Impression comments analysis in consideration of user profile of the online guests' review

Koichi Tsujii [*], Koji Tanaka [†],

Yoshikatsu Fujita [‡], Kazuhiko Tsuda [§]

Abstract

Online hotel reservation is widely used for planning travel in recent years. The number of female guests for leisure usage has been increased more and more, to exceed that of male guests. The female travel demand called "a woman trip" became an important market even for the travel industry. Potential guests usually refer to the guests' reviews before making their reservations. Guest reviews consist of a user profile, numerical evaluation and impression comments. This reflects guest's opinion about the hotel. In this paper, we aim to extract the characteristic expressions from impression comments which the feature of the guests' review caused by gender with text mining. As a result, we have found that an impression comment complements the content of numerical evaluation, and showed the difference in the characteristic expressions by gender.

Keywords: Hotel reservation, Travel, User review, Text mining, Service Innovation and Marketing

1   Growth of Internet hotel reservation

With the widespread use of Internet, most hotel bookings are made by online. Users can make hotel reservations in many ways, and the number of hotel reservations made by online has been increased more than ten-fold in this decade [1]. According to the Survey on travel by Yahoo, the first motivation for going out for travel comes from the conversation with family, friends, and acquaintances, and the second comes from browsing Web sites and Web advertisement [2].

According to the travel annual report 2013, the growth rate of leisure travel is observed higher mostly for female guests [3]. And the number of female travellers went over the number of male in 2013. Female guests' demands for travel called "a woman trip" became an important growth market for the recent travel industry. And this market has been drawn the highest attention from

---

[*]   Nippon Travel Agency Co., LTD, Tokyo, Japan
[†]   Hitachi Government & Public Sector Systems, Ltd, Tokyo, Japan
[‡]   Department of Sociology, Teikyo University, Tokyo, Japan
[§]   Graduate School of Business Sciences, University of Tsukuba, Tokyo, Japan

the travel industry people [4]. Moreover, according to the survey conducted by JTB, the travel reservation method employed by female guests is mostly an internet.

When making an online hotel room reservation, the users are expected to select and to reserve a hotel room by themselves, unlike when using a travel agency. In order to choose the best accommodations, users used to collect information through travel reservation sites. Users are not satisfied with such information as supplied only by hotels but seeking for the reviews from other guests. Such feature information of hotels is important for users [5]. Hotel reviews written by guests are important reference information for choosing appropriate accommodations. Such reviews are helpful for people searching for the reputation of hotels. In addition, hotel managers also analyze such hotel reviews in order to improve their services. Thus, hotel reviews are actively used by both online hotel reservation users and accommodation service providers. While the number of female guest increases, to find the factors which draw female guests' attentions is an important marketing challenge. In this paper, we aim at extracting the difference in the Guests' review by gender.

The rest of the paper is organized as follows: In chapter 2, we briefly describe the related works and give basic concepts. In chapter 3 the features of the numerical evaluation criteria of a Guests' reviews are shown. In Chapter 4 our proposed method for feature extraction from hotel reviews is presented. In Chapter 5, analytical results are shown. Conclusions and future work are presented in Chapter 6.

2   Usage of Guests' review

Online hotel reservation sites usually provide users with criteria for selecting accommodations. Guest reviews are one of those criteria. Guest reviews which reflect the guests' opinions are provided by virtually all hotel reservation websites [6]. Guest reviews consist of a user profile, a numerical evaluation, and an impression comment. That reflects a guest's opinion about the hotel. The age and gender of the user are included in the user profile. The numerical evaluation criteria provided by the websites are loosely classified into five categories: 'meals', 'bath', 'service', 'location' and 'room'. These aspects are important for guests and are rated on a five or ten-point scale. The impression comment column is where guests leave their free impressions of the hotel. Simple impressions and detailed evaluations are expressed in the impression comments.

Many previous studies have shown the importance of such guest reviews [7]- [14]. However, a problem exists with the method by which information is shared. In the case of popular hotels, the number of reviews can become very large, and all numerical evaluations can be fixed at high scores. Therefore, understanding the characteristics of the accommodations is difficult. Moreover, user reviews are written without guidelines; therefore, their content can be varied and ambiguous. Numerical evaluation criteria and impression comments are unrelated items.

Guest reviews have become very important not only for the potential guests themselves but also for the hotels and travel agencies. Managers of accommodation also pay more and more attention for such reviews, because of growing impact of guest reviews. There are some previous researches about the approach of finding evaluation and the improving point from online guests' review [15]- [20]. Kasper et al. presented a web based opinion mining system for hotel reviews and user comments [15]. Ouchi et al. carried out the questionnaire to accommodations, and studied the technique for accommodations evaluation [17]. Tsujii et al. employed text mining for evaluating travelers' impressions and found that their interests varied depending on the areas in which they stayed. They also classified the reviews and the comments on the website review

board into the prepared evaluation categories by analyzing their meanings and assigned positive and negative attributes to each expression. The results revealed that users had some strong opinions that did not accurately reflect the numerical evaluations [21] [22]. These studies revealed that the potential users only paid attention to the website reviews in making reservations and the past studies proposed the new methods to evaluate both a large number of reviews and the importance of review information.

In the previous research, although the feature of impression comments and the feature in the numerical evaluation were investigated, the feature from a profile was not investigated. The guest's gender and age are found on the profile information of the Guests' review. By using these profile information, it is assumed that the difference in the gender of a hotel guest would be revealed. In this paper, we apply the method of Tsujii's previous research, extract characteristic expressions from an impression comment using text mining, and investigate the feature. We divide the Guests' review into the gender which the hotel guest posted, and analyze the feature of numerical evaluation. Next, we analyze the feature of an impression comment. And we extract a feature of guests' review by unifying the numerical evaluation criteria and the characteristic expressions obtained from the impression comment. And we investigate whether the feature of gender exists in a Guests' review.

3   Feature of numerical evaluation

We examine the features of numerical evaluation which is one of the elements of hotel reviews. To characterize the accommodation areas that differ from visiting purpose, we focus on 10 areas, e.g. city direction (Tokyo, Osaka, Nagoya, Hakata, and Sapporo) and leisure direction (Hakone, Yufuin, Kusatsu, Noboribetsu, and Dogo). The review data comes from 100 reviews from 10 frequently posted accommodations (10,000 reviews in total). Table 1 shows the frequency of the five-point score that we obtained from the online hotel reservation site 'jalan.net' (http://www.jalan.net). Numerical evaluation criteria consist of six subjects, which is Entirety, Room, Bath, Meal, Service, Purity. Among 10,000 reviews, the majority of users gave a rating of 4 or 5 out of 5. A low rating of 1 or 2 was less than 10% in all reviews.

Table 1: Frequency of numerical evaluation scoring

Point	Entirety	Room	Bath	Meal	Service	Purity
5	4,314	3,870	3,891	3,700	4,637	4,596
4	4,089	3,910	3,575	2,371	3,342	3,478
3	1,121	1,725	1,922	935	1,521	1,485
2	367	352	396	263	330	329
1	109	143	105	85	170	112
Total	**10,000**	**10,000**	**9,889**	**7,354**	**10,000**	**10,000**

Table 2 shows the frequency of numerical evaluation scores classifying reviews by gender. The number of reviews by gender is roughly the same number. Similar to Table 1, the score of the review is high in both, and it is difficult to find a tendency of evaluation. When using numeric evaluation information, it is almost impossible to provide specific information for distinguishing the characteristics of the subjects analyzed.

Table 2: Frequency of numerical evaluation scoring by gender

Point	Entirety	Room	Bath	Meal	Service	Purity
Female	**5,248**	**5,248**	**5,187**	**3,865**	**5,248**	**5,248**
5	2,429	2,139	2,104	2,048	2,565	2,506
4	2,021	1,939	1,856	1,166	1,693	1,748
3	546	896	956	480	721	747
2	203	205	222	133	191	186
1	49	69	49	38	78	61
Male	**4,752**	**4,752**	**4,702**	**3,489**	**4,752**	**4,752**
5	1,885	1,731	1,787	1,652	2,072	2,090
4	2,068	1,971	1,719	1,205	1,649	1,730
3	575	829	966	455	800	738
2	164	147	174	130	139	143
1	60	74	56	47	92	51

4   Extraction of characteristic expressions for evaluation using text mining

A text mining is the technology which finds new knowledge after doing natural language processing of comments, the tendency and the feature of the comments are picked out using the technology of data mining. We extracted the characteristic expressions considering of gender that appeared in each impression comment using text mining. The analytical flow of text mining using the guest reviews is shown in Figure 1.

```
┌─────────────────────────────────┐
│       Collection of data        │
└─────────────────────────────────┘
                 ↓
┌─────────────────────────────────┐
│     Extraction of future words  │
└─────────────────────────────────┘
                 ↓
┌─────────────────────────────────┐
│ Extraction of characteristic    │
│   expressions including         │
│       feature words             │
└─────────────────────────────────┘
                 ↓
┌─────────────────────────────────┐
│ Association with numerical      │
│ evaluation criteria and         │
│ Separation of positive          │
│ negative evaluation             │
└─────────────────────────────────┘
                 ↓
┌─────────────────────────────────┐
│ characteristic expressions      │
│   analysis by gender and area   │
└─────────────────────────────────┘
```
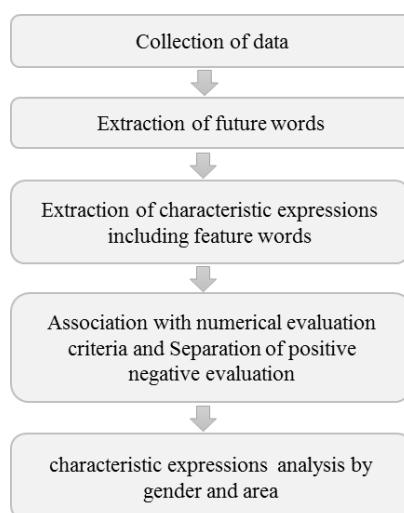
Figure 1: Text mining flow

Using the accommodation review shown in Chapter 3, we counted the frequency of each word using morphological analysis and analyzed the tendency of the feature word that the user empha-sized during the hotel stay. Then, we counted the characteristic expressions that pair of feature words and evaluation words the hotel received through dependency analysis. We separated each characteristic expression into positive and negative components and assigned them to the numer-ical evaluation. Then, we analyzed the evaluation results of hotels by area and gender.

When analysing guest reviews, building of several dictionaries are important tasks. Synonym dictionary: The reviews varied in their means of expression when it came to identical content, such as 'bath room' and 'bath tub', or 'staff' and 'concierge'. We made clusters of attributes that frequently appeared and created a dictionary of synonyms that organized words expressing the same meaning. We registered 83 pairs 698 words in the synonym dictionary. The sample of the synonym dictionary is shown in Table 3.

Table 3. Sample of synonym dictionary

Synonym	alternative impression			
bath	bathroom	bathtub	bath/spa	showerroom
staff	concierge	porter	hotel clerk	bell boy
bed	bedding	mattress		
internet	wifi	wi-fi	lan	

Numerical evaluation dictionary: Guest reviews consist of impression comments and numerical evaluation criteria, and we examine how to associate them using a numerical evaluation dictionary. Confirming each of the synonyms and the future words, we classified each characteristic expression into five kinds of numerical evaluation criteria frequently used at the hotel reservation websites: meal, bath, service, purity, and room. Through these processes, we gave the hierarchical concept that feature words are gathered by the synonym and synonym are gathered by the numerical evaluation criteria. We registered 321 words and synonyms in this dictionary. We did not perform dictionary registration at the time of classification for words and synonyms which unrelated to numerical evaluation criteria. The relationship between synonyms and the numerical evaluation criteria is shown in Figure 2.
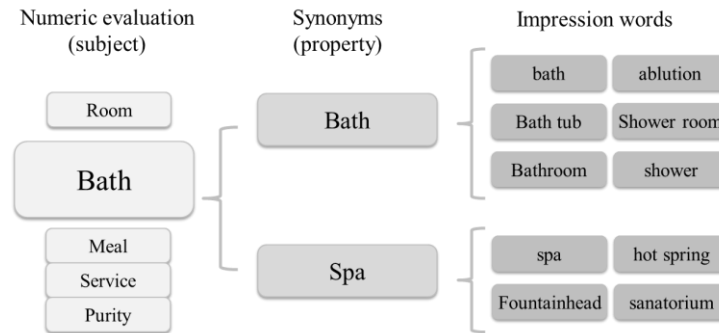

Figure 2. Relationship between synonyms and numerical evaluation

Evaluation attribute dictionary: To evaluate an impression comment, we labelled positive and negative polarities to the dependency relationship of a feature word and an evaluation word, and built the evaluation attribute dictionary. Leveraging Kobayashi et al reputation analysis method in this study, an affirmative expression and a negative expression were classified into the subject, the property, and the evaluation [23]. In this paper, the numerical evaluation criteria are applied to subjects, synonyms are applied to properties, and the words indicating evaluation are applied to evaluation. Using the evaluation attribute dictionary, we connected the feature words of impression comments with the numerical evaluation criteria and categorized them as 'positive' or

'negative' in the evaluation attribute dictionary, as shown in Table 4. 1,111 sets of characteristic expressions were registered in the evaluation attribute dictionary.

Table 4. evaluation attribute dictionary

Characteristic expression	Subject (numerical evaluation)	Property (synonym)	Evaluation
room - large	room	room	positive
staff - helpful	service	staff	positive
hotel - nice	hotel	hotel	positive
breakfast - delicious	meal	breakfast	positive
room - clean	purity	room	positive
room - small	room	room	negative
hotel - old	hotel	hotel	negative
bath - small	bath	bath	negative
room - dirty	purity	room	negative
staff - regret	service	staff	negative

In the classification process, we excluded unclear or neutral expressions and requirement expressions that had no evaluation polarity. We used the abovementioned synonym dictionary, the numerical evaluation dictionary and the evaluation attribute dictionary which allowed us to evaluate the characteristic expressions extracted by text mining according to the numerical evaluation criteria. In the following chapter, we extract the characteristics of the hotel review according to the area from impression comment separated by gender. And we present the result of the analysis.

5 Characteristic expression analysis in consideration of numerical evaluation

As discussed in chapter 3, the numerical evaluation scores can become vague for popular hotels. In this situation, it is difficult to find the evaluated object only by the numerical evaluation. In this section, we analyze the characteristics of evaluation by gender. We extracted the positive and the negative expression from the impression comment using the extraction method of the char-acteristic expressions shown in chapter 4. And the frequency of the characteristic expressions by numerical evaluation criteria in consideration of gender is shown in Table 5.
Table 5 summarizes the frequency of characteristic expressions, and as a result, there are many positive characteristic expressions in both areas. When we summed up the frequency of the char-acteristic expressions by gender, despite almost the same number of male and female guests, it turns out that female guests tend to write 1.5 times characteristic expressions compared with male guests. Female guests tend to carefully evaluate hotels in hotel reviews.

Table 5. Characteristic expression analysis in consideration of gender

		Room	Bath	Meal	Service	Purity	Total
Male	Positive	535	658	1,021	858	271	3,343
	Negative	249	128	118	53	25	573
	Total	**784**	**786**	**1,139**	**911**	**296**	**3,916**
Female	Positive	791	875	1,575	1,365	353	4,959
	Negative	314	184	141	65	34	738
	Total	**1,105**	**1,059**	**1,716**	**1,430**	**387**	**5,697**
Total	Positive	1,326	1,533	2,596	2,223	624	8,302
	Negative	563	312	259	118	59	1,311
	Total	**1,889**	**1,845**	**2,855**	**2,341**	**683**	**9,613**

Table 6. The characteristic expression analysis in consideration of gender (leisure area)

		Room	Bath	Meal	Service	Purity	Total
Male	Positive	198	461	721	502	110	1,992
	Negative	57	89	79	35	14	274
	Total	**255**	**550**	**800**	**537**	**124**	**2,266**
Female	Positive	316	658	1,203	821	165	3,163
	Negative	96	146	107	41	13	403
	Total	**412**	**804**	**1,310**	**862**	**178**	**3,566**
Total	Positive	514	1,119	1,924	1,323	275	5,155
	Negative	153	235	186	76	27	677
	Total	**667**	**1,354**	**2,110**	**1,399**	**302**	**5,832**

Table 7.   The characteristic expression analysis in consideration of gender (city area)

		Room	Bath	Meal	Service	Purity	Total
Male	Positive	337	197	300	356	161	1,351
	Negative	192	39	39	18	11	299
	Total	**529**	**236**	**339**	**374**	**172**	**1,650**
Female	Positive	475	217	372	544	188	1,796
	Negative	218	38	34	24	21	335
	Total	**693**	**255**	**406**	**568**	**209**	**2,131**
Total	Positive	812	414	672	900	349	3,147
	Negative	410	77	73	42	32	634
	Total	**1,222**	**491**	**745**	**942**	**381**	**3,781**

Next, we investigated the difference in the characteristic expression in both areas. The characteristic expression analysis in each area is shown in table 6 and table 7. Table 6 shows the characteristic expressions in leisure area. Comparing the frequency of the characteristic expressions by the area, leisure area is about 1.5 times larger than the city area. This result shows that when staying at leisure area, there are many target properties to be evaluated. And the characteristic expression of the leisure area frequently appears in the order of meal, service, bath numerical evaluation criteria. When using accommodations in leisure, these three types of numeric evaluation criteria are important for the guests. And the frequency of the characteristic expression of the female guest is 1.5 times more than that of the male. Table 7 shows the characteristic expression in the city area. Many characteristic expressions about the room, service, and meal have appeared. The number of characteristic expression of the female guests is 1.3 times more than that of the male guests.

Next, Figure 3 and Figure 4 show the comparison of the frequency of characteristic expressions by area for each gender. Figure 3 shows the frequency of the characteristic expressions in leisure

area. Although the number of guests is almost the same, the female guests wrote many characteristic expressions about meal and service. Female guests regard meals to be more significant rather than male guests. In addition, female guests have negative characteristic expressions for bath compare to the other subjects. Figure 4 shows the evaluation of the frequency of characteristic expression in the city area. At the first, there is a feature in the characteristic expression about room and service. There are differences in the evaluation of these two subjects, while the evaluations for services are mostly positive evaluation, but the evaluation for the room includes some negative evaluations. Guests can freely fill their opinions in the comments and post impressive events during their stay. Therefore, we can investigate criteria of high interest in the area from a different viewpoint from the unified numerical evaluation items.

As mentioned above, female guests posted many comments compared to male guests. And we also clarified that numerical evaluation criteria of high interest can be found from the characteristic expression by the proposed method.
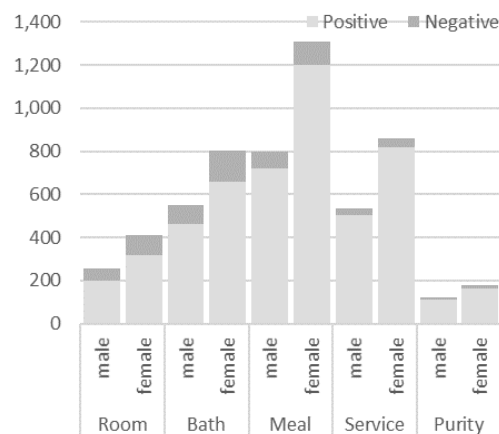

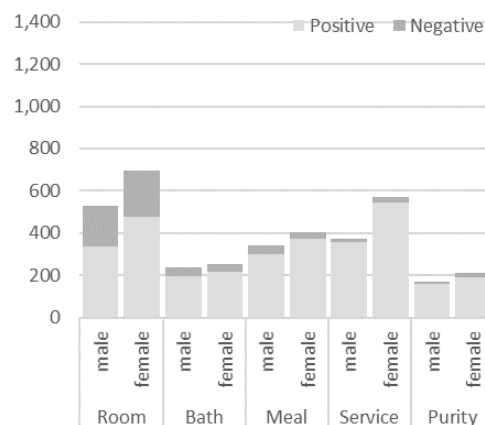Figure 3. Characteristic expression ratio (leisure area)

Figure 4. Characteristic expression ratio (city area)

6 Concluding Remarks

We have analyzed the relationship between numerical evaluation and the characteristics of impression comments presented for online hotel reviews. We have given the most attention for the gender difference, which causes remarkable feature for the online hotel reviews. As a result, we have shown that the gender difference plays an important role in the expression of impression comments. By treating a Guests' review based on a user's profile such as gender, we consider that we can make a Guests' review easier to use and more helpful. In the future, we will increase the efficiency for obtaining evaluation information and will improve information presentation methods to those who are in charge of hotel services.

References

[1] 'The sale 2013 in a tourist industry', Japan Association of Travel Agents, 2013.

[2] Trend survey about travel at New Year's, Yahoo! JAPAN Ads White Paper, 2015 (Japanese).

[3] Annual Report on the Tourism Trends Survey 2013, Japan Travel Bureau Foundation, 2014 (Japanese).

[4] Survey about a woman trip and information collection, JTB Tourism Research & Consulting, 2014 (Japanese).

[5] 'Consumer Trust in Online, Social and Mobile Advertising Grows', Nielsen,2012 http://www.nielsen.com/us/en/newswire/2012/consumer-trust-in-online-social-and-mobile-advertising-grows.html(2013/11/25 accessed).

[6] Tsujii K, Takahashi M, Fujita Y, Tsuda K, 'Extraction of Accommodation Evaluation by Foreign Tourists with Text Mining', Int. J. Trade, Econ. Financ. 5(2), 2013, pp. 130-135.

[7] 'Research: What makes a helpful review', Tripadvisor, 2015, http://www.tripadvisor.co.uk/TripAdvisorInsights/n2617/research-what-makes-helpful-review (2015/4/2 Accessed)

[8] Rosen E, 'The Anatomy of Buzz: How to Create Word of Mouth Marketing', Random House, 2002.

[9] Zhang, L., Pan, B., Smith, W., & Li, X. R. 'An exploratory study of travelers' use of online reviews and recommendations', Information Technology & Tourism, 11(2), 2009, pp. 157-167.

[10] Gretzel U, Kyung H.Y, 'Use and Impact of Online Travel Reviews', Information and Communication Technologies in Tourism, 2008, pp. 35–46.

[11] Vermeulen I.E, Seegers D, 'Tried and Tested: The Impact of Online Hotel Reviews on Consumer Consideration', Tourism Management, Vol. 30(1), 2009, pp. 123–127.

[12] Sparks, B.A. and Browning, V. 'The Impact of Online Reviews on Hotel Booking Intentions

and Perception of Trust', Tourism Management 32(6), 2011, pp. 1310-1323.

[13] O'Connor P, 'User-Generated Content and Travel: A Case Study on Tripadvisor.Com', Information and Communication Technologies in Tourism 2008, 2008, pp. 47-58.

[14] Nagashima K, 'Customer Decision-making in the Service of Hospitality Industry: The Role of Word of Mouth', Toyo University Faculty of Business Administration Management Journal 74, 2009, pp. 91-108.

[15] Kasper W, Vela M, 'Sentiment Analysis for Hotel Reviews', Proc. Comput. Linguist. Conf. 2011, pp. 45-52.

[16] Tanabe W, Goto M, 'One Consideration about the User Review Analysis to Support the Strategic Construction of the Accommodations', the Musashi Institute of Technology Environmental Information Department Information Media center Journal Vol. 9, 2008 (Japanese).

[17] Ouchi A., and Oyanagi Y., 'Study on Hotel Evaluation System of Japan', Kaihatsu Koho (539), 2008 (Japanese).

[18]   O'Mahony M.P, Smyth B, 'Learning to recommend helpful hotel reviews', Proc. Third ACM Conf. Recomm. Syst. - RecSys '09, ACM Press 2009, 2009, pp. 305-308.

[19]   Morikawa T, 'An Evaluation of Using Hotel and Ryo-kan', Nagasaki International University Selection of Treatises 1, 2001, pp. 273-280 (Japanese).

[20] Bjørkelund E, Burnett T.H, Nørvåg K, 'A study of opinion mining and visualization of hotel reviews', Proc. 14th Int. Conf. Inf. Integr. Web-Based Appl. Serv. - IIWAS '2012, 2012, pp. 229-238.

[21] Tsujii K, Tsuda K. 'The attention information extraction method from a stay review using text mining', Journal of Digital Practices. 3, 2012, pp. 289–296 (Japanese).

[22] Tsujii K, Tsuda K, Fujita Y, 'The extraction method of the service improvement information from Guests' review', Adv. Knowledge-Based Intel. Inf. Eng. Syst., 2013, pp. 250–257.

[23] Kobayashi N, Inui K, Matsumoto Y, Tateishi K, Fukushima T, 'Collecting Evaluative Expressions for Opinion Extraction', Journal of natural language processing 12(3), 2005, pp. 203-222 (Japanese).