis of influencial reviews on Web(in Japanese), Proc. Of the 18th Annual Conference of the Association for Natural Language Processing, pp. 731-734.

P. Victor, C. Cornelis, M. De Cock, and A. Teredesai. 2009. "Trust- and distrustbased recommendations for controversial reviews." in Proceedings of the WebSci'09, Society On-Line.

Peter Turney. 2002. "Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews". Proceedings of the Association for Computational Linguistics. pp. 417-424.

Thelwall, Mike; Buckley, Kevan; Paltoglou, Georgios; Cai, Di; Kappas, Arvid. 2010. "Sentiment strength detection in short informal text". Journal of the American Society for Information Science and Technology 61 (12). pp. 2544-2558.