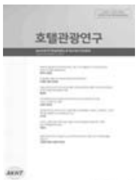
한국호텔관광학회 **호텔관광연구** 2020 제22권 제3호 (통권84호) **pp. 123-136** https://doi.org/10.31667/jhts.2020.9.84.123

www.akht.or.kr

Journal of Hospitality & Tourism Studies

The Academy of Korea Hospitality & Tourism



Big data analysis of online reviews of hotel service quality: solo travelers vs. non-solo travelers

호텔 서비스 평가에 대한 온라인 리뷰 빅데이터 분석: 솔로 여행자와 동반 여행자의 비교

조동석* · 리드얼** · 이해영***
Cho, Dong Seok · Reid, Earl L. · Lee, Hae Young

* 경성대학교 호텔관광외식경영학부 강사, chaosddadda@gmail.com, 관심분야: 호텔관광경영, 빅데이터마케팅
** 경성대학교 호텔관광외식경영학부 조교수, elreidjr@gmail.com, 관심분야: 호텔관광마케팅경영, 온라인마케팅
*** 경성대학교 호텔관광외식경영학부 부교수, hylee1228@ks.ac.kr, 관심분야: 여행항공경영, 호텔관광마케팅(교신저자)

요약

솔로 여행은 1인 가구의 증가로 인해 보편화 되고 있는 새로운 여행현상으로서 여행업계에 새로운 활력을 불어 넣고 있다. 솔로 여행의 특징 중 하나는 여행자의 디지털 콘텐츠 활용도가 매우 높다는 것인데, 이런 측면에서 소셜 미디어를 통해 얻어지는 다양한 정보들은 솔로 여행 현상을 이해하는 데 매우 중요한 요소라 할 수 있다. 이에 본 연구는 온라인 호텔 리뷰 빅데이터 분석을 실시함으로써, 솔로 여행자들의 호텔 서비스에 대한 종합적인 평가를 확인하고 필요한 관리적 아이디어를 얻고자 하였다. 구체적으로는 여행자 감정에 대한 빅데이터 주요 키워드와 분류된 선택 속성의 효과를 솔로 여행자와 동반 여행자 간에 비교하였다. 연구결과를 살펴보면, '객실', '직원', '좋음', '위치', '깨끗함' 이 빈출도가 가장 높은 상위 5개의 단어로 나타났다. 호텔 선택 속성 요인 및 텍스트 키워드 연관 규칙은 솔로 여행자와 동반 여행자 간에 상당한 차이가 있는 것으로 조사되었는데, 솔로 여행자들의 연관 단어들은 모두 부정적 감정과 연결되는 반면, 동반 여행자들의 연관 단어들은 긍정과 부정 감정과 혼재되어 연결되어 있었다. 또한 로지스틱 회귀 분석을 통해 여행자의 만족과 불만족에 영향을 미치는 핵심 키워드의 세부적 영향력이 보다 명확하게 드러났다. 솔로 여행자의 만족에 영향을 주는 주요 단어는 '플로어', '직원', '근처' 등으로 비교적 적었던 반면, 동반 여행자의 만족에 영향을 주는 주요 키워드는 '객실', '뷔페', '수영장', '직원', '깨끗함', '서비스', '아침식사', '음식', '비용', '지불' 등으로 좀 더 다양하게 나타났다. 더불어 연구의 후반부에는 온라인 호텔 리뷰 정보를 활용한 선제적 관리 방안과 적극적 커뮤니케이션 전략이 논의된다.

핵심용어

솔로 여행,
온라인 호텔 리뷰,
소셜 빅데이터,
연관분석

ABSTRACT

Individualized activities, caused by the increase of single-person households, have now materialized in travel, leading to a rise in the number of solo travelers. One of the characteristics of solo travel is the solo traveler's high usage of digital content. In this respect, various types of information obtained through social media can be used as a critical element in understanding the phenomenon of solo travel. Therefore, this study attempted to confirm the comprehensive evaluation of hotel services of solo travelers, in particular, and obtain necessary management ideas by conducting an online hotel review big data analysis. Specifically, the differential effect of categorized attributes and important text derived from big data analysis of traveler emotions were compared between solo travelers and non-solo travelers. Results showed that 'room', 'staff', 'good', 'locate', and 'clean' are the keywords that travelers most often use in hotel reviews. Traveler selection attributes and association rules showed significant differences between the two traveler groups. The entire set of associated texts of solo travelers were linked to negative emotions, while the association words of non-solo travelers were mixed with positive and negative emotions. In addition, a logistic regression analysis more clearly revealed the critical keywords that affect the satisfaction and dissatisfaction of travelers. Texts that influence the satisfaction of solo travelers were relatively few, such as 'floor', 'staff', and 'nearby', while the main keywords affecting the satisfaction of non-solo travelers appeared more diverse such as 'room', 'buffet', 'pool', 'staff', 'clean', 'service', 'breakfast', 'food', 'expense', and 'pay'. Preemptive management plans and communication strategies using social big data such as online hotel reviews are discussed from various perspectives.

KEYWORDS

solo travel,
online hotel review,
social big data,
association analysis

ACCEPTANCE INFO.

2020.09.08. 원고 접수
2020.09.23. 1차 수정본 접수
2020.09.27. 2차 수정본 접수
2020.09.28. 최종수정본 접수
2020.09.28. 최종 게재 확정
3인 익명심사필

ACKNOWLEDGEMENTS

This paper is a revised version of Dong Seok Cho(2018)'s doctoral dissertation, and an earlier version of this research was presented at the 50th KASTLE Conference 2018.

I. Introduction

The increase of solo travelers due to various social changes is a global trend in the travel context (Heimtun & Abelsen, 2013; Solo Traveler, 2019). An active, aging population, the phenomena of childless couples and later marriage, and especially a rising population of singles have combined to produce substantial changes in travel and leisure demands (Allahviranloo & Axhausen, 2018). As the social atmosphere has become more favorable for solo travel, more travelers are embarking on solo journeys. Solo travel accounted for 18% of worldwide reservations and increased by 7% in 2018 (PhocusWire, 2019). One study even predicts that solo travel will be the leading travel category in the post-COVID-19 era (Travel Weekly, 2020).

With regard to solo travel, the importance of online review data generated by travelers is noteworthy (Ahani et al., 2019; Cho & Lee, 2018). The specific reason is that social media developed review data, featuring a broader expression of the daily activities of solo travelers that contain more genuine traveler information, is necessary to understand the new solo travel phenomenon (Gretzel & Yoo, 2008; Lee et al., 2019; Lee & Ko, 2019; Sparks et al., 2013). When preparing an actual trip, solo travelers refer to other travelers' reviews of hotels, restaurants, or tourist attractions. Such social generated reviews (social reviews) are considered more reliable information sources than commercial information as they are based on personal experiences of actual travelers (Chang et al., 2019; Kang & Moon, 2019; Xia & Bechwati, 2008).

However, despite the practical value of review data related to solo travel, such information has not been fully exploited in previous studies. Indeed, limited attempts using social big data analysis techniques have been made to explore various forms of information such as solo traveler's interests, preferences, or values embedded in reviews. The role of online reviews as a travel reference source has been repeatedly confirmed through the responses of travelers to reviews (Kang & Moon, 2019; Lee & Ko, 2019; Lei & Law, 2015; Nicoli & Papadopoulou, 2017). Most studies, however, have mainly focused on revealing the characteristics of reviews, confirming their influence on traveler behavior, or revealing differences in review influence based on traveler profiles.

Through the social big data analysis technique, therefore, this study attempts to analyze online review content produced by solo travelers in pursuit of more effective business strategies. To better ascertain the characteristics of the solo travel phenomenon, this study seeks to distinguish between solo travelers and non-solo travelers and to explain their characteristic differences (Chen et al., 2014). Specifically, by collecting hotel service textual data included in online review data, several keywords will be identified and categorized as text-based attributes. Furthermore, this study intends to identify and predict the emotional responses of travelers by clarifying the association rules between individual words, and to confirm the influence of individual words on traveler satisfaction/dissatisfaction (Lee et al., 2019; Song et al., 2014; Yang & Yang, 2015). This new approach based on social big data analysis techniques expands the scope of research and analysis in the field of hotel and tourism. Such an approach may provide more practical information, enabling a broader understanding of the solo travel phenomenon, and establishing effective corporate response strategies.

II. Literature review

1. Solo travel

Planning and decision-making activities that are done alone and not shared with others, as well as enjoyment in being independent, can be called 'solo traveling' (Bianchi, 2015). Solo travelers are a part of a separate group of travelers that refers to people who arrive in a destination country by themselves (Foo, 1999). Laesser et al. (2009) defined a solo traveler as a person who travels alone or a person from a larger household who decides to interact with society and look for something new by themselves. A solo traveler is defined in this study as a traveler who plans and enjoys traveling alone without a companion.

Solo travelers are distinguished from non-solo travelers in terms of their motivation, purpose, or values. Solo travelers are primarily highly educated and have rich travel experiences. They travel alone but tend to value social contact and have strong, adventurous personalities (Solo Traveler, 2019). By proposing the classification of solo travelers, Laesser et al. (2009) found some differences in profiles based on income, occupation, age, destination familiarity, accommodation type, and travel expenditures. Bianchi (2015) added that while solo travelers pursue freedom, relaxation, and discovery, they are sensitive to safety issues and hate unfriendly services. Ji et al. (2020) suggested that escapism motivation and risk perception are also useful in explaining solo travel behavior.

Notably, the high proportion of female travelers is another salient feature of the solo travel phenomenon (Solo Traveler, 2019). Despite various travel constraints, the number of adventure-oriented women who want to enjoy freedom through travel continues to grow. Chiang and Jogaratnam (2006) identified solo female traveler travel motivations with the terms of experience, escape, relaxation, social, and self-esteem. The importance of each motivation was based on specific traveler demographic and trip characteristics. Wilson and Little (2005) contended that solo female traveler travel constraints are grouped into socio-cultural, personal, practical, and spatial factors, with related sub-factors of social expectations, perceptions of others, and doubts and fear. Uatay et al. (2019), in an examination of the travel constraints of female solo travelers, found that solo travelers are primarily affected by intrapersonal and interpersonal constraints as opposed to structural constraints.

2. Hotel selection attributes

Hotel selection attributes are features that a guest prefers or takes into consideration when selecting a hotel. These attributes can, therefore, be employed as tools for evaluating traveler satisfaction and dissatisfaction for the hotel (Knutson, 1988). Generally, hotel selection attributes are more diverse and complex than general consumer products because hotels comprise both tangible features such as rooms and amenities, as well as intangible features such as human services (Masiero et al., 2015). Hotel selection attributes have long been considered as the determinants affecting the guest decision-making process. An understanding of hotel selection attributes better meets the needs of the guest, thereby improving the satisfaction of the guest and ultimately contributing significantly to the management performance of the hotel (Liu et al., 2017).

The detailed types of selection attributes and their roles have been discovered in various areas. Most research on hotel selection attributes focuses primarily on identifying important determinant attributes of various types of hotels (i.e., business hotel, luxury hotel, and resort hotel). For instance, Dolnicar and Otter (2003) simplified 173 attributes while reviewing existing studies (From 1984 to 2000) on hotel selection attributes. While attributes related to ‘hotel’, ‘room’, and ‘services’ were widely mentioned, ‘image’, ‘price/value’, ‘location’, ‘security’, ‘marketing’, and ‘food’ were mentioned in a limited way. Clow et al. (1994) grouped specific selection attributes into ‘safety’, ‘service quality’, ‘reputation/familiarity’, ‘physical environment’, ‘location’, and ‘price’, and classified them into tangible and non-tangible attributes. Sohrabi et al. (2012) presented a comprehensive hotel selection attribute model in which ‘hotel staff services’, ‘network services’, ‘promenade’, and ‘comfort’ were classified into the hotel comfort dimension, and ‘expenditure’, ‘security and protection’, ‘news’, and ‘recreational information’ were classified into the hotel compensation dimension.

In addition, some studies have compared the importance of selection attributes according to various demographic characteristics such as traveler purpose, travel type, nationality, and gender, to understand hotel selection attributes from the traveler's point of view. Chu and Choi (2000) compared the hotel selection attributes of business and leisure travelers.

They found that ‘service quality’, ‘business facilities’, ‘value’, ‘room and front desk’, ‘food and recreation’, and ‘security’ were the main attributes in both groups. McCleary et al. (1998) also found meaningful differences and considerable agreement regarding the importance of hotel selection attributes between US and Korean business travelers. Respondents, irrespective of nationality, placed the greatest importance on ‘cleanliness’, ‘comfort’, ‘staff’, and ‘security’ attributes. ‘Convenient location’ and ‘service availability’ were more welcomed by Korean travelers, whereas the ‘non-smoking room’ attribute was more appealing to US travelers. Crnojevac et al. (2010) explored the potential differences in a guest’s hotel selection attribute preference according to the reservation mode, such as online and offline bookings. Although the reservation mode varies depending on the nature of travel (leisure vs. business or group), the importance of the major hotel selection attributes did not differ significantly depending on the reservation mode.

3. Online hotel review and big data research

Online hotel reviews, as personal daily records, refer to hotel-use experience information exchanged by travelers in cyberspace. Such shared personal experienced-based review information is used as essential data in travelers’ preliminary information search activities (Gretzel & Yoo, 2008; Oh, 2018). Many travelers search for various travel information, such as on hotels, to reduce the risk involved in choosing travel services whose qualities are difficult to assess before purchase (Cho & Lee, 2018; Sparks et al., 2013).

Furthermore, such review information is perceived as more reliable as it is based on personal experience compared to commercial information specifically written to induce travelers (Giglio et al., 2019; Xia & Bechwati, 2008). Various evaluations, including emotions such as satisfaction or dissatisfaction with specific hotel services, are quickly and widely delivered through various platforms such as blogs, cafes, or social media sites. Hotel reviews, in particular, provide beneficial information for businesses since the reviews contain consumer evaluations. In reality, online reviews include a variety of criteria used as selection attributes that have been evaluated by travelers after using hotel services (Jang, 2011; Mudambi & Schuff, 2010). Therefore, an analysis of review contents allows for a deeper understanding of traveler interests in or needs for hotel services and provides critical information necessary for service improvement.

Recently, various big data analysis techniques have emerged, and studies have been carried out to derive selection attributes for hotels by utilizing traveler review data. Tussyadiah and Zach (2017) demonstrated that ‘location’ (proximity to the point of interest and characteristics of the neighborhood), ‘host’ (service and hospitality), and ‘property’ (facilities and atmosphere) are frequently mentioned in traveler reviews. Ukpabi et al. (2018) also used online traveler reviews to compare the hotel selection attributes in African countries. ‘Breakfast/food’, ‘room service’, and ‘security/safety’ were the main selection attributes for visitors to Egypt visitors, ‘room service’ and ‘staff’ for travelers to Kenya, and ‘room service’ and ‘location’ to for visitors to South Africa.

Some recent studies have designated TripAdvisor as a data source, containing a great deal of review information on hotel experiences. Liu et al. (2017) studied hotel selection attributes in TripAdvisor reviews written by travelers who had stayed at hotels. The selection attributes that appeared at high frequency in traveler reviews were ‘cleanliness’, ‘location’, ‘room’, ‘service’, and ‘value’. Drawing on hotel reviews of business travelers in TripAdvisor, Chang et al. (2019) found that business travelers were more generous in hotel ratings than couples and tended to rate hotels higher in winter than in summer. Business travelers used negative keywords more often, such as ‘rude’, ‘terrible’, ‘horrible’, ‘broken’ and ‘dirty’ to express unsatisfactory feelings about hotel stays. Ahani et al. (2019) also focused on spa hotel guests to analyze health-related features within TripAdvisor. Several keywords such as ‘mineral water’, ‘treatment’, ‘massage’, ‘water’, ‘swimming pool’, ‘sauna’, ‘mineral bath’, and so forth were identified through online reviews. Similarly, Giglio et al. (2019), through the non-textural elements of the hotel experience, examined that the interior features of ‘rooms’, ‘restaurants’, and ‘staff services’ are mainly expressed in travelers’ pictural information. ‘Bedroom’, 'dinner', 'bathroom', 'living room', 'restaurant', 'washbasin', 'person' were identified as the most frequent elements in TripAdvisor.

In addition to analyses that review content, other studies have attempted to verify the relationship between textual data identified in online review data and traveler satisfaction. Li et al. (2013) stated that 'transportation convenience', 'food and beverage management', 'convenience of tourist destinations', and 'value for money' are significant attributes that affect traveler satisfaction. Xiang et al. (2015) suggested that words such as 'deals', 'family friendliness', and 'core product' and 'staff' have positive impacts on traveler satisfaction ratings of Expedia hotel reviews. Additionally, Jang et al. (2018) collected hotel reviews over six years and derived important hotel attributes: 'staff' , 'room', 'service', 'space', 'view', 'reservation', 'quietness', and 'internet/WiFi'. In particular, 'staff' was the most important hotel choice attribute in the satisfaction of travelers.

III. Research Methods

1. Data collection and subject

TripAdvisor English reviews written by travelers with experience staying at hotels in Seoul were selected for the study analysis. TripAdvisor has collected and stored detailed information about many hotels from travelers around the world, possessing around 75 million traveler reviews (Kim, 2016). Among them, reviews were collected that included one or more of the words 'hotel' and 'accommodation'. Hotel-related topics were collected from traveler reviews on TripAdvisor over 60 months from January 1, 2013, through December 31, 2017. A total of 16,886 traveler reviews were collected, of which 15,392 reviews were used for the final analysis, after excluding reviews written for advertising and those deemed inappropriate. Solo travelers, the subject of data collection, represent a progressively increasing number, and present promising business opportunities. Recent surveys reported that 18% of global bookings in 2019 were made by solo travelers, up 7% from the previous year. Regardless of age, gender, or nationality, 76% of the surveyed travelers have experienced or are considering solo travel (PhocusWire, 2019, 2020). Google searches related to solo travel in 2019 also increased by 230% compared to the previous year (Industry Global News 24, 2020. 1. 14).

2. Data process

Big data processing can be divided into bottom-up and top-down processes (Inzalkar & Sharma, 2015). First, the top-down method is a process in which the ontology is developed by analyzing the theoretical background, and then the keywords of the ontology are collected and classified. Conversely, the bottom-up process is a method in which topics are collected by a web crawler and classified into a general dictionary or a user dictionary. Traveler reviews used in this study were classified as bottom-up, in which the textual data were collected by a web crawler and classified into a general dictionary or a user dictionary (Inzalkar & Sharma, 2015). Keywords extracted from text mining were grouped into four higher concepts, of 'facility', 'service', 'location and transportations', and 'price'. Further, 28 hotel emotional keywords were classified using the opinion mining technique into 'satisfaction' (e.g., good, great, nice, friendly, helpful, etc.) or 'dissatisfaction' (e.g., small, few, bit, problem, bad, etc.). Reviews designated as showing 'satisfaction' possessed positive keywords that travelers used after using a hotel, whereas 'dissatisfaction' reviews included negative keywords. Reviews with an equal number of expressions of satisfaction and dissatisfaction were classified as having neutral emotions.

3. Data analysis

The collected data were analyzed using SPSS 20.0 for Windows statistical package program and R version 3.5. First, the text mining process was utilized to extract important keywords from the entire hotel review data sample. Keyword extraction criteria were frequency and degree centrality. The WordArt program was then employed to visualize words extracted by year to investigate the interests and trends of travelers. Third, an association analysis was performed using R version 3.5 to identify the association rules between the four text-based attributes and traveler emotions. The association rules derived from the analysis were evaluated based on support, confidence, and lift (Yang & Yang, 2015). Generally, association analyses find that the interrelation of two or more words exists in a single online transaction (Song et al., 2014). Lastly, a logistic regression analysis was conducted to verify the differential effect of major keywords of categorized attributes on the satisfaction/dissatisfaction of travelers. As different travel companion types, solo travelers and non-solo travelers were analyzed separately, and results identified text-based attributes and detailed words that directly affect each group's satisfaction/dissatisfaction.

IV. Results

1. Descriptive statistics

The results of the descriptive statistics of the entire data sample used in this study are as follows. Of the total 15,392 traveler reviews collected, 9.1% (1,403 reviews) were by solo travelers, and 90.9% (13,989 reviews) were by non-solo travelers. The fact that solo traveler's reviews are fewer in number than non-solo traveler reviews seems reasonable considering that around 10% of world travelers are actually solo travelers (Hana Tour, 2017). Of all of the traveler hotel reviews collected, 12% (1,852 reviews) were from 2013, 15.5% (2,392 reviews) from 2014, 18.3% (2,811 reviews) from 2015, 28.5% (4,390 reviews) from 2016 and 25.6% (3,927 reviews) from 2017. The hotel-related emotional keywords of 'satisfaction' were found in 84.1% (12,273 reviews) of the reviews, while 'neutral' 9.2% (1,344 reviews) and 'dissatisfaction' 6.7% (976 reviews) comprised the remainder of the reviews.

<Table 1> Number of hotel review

	2013 n (%)	2014 n (%)	2015 n (%)	2016 n (%)	2017 n (%)	Total n (%)
Solo traveler	251 (1.6)	314 (2.0)	172 (1.1)	134 (0.9)	532 (3.5)	1,403 (9.1)
Non-solo traveler	1,601 (10.4)	2,078 (13.5)	2,639 (17.1)	4,256 (27.7)	3,415 (22.2)	13,989 (90.9)
Total	1,852 (12.0)	2,392 (15.5)	2,811 (18.3)	4,390 (28.5)	3,947 (25.6)	15,392 (100.0)

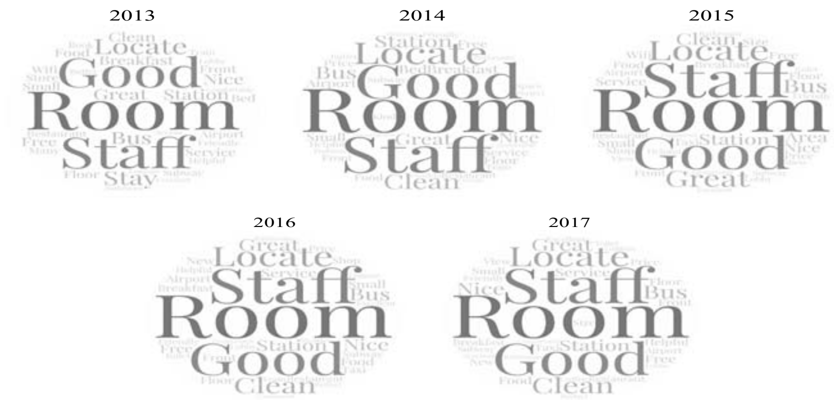
2. Text mining analysis

Text mining the collected hotel reviews resulted in a total of 15,392 words, from which the top 100 words were extracted based on frequency. After excluding words deemed unnecessary or redundant, a total of 68 words were used for final analysis. In general, the frequency and degree centrality of words are the key indicators in determining important words in a network (Kang et al., 2012). The word with the highest frequency in total data was 'room', followed by 'staff', 'good', 'locate', 'clean', 'station', 'great', 'breakfast', 'bus' and so on. The degree centrality did not exactly match the frequency, but it showed a similar ranking. Therefore, these results explain that the words mentioned above are important keywords in traveler reviews of hotels.

Figure 1 is a visualization of the words extracted from the entire review data sample based on frequency. The size of the letters to spell a word corresponds to the frequency of the word appearing in the review data: the larger the letters, the higher the frequency (Song et al., 2014). Looking at the keyword changes over the last five years, the interest of travelers changed a little. However, the most important factors for selecting a hotel seemed to have remained largely the same. From 2013 to 2017, the words that travelers most often use in hotel reviews were almost unchanged: ‘room’, ‘staff’, ‘locate’, ‘clean’, ‘airport’, and ‘station’.

<Table 2> Frequency and degree centrality of extracted words

Rank	Word	Frequency	Degree Centrality	Rank	Word	Frequency	Degree centrality
1	Room	25,271	.208	35	Toilet	1,509	.044
2	Staff	10,353	.125	36	Shuttle	1,434	.024
3	Good	9,995	.120	37	Perfect	1,273	.038
4	Locate	8,911	.104	38	Bed	1,267	.032
5	Clean	6,418	.073	39	Buffet	1,225	.033
6	Station	6,386	.073	40	Lounge	1,222	.035
7	Great	6,326	.095	41	Train	1,221	.028
8	Breakfast	5,490	.095	42	Problem	1,164	.045
9	Service	5,431	.091	43	Access	1,122	.029
10	Bus	5,009	.053	44	Quiet	1,093	.030
11	Nice	4,924	.087	45	Convenience	1,092	.032
12	Airport	4,866	.059	46	Facility	1,027	.034
13	Restaurant	4,790	.077	47	Bar	1,010	.036
14	Small	4,217	.077	48	Kind	952	.042
15	Subway	4,185	.053	49	Local	861	.031
16	Friendly	4,148	.069	50	Pool	853	.029
17	Helpful	4,075	.062	51	Money	843	.026
18	Food	3,812	.071	52	Concierge	778	.032
19	Floor	3,691	.073	53	Wonderful	767	.029
20	Comfortable	3,590	.061	54	Bad	752	.035
21	Front	3,001	.060	55	Happy	750	.029
22	Bathroom	2,615	.062	56	Pleasant	743	.027
23	Excellent	2,459	.055	57	Fantastic	718	.026
24	Price	2,212	.055	58	Love	707	.031
25	Big	1,974	.053	59	Nearby	696	.026
26	Distance	1,893	.039	60	Expense	668	.029
27	Few	1,741	.048	61	Noise	645	.027
28	Taxi	1,723	.045	62	Drink	624	.024
29	Commend	1,691	.042	63	Cold	610	.026
30	Bit	1,677	.054	64	Safe	550	.025
31	Spacious	1,674	.031	65	Pay	543	.025
32	Lobby	1,672	.054	66	Housekeep	511	.024
33	Size	1,551	.034	67	Upset	420	.021
34	WiFi	1,542	.038	68	Terrible	403	.021



<Figure 1> Keyword visualization by year

3. Association analysis

An association analysis was conducted in this study to find conditions and rules for a set of specific words occurring at the same time. The association rule with the highest confidence in the relationship between solo travelers' hotel facility attributes and hotel-related emotions was {room, lounge, pool} => {dissatisfaction}. When a traveler mentions a 'room', 'lounge', and 'pool' together through a review, their chances of being dissatisfied are 60%. Also, if all three words are included, the probability of a dissatisfied traveler is 6.33 times higher than if they are not included. In addition, the highest confidence association rule between hotel service attribute and emotion was {front, drink, safe} => {dissatisfaction}. For location and transportation keywords, {bus, distance, local} => {dissatisfaction} was found to be the association rule with the highest confidence. Finally, the association rule with the highest confidence between price attribute and emotion was {price, money} => {dissatisfaction}.

<Table 3> Association analysis of solo traveler

Variable	Rules	Support	Confidence	Lift
Facility	{room, lounge, pool} = > {dissatisfaction}	.00214	.60000	6.32932
	{restaurant, lobby, lounge} = > {dissatisfaction}	.00214	.60000	6.32932
	{room, restaurant, lobby, lounge} = > {dissatisfaction}	.00214	.60000	6.32932
	{bed, lounge, pool} = > {dissatisfaction}	.00143	.50000	5.27444
Service	{front, drink, safe} = > {dissatisfaction}	.00143	1.00000	10.54887
	{breakfast, front, drink, safe} = > {dissatisfaction}	.00143	1.00000	10.54887
	{breakfast, food, front, drink} = > {dissatisfaction}	.00143	1.00000	10.54887
	{staff, breakfast, front, WiFi, safe} = > {dissatisfaction}	.00143	1.00000	10.54887
Location and transportation	{bus, distance, local} = > {dissatisfaction}	.00214	1.00000	10.54887
	{locate, subway, train, local} = > {dissatisfaction}	.00143	1.00000	10.54887
	{locate, airport, distance, local} = > {dissatisfaction}	.00143	1.00000	10.54887
	{locate, subway, distance, local} = > {dissatisfaction}	.00143	1.00000	10.54887
Price	{price, money} = > {dissatisfaction}	.00143	.14286	1.50698
	{price, pay} = > {dissatisfaction}	.00214	.13043	1.37594
	{money} = > {dissatisfaction}	.00570	.11268	1.18861
	{price} = > {dissatisfaction}	.01711	.10435	1.10075

The association rule with the highest confidence in the relationship between non-solo travelers' hotel facility attributes and hotel-related emotions was {floor, toilet} => {dissatisfaction}. When a traveler mentions 'floor' and 'toilet' together in a review, their chances of being dissatisfied are 100%. Also, if both words are included, the probability of traveler dissatisfaction is 1.41 times higher than if they are not included. The highest confidence association rule between non-solo travelers' hotel service attributes and hotel-related emotions was {housekeep, safe} => {satisfaction}. For non-solo traveler hotel location and transportation attributes and hotel-related emotions, {taxi, train, local} => {satisfaction} had the highest confidence level. Finally, the association rule with the highest confidence between hotel price attribute and emotion was {price, money} => {satisfaction}.

<Table 4> Association analysis of non-solo traveler

Variable	Rules	Support	Confidence	Lift
Facility	{floor, toilet} = > {dissatisfaction}	.00257	1.00000	1.41487
	{room, floor, toilet} = > {dissatisfaction}	.00250	1.00000	1.43793
	{bed, buffet, pool} = > {satisfaction}	.00222	1.00000	1.25733
	{room, bed, buffet, pool} = > {satisfaction}	.00222	1.00000	1.25733
Service	{housekeep, safe} = > {satisfaction}	.00179	1.00000	1.25733
	{clean, housekeep, safe} = > {satisfaction}	.00114	1.00000	1.25733
	{staff, housekeep, safe} = > {satisfaction}	.00129	1.00000	1.25733
	{front, concierge, housekeep} = > {satisfaction}	.00122	1.00000	1.25733
Location and transportation	{taxi, train, local} = > {satisfaction}	.00107	1.00000	1.25733
	{taxi, local, nearby} = > {satisfaction}	.00107	1.00000	1.25733
	{station, taxi, train, local} = > {satisfaction}	.00100	1.00000	1.25733
	{locate, station, taxi, local} = > {satisfaction}	.00136	1.00000	1.25733
Price	{price, money} = > {satisfaction}	.00858	.85106	1.07006
	{price, pay} = > {satisfaction}	.01122	.83957	1.05562
	{expense} = > {satisfaction}	.00172	.82759	1.04055
	{money} = > {satisfaction}	.03982	.81195	1.02089

4. Logistic regression analysis

This study used logistic regression analysis to determine the hotel selection attributes that affect the satisfaction of travelers. The model that predicts the satisfaction of solo travelers by the hotel selection attribute was statistically significant (-2 Log L = 456.193, $\chi^2[40] = 68.02$, $p< .01$). Among the facility attributes, only 'floor' had a significant (Wald = 4.28, $p< .05$) effect on the satisfaction of travelers. This result is explained by a 2.75 times increase in the likelihood of a traveler being satisfied whenever 'floor' (odd ratio = 2.75) is mentioned in a traveler review. In the service attributes, only 'staff' had a significant (Wald = 14.55, $p< .001$) effect on the satisfaction of travelers. There is a 3.08 times increase in the likelihood that a traveler will be satisfied when 'staff' (odd ratio = 3.08) is mentioned in a traveler's review. Only the 'nearby' factor had a significant (Wald = 4.78, $p< .05$) effect on the satisfaction of solo travelers in location and transportation attributes. Therefore, every time 'nearby' (odd ratio = .41) is mentioned in a hotel review, the probability of a satisfied solo traveler is reduced by 0.59 times. Finally, all the keywords of the price attribute were found not to affect the satisfaction of solo travelers.

The model that predicts the satisfaction of non-solo travelers by the hotel selection attribute was also statistically significant (-2 Log L = 605.33, $\chi^2[40] = 385.39$, $p< .01$). In the facility attributes, 'room' (Wald = 47.00, $p< .001$), 'buffet' (Wald = 3.96, $p< .05$) and 'pool' (Wald = 5.91, $p< .05$) had significant effects on the satisfaction of non-solo travelers. This result shows that when 'room' (odd ratio = .53) is mentioned in a hotel review, non-solo traveler satisfaction is reduced by .47 times. On the other hand, when 'buffet' (odd ratio = 1.43) and 'pool' (odd ratio = 1.72) are mentioned in a review, their satisfaction increases by 1.43 times and 1.72 times, respectively. In the service attributes, 'staff' (Wald = 162.08, p< .001), 'clean' (Wald = 24.78, p< .001), 'service' (Wald = 18.48, p< .001), 'breakfast' (Wald = 13.08, p< .001) and 'food' (Wald = 6.71, p< .05) had significant effects on the satisfaction of non-solo travelers. Every time a traveler mentions 'staff' (odd ratio = 2.74), 'clean' (odd ratio = 1.46), 'service' (odd ratio = 1.47), 'breakfast' (odd ratio = 1.42) and 'food' (odd ratio = 1.30), non-solo traveler satisfaction increases. All keywords related to location and transportation were found to have no effect on the satisfaction of non-solo travelers. 'Pay' and 'expense' appeared to have significant (Wald = 24.18, p< .001) effects on the satisfaction of non-solo travelers in price attributes. Therefore, every time 'pay' (odd ratio = .49) and 'expense' (odd ratio = .82) is mentioned in a hotel review, the probability of non-solo traveler satisfaction is reduced by .51 times and .18 times, respectively.

<Table 5> Logistic regression results of solo traveler

	Variable	B	S.E.	Wald	Exp(B)	p
Facility	Restaurant	.41	.35	1.37	1.50	.24
	Floor	1.01	.49	4.28	2.75	.04
	Lobby	-.59	.47	1.52	.56	.22
	Bed	.42	.38	1.21	1.52	.27
	Size	.93	.64	2.11	2.52	.15
Service	Staff	1.13	.30	14.55	3.08	.00
	Clean	-.21	.27	.59	.81	.44
	Breakfast	.35	.40	.74	1.41	.39
	Food	.51	.44	1.36	1.66	.24
	WiFi	.79	.65	1.50	2.21	.22
Location and transportation	Locate	-.44	.31	1.99	.65	.16
	Airport	-.45	.37	1.54	.64	.21
	Access	1.36	.75	3.28	3.89	.07
	Local	-.79	.46	2.93	.45	.09
	Nearby	-.88	.40	4.78	.41	.03
Price	Price	.09	.38	.06	1.09	.81
	Money	-.93	.49	3.58	.39	.06
	Expense	17.91	40192.97	.00	59932532.93	1.00
	Pay	1.32	1.05	1.59	3.75	.21
	-	-	-	-	-	-
-2 Log L = 456.193, $\chi^2[40]$ = 68.02, p< .01						

Note: Dependent variable: Satisfaction (0 = dissatisfaction, 1 = satisfaction)

<Table 6> Logistic regression results of non-solo traveler

	Variable	B	S.E.	Wald	Exp(B)	p
Facility	Room	-.63	.09	47.00	.53	.00
	Lobby	-.21	.12	3.10	.81	.08
	Buffet	.36	.18	3.96	1.43	.05
	Lounge	.39	.20	3.61	1.47	.06
	Pool	.54	.22	5.91	1.72	.02
Service	Staff	1.01	.08	162.08	2.74	.00
	Clean	.38	.08	24.78	1.46	.00
	Service	.39	.09	18.48	1.47	.00
	Breakfast	.35	.10	13.08	1.42	.00
	Food	.26	.10	6.71	1.30	.01
Location and transportation	Bus	.10	.89	1.17	1.10	.28
	Distance	.14	.12	1.33	1.15	.25
	Taxi	.24	.16	2.41	1.27	.12
	Local	-.13	.16	.60	.88	.44
	Nearby	-.18	.12	2.10	.84	.15
Price	Price	-.03	.10	.09	.97	.77
	Money	.06	.17	.13	1.06	.72
	Expense	-.20	.64	.10	.82	.00
	Pay	-.72	.15	24.18	.49	.00
	-	-	-	-	-	-
-2 Log L = 605.33, χ^2[40]=385.39, p<.01						

Note: Dependent variable: Satisfaction (0 = dissatisfaction, 1 = satisfaction)

V. Discussion

Solo travel is among the fastest-growing segment, driven by shifts in social structures as well as lifestyles (Solo Traveler, 2019). Based on online review social big data analysis, this study examined the overall evaluation of hotel services by solo travelers. Specifically, by focusing on various keywords derived from online reviews, this study attempted to categorize hotel selection attributes and to clarify the rules of association between words that influence traveler emotions. Furthermore, the relative influence of detailed word variables in text-based attributes on traveler satisfaction or dissatisfaction was compared between solo travelers and non-solo travelers. This study expanded the paradigm of data analysis and broadened the scope of academic understanding related to solo travel through a comprehensive evaluation centered on unstructured textual data of solo traveler responses to hotel services. In particular, the comparison of the association rules and the effects of keywords related to satisfaction, by traveler type (solo traveler vs. non-solo traveler), is expected to provide useful information to practitioners seeking to create new travel demands in the future through segmented service design and provisions.

Results of the text mining process revealed 68 keywords extracted based on frequency and degree centrality. The word with the highest frequency in the total data sample was 'room', followed by 'staff', 'good', 'locate', 'clean', and 'station'. Regarding the association rules, there was a difference in the association rules of detailed keywords for overall emotions between solo travelers and non-solo travelers. Solo traveler association rules were mainly related to satisfaction, while non-solo traveler association rules were linked to dissatisfaction. Such results imply that compared to non-solo travelers, solo travelers use online reviews as a place to communicate negative hotel service experiences.

Therefore, if a hotel improves the level of attributes included in the association rules, it would be a good way to reduce the dissatisfaction of solo travelers. Above all, it is necessary to periodically monitor social information on social media (i.e., blog, cafe, Facebook, etc.) and to communicate with travelers actively. Further, clarification of the contextual situation through related rules can be a prerequisite for reorganizing the service levels of each specific situation.

Logistic regression analysis showed more clearly the important keywords that affect the satisfaction of travelers. Major keywords that affect solo traveler satisfaction were relatively limited, and the price attribute had no influence. In contrast, the influence of the words was remarkable in factors of facility, service, and price factors for non-solo travelers. Therefore, the establishment of a segmentation strategy for each traveler type can further refine each traveler's specific needs, and the causes of satisfaction and dissatisfaction should be identified for a more competitive service. Specifically, for solo travelers, managers must design clean, sensible floors and train employees to maintain a friendly attitude. In addition to providing a space for travelers to stay, a hotel can also improve traveler satisfaction by recommending nearby attractions and providing transportation. Managers can also improve the satisfaction of non-solo travelers by providing hotel facilities (i.e., buffet and pool) and high-quality human services through staff training. The interior and convenience of rooms and the introduction of a convenient payment system (i.e., mobile payment) need to be carefully designed to reduce the dissatisfaction of non-solo travelers.

Despite some practical insight, this study is limited due to generalization as the scope of this study being confined to Seoul. Depending on the nature of the country or tourist destination, there may be differences in hotel selection attributes and their roles. Therefore, regional expansion or cross-national comparisons may be considered in future studies. Second, hotel service selection attributes in this study were generally compared without subdividing solo travelers and non-solo travelers by their travel characteristics such as preference, motivation, or mode. Future research to understand differences in evaluation processes through detailed traveler type comparison can further refine travelers' responses to hotel services and provide more practical information. Lastly, this study did not distinguish hotels by type or by grade. Given that traveler evaluation toward hotels may vary depending on hotel star ratings and classifications, future research can provide more useful information by analyzing traveler reviews considering various types of hotels.

REFERENCES

Ahani, A., Nilashi, M., Ibrahim, O., Sanzogni, L., & Weaven, S.(2019). Market segmentation and travel choice prediction in spa hotels through TripAdvisor's online reviews. *International Journal of Hospitality Management,* 80(July), 52-77.

Allahviranloo, M., & Axhausen, K.(2018). An optimization model to measure utility of joint and solo activities. *Transportation Research Part B: Methodological,* 108(February), 172-187.

Bianchi, C.(2015). Solo holiday travellers: Motivations and drivers of satisfaction and dissatisfaction. *International Journal of Tourism Research,* 18(2), 197-208.

Chang, Y. C., Ku, C. H., & Chen, C. H.(2019). Social media analytics: extracting and visualizing Hilton hotel ratings and reviews from TripAdvisor. *International Journal of Information Management,* 48(October), 263-279.

Chen, M., Mao, S., & Liu, Y.(2014). Big data: a survey. *Mobile Networks and Applications,* 19(2), 171-209.

Chiang, C. Y., & Jogaratnam, G.(2006). Why do women travel solo for purposes of leisure? *Journal of Vacation Marketing,* 12(1), 59-70.

Cho, D. S., & Lee, H. Y.(2018). Predictive models for solo traveler's online hotel review using big data analysis. Paper presented at the 50th KASTLE Conference, Busan, Korea.

Chu, R. K., & Choi, T.(2000). An importance-performance analysis of hotel selection factors in the Hong Kong hotel industry: a comparison of business and leisure travellers. *Tourism Management,* 21(4), 363-377.

Clow, K. E., Garretson, J. A., & Kurtz, D. L.(1994). An exploratory study into the purchase decision process used by leisure travelers in hotel selection. *Journal of Hospitality & Leisure Marketing,* 2(4), 53-72.

Crnojevac, I. H.,, Gugić, J., & Karlovčan, S.(2010). eTourism: a comparison of online and offline bookings and the importance of hotel attributes. *Journal of Information and Organizational Sciences,* 34(1), 41-54.

Dolnicar, S., & Otter, T.(2003). Which Hotel attributes Matter? a review of previous and a framework for future research. Retrieved March 13, 2016, from https://ro.uow.edu.au/commpapers/268

Foo, L. M.(1999). International solo travelers to Australia. *Bureau of Tourism Research Report,* 1(2), 1-5.

Giglio, S., Pantano, E., Bilotta, E., & Melewar, T. C.(2019). Branding luxury hotels: evidence from the analysis of consumers' "big" visual data on TripAdvisor. *Journal of Business Research.* Doi: 10.1016/j.jbusres.2019.10.053

Gretzel, U., & Yoo, K.(2008). Use and impact of online travel reviews. In P. O'Connor, W. Hopken, & U. Gretzel (Eds.), *Information and Communication Technologies in Tourism 2008* (pp. 35-46). Vienna, Austria: Springer-Verlag Wien.

Hana Tour(2017). 2017 Travel trends. Retrieved February 26, 2017, from: http://naver.me/GyflswR2

Heimtun, B., & Abelsen, B.(2013). Singles and solo travel: gender and type of holiday. *Tourism Culture & Communication,* 13(3), 161-174.

Industry Global News 24(2020). Over 85% of solo travelers are women-adventure travel reports. Retrieved July 10, 2020, from: https://industryglobalnews24.com/over-85-of-solo-travelers-are-women-adventure- travel-reports

Inzalkar, S. M., & Sharma, J.(2015). A survey on text mining-techniques and application. *International Journal of Research in Science & Engineering,* 24, 1-14.

Jang, D. S.(2011). Effects of brand reputation, positive and negative reviews, and gender on online hotel reviews: perspectives of US consumers for Korean hotels. *Northeast Asia Tourism Research,* 7(2), 215-231.

Jang, S., Liu, T., Kang, J. H., & Yang, H.(2018). Understanding important hotel attributes from the consumer perspective over time. *Australasian Marketing Journal,* 26(1), 23-30.

Ji, Y., Byun, U., & Park, M.(2020). A study on the intention of solo traveling in single consumer: using extended theory of planned behavior. *Journal of Tourism & Leisure Research,* 32(5), 149-166.

Kang, M. M., Kim, S. L., & Park, S. M.(2012). Analysis and utilization of big data. *Communications of the Korean Institute of Information Scientists and Engineers,* 30(6), 25-32.

Kang, S., & Moon, H.(2019). The study on hotel assessment of inbound tourists in Pyongyang, North Korea: based on TripAdvisor reviews. *Journal of Hospitality & Tourism Studies,* 21(2), 75-89.

Kim, Y. P.(2016). An importance-performance analysis of hotel selection attributes by using big data: comparison of deluxe and mid-price hotel. Catholic University of Daegu, Ph.D dissertation.

Knutson, B. J.(1988). Frequent travelers: making them happy and bringing them back. *Cornell Hotel and Restaurant Administration Quarterly,* 29(1), 82-87.

Laesser, C., Beritelli, P., & Bieger, T.(2009). Solo travel: explorative insights from a mature market. *Journal of Vacation Marketing,* 15(3), 217-227.

Lee, H. J., & Ko, J. Y.(2019). A netnography analysis on foreign customer's online review of Michelin restaurant in Seoul. *Journal of Hospitality & Tourism Studies,* 21(3), 189-205.

Lee, S. A., Cho, D. S., & Lee, H. Y.(2019). Prediction and estimation in location-based travel service using social big data. *Journal of Hospitality & Tourism Studies,* 21(3), 1-17.

Lei, S., & Law, R.(2015). Content analysis of TripAdvisor reviews on restaurants: a case study of Macau. *Journal of Tourism,* 16(1), 17-28.

Li, H., Ye, Q., & Law, R.(2013). Determinants of customer satisfaction in the hotel industry: an application of online review analysis. *Asia Pacific Journal of Tourism Research,* 18(7), 784-802.

Liu, Y., Teichert, T., Rossi, M., Li, H., & Hu, F.(2017). Big data for big insights: investigating language-specific drivers of hotel satisfaction with 412,784 user-generated reviews. *Tourism Management,* 59(April), 554-563.

Masiero, L., Heo, C. Y., & Pan, B.(2015). Determining guests' willingness to pay for hotel room attributes with a discrete choice model. *International Journal of Hospitality Management,* 49(August), 117-124.

McCleary, K. W., Choi, B. M., & Weaver, P. A.(1998). A comparison of hotel selection criteria between US and Korean business travelers. *Journal of Hospitality & Tourism Research,* 22(1), 25-38.

Mudambi, S. M., & Schuff, D.(2010). What makes a helpful online review: a study of customer reviews on Amazon.com. *MIS Quarterly,* 34(1), 185-200.

Nicoli, N., & Papadopoulou, E.(2017). TripAdvisor and reputation: a case study of the hotel industry in Cyprus. *EuroMed Journal of Business,* 12(3), 316-334.

Oh, H. C.(2018). Semantic network analysis of foreign visitors' online review of Seoul business hotel by TripAdvisor's big data. *Northeast Asia Tourism Research,* 14(1), 143-166.

PhocusWire(2019). What billions of GDS data signals tells us about traveler booking behavior. Retrieved July 1, 2020, from: https://www.phocuswire.com/Travel-booking-behavior-data-signals

PhocusWire(2020). Solo travel and its promising business opportunity. Retrieved July 11, 2020, from: https://www.phocuswire.com/Solo-travel-deep-dive-Klook

Sohrabi, B., Vanani, I. R., Tahmasebipur, K., & Fazli, S.(2012). An exploratory analysis of hotel selection factors: a comprehensive survey of Tehran hotels. *International Journal of Hospitality Management,* 31(1), 96-106.

Solo Traveler(2019). Solo travel statistics and data: 2019-2020. Retrieved January 8, 2020, from https://solotravelerworld.com/about/solo-travel-statistics-data/

Song, T. M., Song, J., An, J. Y., Hayman, L. L., & Woo, J. M.(2014). Psychological and social factors affecting Internet searches on suicide in Korea: a big data analysis of Google search trends. *Yonsei Medical Journal,* 55(1), 254-263.

Sparks, B. A., Perkins, H. E., & Buckley, R.(2013). Online travel reviews as persuasive communication: the effects of content type, source, and certification logos on consumer behavior. *Tourism Management,* 39(December), 1-9.

Travel Weekly(2020). Solo travel is the second-most popular category for post-lockdown trips. Retrieved July 13, 2020, from https://www.travelweekly.co.uk/articles/378221/solo-travel-is-the-second-most-popular-category-for-post-lockdown-trips-says-cox-kings

Tussyadiah, I. P., & Zach, F.(2017). Identifying salient attributes of peer-to-peer accommodation experience. *Journal of Travel & Tourism Marketing,* 34(5), 636-652.

Uatay, G., Reid, E. L., & Lee, H. Y.(2019). The impact of female travelers' travel constraints on attitude toward solo travel and travel intention. *Culinary Science & Hospitality Research,* 25(9), 102-110.

Ukpabi, D., Olaleye, S., Mogaji, E., & Karjaluoto, H.(2018). Insights into online reviews of hotel service attributes: a cross-national study of selected countries in Africa. Paper presented at the International Conference in Jönköping, January, 24-26.

Wilson, E., & Little, D. E.(2005). A relative escape? the impact of constraints on women who travel solo. *Tourism Review International,* 9(2), 155-175.

Xia, L., & Bechwati, N. N.(2008). Word of mouse: the role of cognitive personalization in online consumer reviews. *Journal of interactive Advertising,* 9(1), 3-13.

Xiang, Z., Schwartz, Z., Gerdes, J. H., & Uysal, M.(2015). What can big data and text analytics tell us about hotel guest experience and satisfaction? *International Journal of Hospitality Management,* 44(January), 120-130.

Yang, H., & Yang, C. C.(2015). Using health-consumer-contributed data to detect adverse drug reactions by association mining with temporal analysis. *ACM Transactions on Intelligent Systems and Technology,* 6(4), 55-84.