

## COMPARING COMPETITIVE ADVANTAGES OF HOTEL SERVICES USING OPINION MINING OF ONLINE CUSTOMER REVIEWS: A STUDY FOCUSING ON 5-STAR HOTELS IN SEOUL, KOREA

SOHYUN PARK, MINJUNG KWAK\* AND JIYU CHOI

Department of Industrial and Information Systems Engineering  
Soongsil University

369 Sangdo-ro, Dongjak-gu, Seoul 06978, Korea

{sohyunpark; jiyu126}@soongsil.ac.kr; \*Corresponding author: mkwak@ssu.ac.kr

Received February 2020; accepted May 2020

**ABSTRACT.** *To survive and succeed in the hotel industry, it is critical for a hotel to understand its customers' needs and their satisfaction levels. Opinion mining of online customer reviews can be a promising solution to this end. However, opinion mining for a single hotel alone is insufficient to analyze the strengths and weaknesses of the hotel. For a better understanding of the customer satisfaction and the current service quality, the quality of competitors should be considered together. This paper presents a competitive study that identifies and compares the competitive advantages of hotel services using opinion mining. The top 10 most reviewed 5-star hotels in Seoul are chosen for the analysis, and their 3,313 Korean reviews posted on TripAdvisor.com in 2015-2018 are crawled and analyzed. A lexicon-based attribute-level opinion mining is conducted for nine service attributes (i.e., room environment, service, basic facility, dining, location, auxiliary facility, furniture and appliances, bathroom, and price), and the percentages of positive and negative reviews are extracted for each attribute which are in turn converted to three comparison indices: Positive Rate, Satisfaction Score, and Integrated Index. The proposed analysis enables a hotel to identify its relative performance and rank in the market at the attribute level and helps understand its strengths and weaknesses compared to the competitors.*

**Keywords:** Opinion mining, Review mining, Text mining, Sentiment analysis, Hotel management, Service quality, Customer relationship management

**1. Introduction.** The hotel industry is one of the major sectors of the hospitality industry where understanding customer needs and their satisfaction levels is essential to achieve success in the market. Recently, as online customer reviews have emerged as an effective means of assessing and analyzing customer satisfaction [1], the hotel industry is showing increasing interest in the application of opinion mining, i.e., a natural language processing (NLP) technique that extracts people's sentiments (opinion, attitude, subjectivity, emotion, etc.) from written text [2-4]. Numerous studies have been presented targeting the hotel industry in the last decade, including the studies [5-11].

Opinion mining is usually conducted for a single hotel and classifies a hotel's reviews mainly into two groups: positive and negative. The classification can be conducted not only at the review level but also at the attribute level. Thus, it enables to measure the service quality of the hotel and individual attributes and is useful in identifying factors of satisfaction and dissatisfaction. However, such opinion mining for a single hotel is insufficient to understand where the hotel stands in the market with regards to customer satisfaction. For a better assessment of the service quality, the quality of competitors should be considered together. This is why this paper proposes a competitive analysis.

Focusing on 5-star hotels in Seoul, Korea, this paper conducts a competitive analysis based on opinion mining. Top 10 most reviewed hotels are chosen for the analysis, and their 3,313 Korean reviews created in 2015 through 2018 on the world's largest travel site TripAdvisor ([www.tripadvisor.com](http://www.tripadvisor.com)) are crawled and analyzed. In terms of the method, opinion mining is conducted based on the authors' previous work [11]. A lexicon-based, attribute-level analysis is performed for each hotel with regard to nine service attributes (i.e., room environment, service, basic facility, dining, location, auxiliary facility, furniture and appliances, bathroom, and price) and their 34 sub-attributes, as shown in Table 1. For each attribute and sub-attribute, the percentages of positive and negative reviews are identified.

TABLE 1. The dictionary of hotel service attributes [11]

Attribute	Sub-attribute
Room environment	Room, soundproof, temperature, floor/wall, window, condition/cleanliness, smell, view, convenience item
Service	Staff, receptionist/concierge, Wi-Fi connection
Basic facility	Mood, facilities/interior, lobby, lounge, parking lot, landscape/garden
Dining	Food, room service
Location	Location, transportation
Auxiliary facility	Auxiliary facilities, fitness, swimming pool/sauna/spa, business center/banquet hall
Furniture/appliances	Furniture, bed, curtain/carpet, appliances/electronics
Bathroom	Bathroom, amenity, water
Price	Price

In this study, the competitive analysis is conducted at the attribute level. For each hotel and each attribute, the level of customer satisfaction is assessed using three indices: *Positive Rate*, *Satisfaction Score*, and *Integrated Index*, which will be discussed later in detail. The proposed analysis enables hotels to identify their relative performance and rank in the market at the attribute level. It helps them understand their strengths and weaknesses compared to the competitors, and based on the results, the hotels can establish management strategies to improve their competitive advantages.

The rest of the paper is organized as follows. Section 2 briefly introduces the opinion mining model and indices used for the competitive analysis. Section 3 shows the main results from the analysis. Section 4 summarizes and concludes the paper.

**2. Research Method.** In this study, the opinion mining is conducted based on the model presented in the authors' previous work [11]. The model uses a dictionary-based approach and conducts sentiment analysis for the predefined sub-attributes in Table 1. To be more specific, for each review, the model identifies whether or not the review refers each sub-attribute, and if so, what sentiment (between positive and negative) is behind. After the sentiment for 34 sub-attributes is extracted, the sentiment for 9 attributes is calculated. Here, the strength of sentiment (e.g., very positive vs. positive) is not considered in the model; if a review shows both positive and negative feelings for the same term, both sentiments are recorded. The model also identifies the most referred sub-attributes in each sentiment group, which gives an idea about the reason why the customers have such opinions. For more detailed description of the model and the method, refer to [11].

Table 2 shows an example of the results obtained by the model. Hotel A is the most reviewed hotel among the 10 hotels. For a single hotel, the model first shows the reference rate of each attribute (i.e., the percentage of the reviews that refers the attribute at least

TABLE 2. An example of opinion mining for a single hotel: Hotel A

Attribute	% of referred	% of positive and negative reviews	Most referred sub-attribute (% referred in each sentiment)
Room environment	73%	Positive 47% Negative 6%	Condition/cleanliness (74%), view (64%) Soundproof (54%), smell (40%)
Service	82%	Positive 70% Negative 4%	Staff (87%), receptionist/concierge (35%) Wi-Fi connection (22%), staff (4%)
Basic facility	58%	Positive 39% Negative 2%	Mood (74%), facilities/interior (71%) Parking lot (33%), lounge (4%)
Dining	51%	Positive 34% Negative 2%	Food (66%), room service (56%) Food (4%)
Location	45%	Positive 37% Negative 1%	Location (85%), transportation (62%) Transportation (4%)
Auxiliary facility	24%	Positive 18% Negative 2%	Auxiliary facilities (89%), fitness (79%) Swimming pool/Sauna/Spa (12%)
Furniture/appliances	17%	Positive 10% Negative 1%	Bed (67%), furniture (56%) Furniture (12%)
Bathroom	16%	Positive 6% Negative 1%	Bathroom (62%), amenity (52%) Bathroom (6%), water (6%)
Price	11%	Positive 6% Negative 3%	Price (55%) Price (26%)

once) that reflects the relative importance of the attribute [12]. For instance, the result shows that 73% and 82% of the Hotel A’s reviews refer the room environment and the service, respectively. More specifically, the model shows the ratios of positive and negative reviews for each attribute and the most referred sub-attributes regarding each sentiment. For instance, 47% of the review mentions the room environment in a positive way while 6% of the review leaves a negative opinion. Among the positive reviews, 74% referred the room condition and cleanliness as positive.

To assist in effective and fair comparison of the subject hotels, this paper proposes three comparison indices that are assessed for individual hotel ( $i \in I$ ) and attribute ( $j \in J$ ). These are *Positive Rate*, *Satisfaction Score* [11,13], and *Integrated Index*, each of which is calculated by Equations (1)-(3), respectively.

$$\text{Positive Rate}_{ij} = \frac{\text{Number of hotel } i\text{'s reviews referring attribute } j \text{ positively}}{\text{Total number of entire reviews on hotel } i} \times 100 \quad (1)$$

$$\text{Satisfaction Score}_{ij} = \frac{\text{Positive rate}_{ij}}{\text{Positive rate}_{ij} + \text{Negative rate}_{ij}} \times 100 \quad (2)$$

$$\text{Integrated Index}_{ij} = \frac{\text{Normalized Positive Rate}_{ij} + \text{Normalized Satisfaction Score}_{ij}}{2}$$

$$\text{where Normalized Positive Rate}_{ij} = \frac{\text{Positive Rate}_{ij} - \min_{i \in I} \text{Positive Rate}_{ij}}{\max_{i \in I} \text{Positive Rate}_{ij} - \min_{i \in I} \text{Positive Rate}_{ij}}, \quad (3)$$

$$\text{Normalized Satisfaction Score}_{ij} = \frac{\text{Satisfaction Score}_{ij} - \min_{i \in I} \text{Satisfaction Score}_{ij}}{\max_{i \in I} \text{Satisfaction Score}_{ij} - \min_{i \in I} \text{Satisfaction Score}_{ij}}$$

Positive Rate helps Hotel A diagnose its service quality and customer satisfaction level. In addition to the absolute number of positive rate, how much the positive rate is relatively stronger than the negative rate [11,13] can also provide useful information. Satisfaction Score measures such relative strength of positive opinions. Positive Rate and Satisfaction

Score are useful when comparing multiple hotels in terms of an attribute. However, there exists one problem that they cannot support direct comparison of different attributes of a hotel (i.e., which attributes are in good performance and which need more improvement). To clarify the relative strengths and weaknesses of a hotel, an index involving other hotel's results together is needed.

Integrated Index is an index newly proposed in this paper. It enables direct comparison of different attributes as well as different hotels. It first normalizes the Positive Rate and the Satisfaction Score to resale them into 0-100 values. Here, 0 corresponds to the minimum amongst all subject hotels (i.e., the worst), whereas 100 corresponds to the maximum (i.e., best). By averaging the two normalized values, the Integrated Index serves as a single measure that enables efficient comparison among the hotels. Integrated Index of 100 means that the hotel is the best hotel in terms of both Positive Rate and Satisfaction Score. In contrast, 0 means that the hotel is the worst in both perspectives.

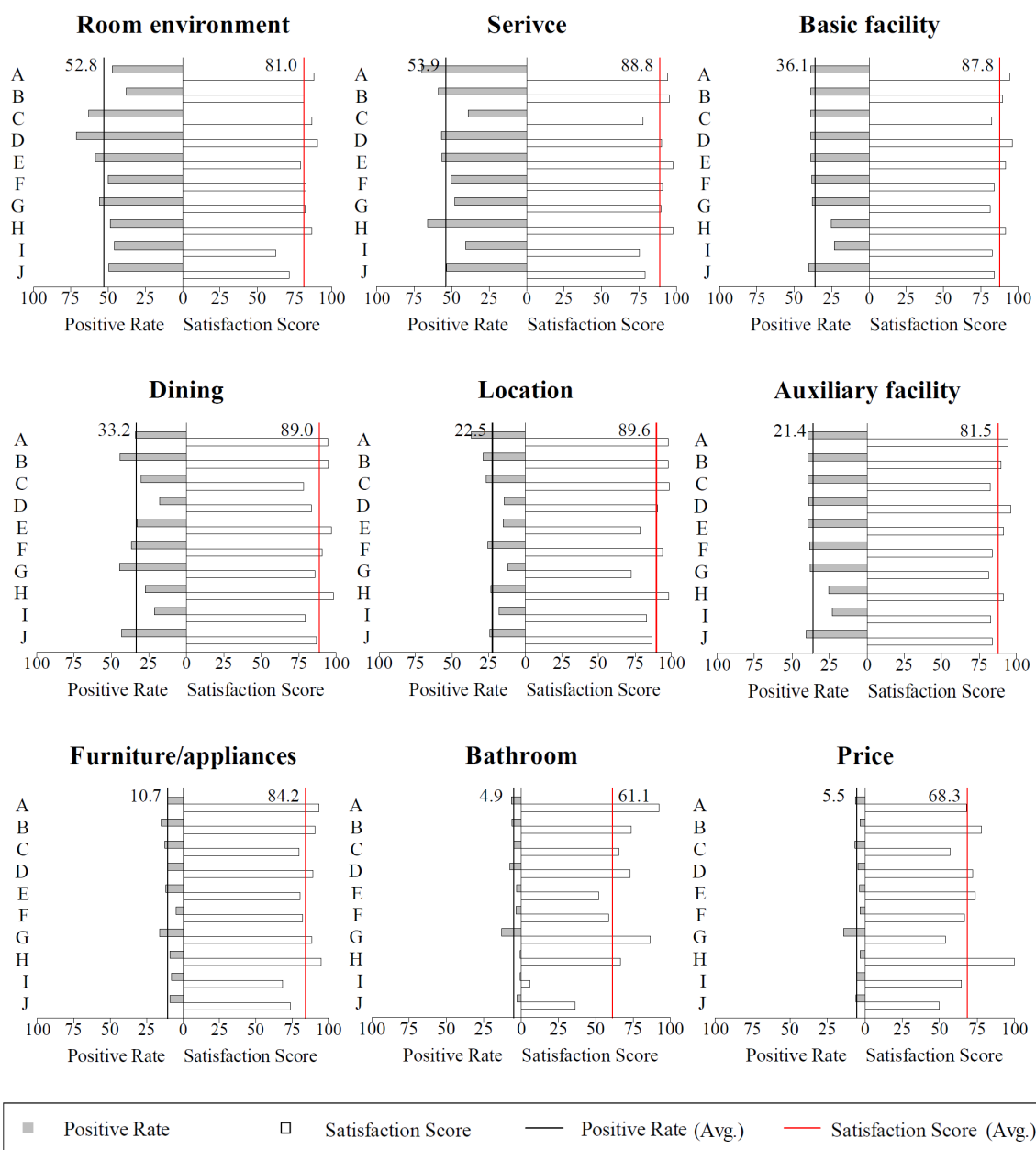


FIGURE 1. Positive Rate and Satisfaction Score of the Hotels A to J

**3. Main Results.** Figure 1 and Table 3 show the main results of the competitive study. Figure 1 compares the 10 subject hotels in terms of the Positive Rate and Satisfaction Score from the viewpoint of each service attribute. The solid lines in each bar graph indicate the average values of the indices. For example, as for the room environment, the 10 subject hotels achieve 52.8% of Positive Rate and 81 points of Satisfaction Score on average. Hotel D achieves the highest scores in both indices, which implies that it is the best hotel in the market from the room environment perspective. Hotel C and Hotel A are the next with respect to the Positive Rate and the Satisfaction Score, respectively. Similarly, a hotel can identify the best and the worst hotels in the market as well as its rank from various perspectives. This enables the hotel to diagnose its service quality clearly and effectively.

As mentioned in Section 2, however, Positive Rate and Satisfaction Score are insufficient to compare attributes with each other. The results of room environment and location are one example showing why Integrated Index is necessary. For the room environment and location, Hotel A is given 47% and 37% of Positive Rates, respectively. One may argue that the location has a lower score, so Hotel A needs to improve the location than the room environment. However, the location is actually above the market average, while the room environment is below the average, which implies the room environment is actually in a worse position. As shown in Table 3, Integrated Index can address such an issue. For Hotel A, the Integrated Index scores of the room environment and the location are 59 and 99, respectively, revealing the relative strength of the location.

TABLE 3. Main results

	Positive Rate									Satisfaction Score									Integrated Index								
	Room environment	Service	Basic facility	Dining	Location	Auxiliary facility	Furniture/appliances	Bathroom	Price	Room environment	Service	Basic facility	Dining	Location	Auxiliary facility	Furniture/appliances	Bathroom	Price	Room environment	Service	Basic facility	Dining	Location	Auxiliary facility	Furniture/appliances	Bathroom	Price
<b>A</b>	47	<u>70</u>	39	34	<u>37</u>	18	10	6	6	88	94	94	95	98	88	93	<u>92</u>	68	59	92	89	71	<u>99</u>	54	72	73	33
<b>B</b>	38	59	39	44	29	31	15	6	3	81	95	90	95	98	85	91	74	78	34	77	73	<u>91</u>	83	73	88	61	29
<b>C</b>	<u>63</u>	39	39	30	27	16	12	5	7	86	77	82	78	<u>98</u>	65	80	66	57	81	5	50	23	79	12	55	53	23
<b>D</b>	<u>71</u>	57	39	18	15	27	11	8	5	<u>90</u>	90	<u>96</u>	84	91	92	89	73	72	<u>100</u>	62	<u>95</u>	13	40	<u>78</u>	65	66	29
<b>E</b>	58	57	39	33	15	16	12	3	4	79	98	92	97	78	81	80	52	74	60	78	80	75	17	38	53	37	27
<b>F</b>	50	50	38	37	26	17	5	3	3	83	91	84	91	94	76	82	59	67	54	53	52	67	69	33	26	41	19
<b>G</b>	56	48	38	<u>44</u>	12	28	<u>16</u>	<u>13</u>	<u>14</u>	82	90	81	86	73	88	89	86	54	62	48	42	69	0	73	<u>88</u>	<u>97</u>	<u>54</u>
<b>H</b>	49	66	35	27	23	10	9	1	3	86	<u>98</u>	91	<u>98</u>	98	<u>95</u>	<u>95</u>	67	<u>100</u>	59	<u>94</u>	40	68	72	50	69	37	50
<b>I</b>	46	41	23	21	18	18	8	1	5	63	75	83	80	83	67	68	6	64	12	3	4	10	32	19	14	0	26
<b>J</b>	50	54	<u>41</u>	43	25	<u>35</u>	9	3	6	71	79	84	87	86	77	74	36	50	33	32	59	70	52	71	29	25	13
<b>Max.</b>	71	70	41	44	37	35	16	13	14	90	98	96	98	98	95	95	92	100	100	94	95	91	99	78	88	97	54
<b>Min.</b>	38	39	23	18	12	10	5	1	3	63	75	81	78	73	65	68	6	50	12	3	4	10	0	12	14	0	13
<b>Avg.</b>	53	54	37	33	23	22	11	5	6	81	89	88	89	90	81	84	61	68	55	54	58	56	54	50	56	49	30
<b>Best</b>	D	A	J	G	A	J	G	G	G	D	H	D	H	C	H	H	A	H	D	H	D	B	A	D	G	G	G

By comparing the Integrated Index scores of individual attributes, a hotel can easily identify its competitive advantages in the market. Taking Hotel A as an example, the results in Table 3 and Figure 2 show that it has the greatest competitive advantage in terms of the location. The service, basic facility, and bathroom also show good performance, taking the second place in the market. In contrast, the room condition and auxiliary facility are revealed as weaknesses of the hotel. Considering the high reference rate of the

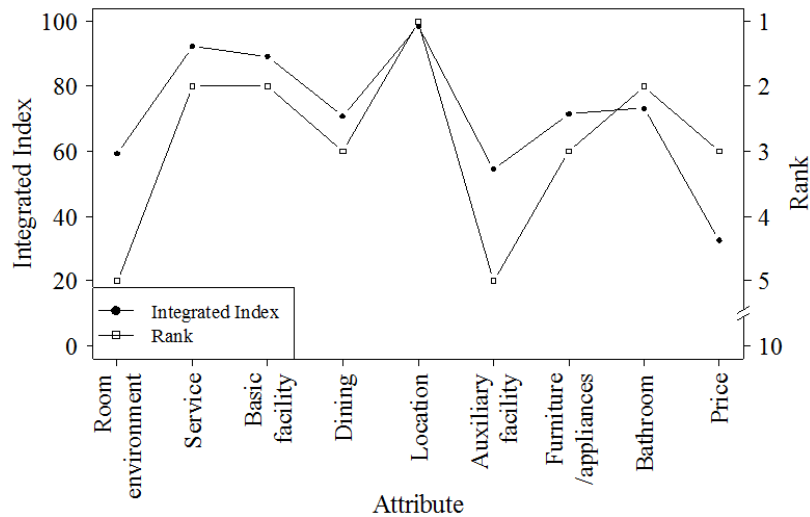


FIGURE 2. Integrated Index scores and the ranks in the market: Hotel A

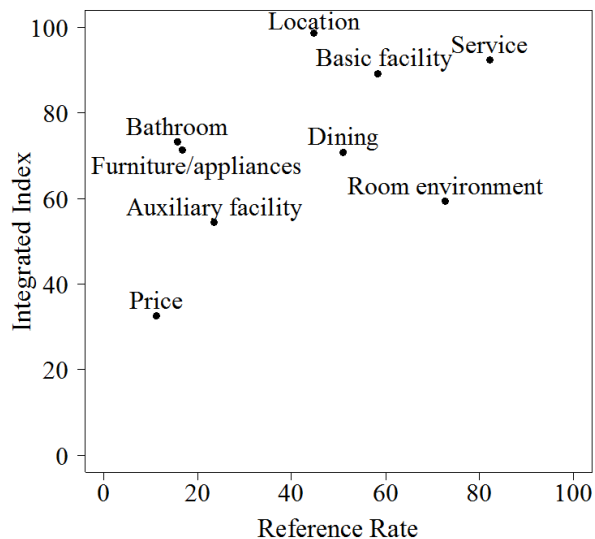


FIGURE 3. Scatter plot of reference rate and Integrated Index scores and the ranks in the market: Hotel A

room environment (see Table 2 and Figure 3), it can be inferred that the room environment is of great importance to customers, and thus, further improvement seems necessary and urgent.

**4. Conclusions.** This paper presented how opinion mining can contribute to a competitive analysis of hotel services. Three comparison indices were suggested: Positive Rate, Satisfaction Score, and Integrated Index. To demonstrate, the competitive analysis was performed for 10 5-star hotels in Seoul and their competitive advantages were evaluated and discussed for nine different service attributes. The proposed analysis enables a hotel to identify its relative performance and rank in the market at the attribute level and helps understand its strengths and weaknesses compared to the competitors. Such information can be useful for service quality management and benchmarking.

Although the current study was conducted for hotel services, the proposed approach and the comparison indices are expected to be applicable to many other cases. One limitation of the current approach, however, is that the strength of sentiments was not considered in the opinion mining. Making the opinion mining model incorporate sentiment strengths is

possible future work. Analyzing sentiments by customer type is another line of potential research.

**Acknowledgment.** This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (Ministry of Science and ICT, MSIT) (No. NRF-2019R1F1A1041099). The authors also gratefully acknowledge the helpful comments and suggestions of the reviewers, which have improved the presentation.

## REFERENCES

- [1] J. Jin, Y. Liu, P. Ji and C. K. Kwong, Review on recent advances in information mining from big consumer opinion data for product design, *Journal of Computing and Information Science in Engineering*, vol.19, no.1, pp.010801.1-010801.19, 2019.
- [2] B. Liu, *Sentiment Analysis and Opinion Mining*, Morgan & Claypool Publishers, 2012.
- [3] K. Khan, B. Baharudin, A. Khan and A. Ullah, Mining opinion components from unstructured reviews: A review, *Journal of King Saud University – Computer and Information Sciences*, vol.26, no.3, pp.258-275, 2014.
- [4] Y. Kim, R. Dwivedi, J. Zhang and S. R. Jeong, Competitive intelligence in social media Twitter: iPhone 6 vs. Galaxy S5, *Online Information Review*, vol.40, no.1, pp.42-61, 2016.
- [5] E. Björkelund, T. H. Burnett and K. Nørvåg, A study of opinion mining and visualization of hotel reviews, *ACM International Conference Proceeding Series*, pp.229-238, 2012.
- [6] D. Gräbner, M. Zanker, G. Fliedl and M. Fuchs, Classification of customer reviews based on sentiment analysis, *Information and Communication Technologies in Tourism*, vol.13, no.10, pp.460-470, 2012.
- [7] C. A. Hargreaves, Analysis of hotel guest satisfaction ratings and reviews: An application in Singapore, *American Journal of Marketing Research*, vol.1, no.4, pp.208-214, 2015.
- [8] X. Tian, W. He, R. Tao and V. Akula, Mining online hotel reviews: A case study from hotels in China, *AMCIS 2016: Surfing the IT Innovation Wave – The 22nd Americas Conference on Information Systems*, 2016.
- [9] A. Valdivia, M. V. Luzon and F. Herrera, Sentiment analysis in TripAdvisor, *IEEE Intelligent Systems*, vol.32, no.4, pp.72-77, 2017.
- [10] J. W. Bi, Y. Liu, Z. P. Fan and J. Zhang, Wisdom of crowds: Conducting importance-performance analysis (IPA) through online reviews, *Tourism Management*, vol.70, pp.460-478, 2019.
- [11] M. Kwak, J. Choi and S. Park, Sentiment analysis of online reviews for evaluating customer satisfaction on hotel service attributes, *Tourism Management Research Organization*, vol.90, no.4, pp.1-25, 2019.
- [12] J. Choi, S. Park, S. Kim, G. Lee and M. Kwak, Key attributes of hotel service quality in Korea: An analysis based on online review mining, *Tourism Management Research Organization*, vol.84, no.5, pp.1069-1093, 2018.
- [13] H. Kong and E. Song, A study on hotel customer reputation analysis based on big data, *Journal of Digital Contents Society*, vol.15, no.2, pp.219-225, 2014.