

## Research Article

# The Identification of Marketing Performance Using Text Mining of Airline Review Data

Jae-Won Hong<sup>1</sup> and Seung-Bae Park<sup>2</sup> 

<sup>1</sup>Department of Global Trade, Gyeongnam National University of Science and Technology, 33 Dongjin-ro, Jinju-si, Gyeongnam 52725, Republic of Korea

<sup>2</sup>Department of Industrial Management, Seoil University, 28 Youngmasan-ro 90-gil, Jungnang-gu, Seoul, Republic of Korea

Correspondence should be addressed to Seung-Bae Park; [sbpark@seoil.ac.kr](mailto:sbpark@seoil.ac.kr)

Received 1 October 2018; Accepted 13 November 2018; Published 2 January 2019

Guest Editor: Jaegeol Yim

Copyright © 2019 Jae-Won Hong and Seung-Bae Park. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

We aim firstly to extract major keywords using text mining method, secondly to identify prominent keyword from the keywords extracted from text mining analysis, and then to confirm differences in influences of the keywords which affect corporate performance. Results were as following. First, keywords have been found to show distinctive features. Since the keywords posted from the clients showed certain tendency, airlines accordingly need service management by identifying the service property through keyword analysis. Second, prominent keywords have been found out of the keyword extracted from text mining. Some of the keywords have significantly correlated with marketing performance, but others not. This implies that the company could uncover consumers' needs through the prominent keywords and managing the properties related to the prominent keywords would help with improving corporate performance. Third, "recommend" should be treated distinctively with "satisfaction" in terms of service management through the keywords. Results suggest strategic implications to the practical business environment by analyzing keywords around the industry using text mining. We believe this work, which aims to establish common ground for understanding these analyses across multiple disciplinary perspectives, will encourage further research and development of service industry.

## 1. Introduction

With rapid development of Internet environment and mobile devices, there have been inevitable changes in way of consumers' communication. Consumers form their own community connected with SNS, which refers to social networking service such as Facebook, Twitter, blog, etc. And those network services are treated as reliable message sources over the mass media. As social networks have become increasingly influential, and consumers exploit much WOM information through the social network, the information of online communication from social network plays a critical role in purchase decision making [1]. Moreover, when it is said that traditional WOM communication is implicit or personal, online WOM is more open and collective. Due to the characteristics of the social network, reach of the information is more extensive and the speed of the information

is also even faster. Importance of active responses to consumers' online WOM, therefore, increases in government and social organization as well as corporations.

In this context, online review that consumers produce on SNS can be influential means to grasp a tendency of consumers WOM and recommendation. Consumers, for instance, even read product review to buy a book, which belongs to the section of experiential products. Preceding researches on online review data mainly hire survey method or text mining method [2]. Survey method is used since it gets direct and clear opinions from consumers, but is limited to collect a variety of consumers' opinions and latent opinions. On the other hand, using the text mining method, the limitation of survey method can be overcome by collecting practical opinions from consumers, and through refining data sourced from online, it can also help with deducing consumers' underlying latent opinions as well.

Text mining, on the other hand, is a method of extracting unknown and valuable information from randomly organized text data [3]. Thus, it is described as an automated tool that extracts undisclosed information from text data which are of unstructured format such as mail, reviews, web documents, video clips, or images [4]. Recently, major statistical packages and data mining programs include text mining function to facilitate or simplify the analysis process of those kinds of unstructured data. They provide with function of preprocess to conduct text mining and function of summarizing and categorizing to identify the pattern of the data, but also provides with a variety analysis function of associate analysis, cluster analysis, etc.

It is important to reduce data in handling big data. Large data are not necessarily required to discover important implications. The important thing is the appropriate data. The academic significance of our research is to derive ways to reduce to appropriate data. Current text mining method tends to complement these limitations, accordingly useful to deduce managerial implication on practical decision-making [5]. In this context, by firstly deducing major variables based on the keywords extracted by text mining to reduce large data that is not necessary to discover important implication, and secondly combining them with items in questionnaire, this research will help with suggesting practical implication in the context of aviation industry.

## 2. Related Works

*2.1. Text Mining.* Text mining refers to automated methods that extract undiscovered and valuable information from unstructured text by categorizing or structuralizing the text [6]. By extracting information from big data in a variety of field, connectivity within information will be uncovered. This overcomes limitations occurred by simple data analysis and enables to identify underlying meanings from massive text data. In this point, the importance of these methods increases in terms of that the method can be utilized for suggesting practical future strategies.

Through the method of text mining, researchers would take advantage of not only extracting concepts of the text, but also identifying relationship with other concepts and visualizing the relationship among the concepts. Current content analysis relies on items that researchers have arbitrarily selected; accordingly, extensive analysis on gathered data, therefore, is limited, and also external validity is not secured since it relies on coders of the data. Text mining, however, has been considered to surpass the limitation of traditional content analysis and used in a variety of fields using big data analysis, social network analysis, consumer product review analysis, and other useful methods. In other words, text mining extracts the appropriate variables to limit the breakdown of content analysis. Netzer et al. [7] examined the relationship between automobile brands through text mining and analyzed the market structure using the multidimensional scaling method. In addition, Mostafa [8] has classified the lexicon through 3D Map in the research that confirmed the brand sentiments of famous brands such as Nokia, IBM, and DHL through social network text mining.

Text consists of words, and analyzing text can be described as analyzing relationship among the words. In terms of it, text network analysis is also called semantic network analysis. That is to say, depending on the research, it can be called networks of words, network text analysis, semantic nets, networks of concepts, networks of centering words, text network analysis, or semantic networks [9].

Text network analysis, as mentioned above, complements the limitation of traditional content analysis, and extracts underlying meaning that the text delivers. Moreover, the pattern of text can be structurally analyzed to identify the relationship among the meanings and the relationship accordingly can be visualized through the analysis [10]. Text mining process has two phases including the data process phase and data analysis phase. The data process phase is relevant to data gathering and preprocess, while data analysis phase is relevant to text analysis that extracts significant information from the text, and visualizing information and extracting knowledge from the former analysis [6].

Through this process, large-volume data can be made more data-suitable for analysis, enabling continuous research to compare experiments.

*2.2. Consumers' Evaluation Criteria on Airlines.* Aviation industry can be described as a field where degree of interaction, customization, and labour intensiveness is relatively lower than that of other field [11]. Service, compared to products, has shown distinctive features of intangibility, heterogeneity, inseparability, and perishability [12]. In other words, consumers cannot see or touch purchased "services". In addition, production and consumption take place at the same time, and it cannot be stored, but even extinguishes once it is unused after production. In that properties of service, therefore, perceived service evaluation or recommendation can be crucial indicators in the field of aviation industry. Jia [13] has conducted text mining through a Chinese crowd-sourced online review community, and 49,080 pairs of restaurant ratings and reviews were examined, with high-frequency words, major topics, and sub-topics identified. After text mining, multilinear regression was employed to screen out the most impactful factors that influence taste, environment, and service ratings. Managerially, the idea of triggering the synergistic benefit from customer ratings and reviews is referential for market practitioners both within and beyond the catering industry.

In case of aviation industry, these elements including flight schedule, fare, services, punctuality, comfort of seats, safety, and frequent flyer program can be determinants of service satisfaction on airlines (IATA: International Air Transport Association). And in research of Hong and Park [14], factors that determine consumers' satisfaction on airline services include punctuality, safety, courteous agents, clean equipment, space, desirable schedule, profitability of the airline, reliability on the provided service, and financial costs. In addition, flight fare, boarding process, space and comfort of seats, in-flight meal, baggage delivery, and ticketing process also can affect consumers'

satisfaction. Based on researches, passengers who experience airline services tend to compare and evaluate its overall quality on the basis of tangible and intangible services provided by airlines. Will these factors be derived from the customers' texts in Internet? Internet comments based on anonymity reveal the desire of consumers and reveal more clear facts.

Aviation industry itself can be defined as a service industry where degree of interaction, customization, and labour intensiveness is relatively lower than that of other field [15]. In-flight service which is the interest of this study, however, can be described with high interaction, high customization, and high labour intensiveness. Thus, assessment and recommendation on services which consumers directly have experienced will be important indicator for the industry. In general, in-flight service consists of tangible service and human service that help with passengers' travel experiences; comfort of seats, in-flight employee services, in-flight food and beverage, entertaining services, etc. These elements, which are important in the aviation industry, are identified through text mining and classified through cluster analysis. These studies provide a possibility to become more realistic studies by combining text mining and statistical analysis.

### 3. Methodologies

**3.1. Research Process.** In this study, we conducted the following process to find out whether the core keywords can be identified through text mining and whether the core keywords are important for the performance of a company (see Figure 1). First, we gathered online reviews of customers and extracted core keywords from it by text mining method. And, we conducted text clustering analysis to explore the meaning of extracted core keywords. After that, we conducted empirical test for demonstrating the impact of core keywords through analysis of the relevance of the company's marketing performance, such as satisfaction, recommendation.

**3.2. Data.** This study used the online review data of airline customers, which was provided by global air service evaluation agency Skytrax in United Kingdom. This study set two large air carriers in Korea and Japan. The main contents of the data consist of two types of data. One type is the text data containing the customer's experience after using the air service. The other type is the survey data which include the evaluation of services, satisfaction, and recommendation after using the air service. The questionnaire survey was conducted online for customers who have recently used airline. The respondents are confirmed by presenting their plane ticket. The item of customer satisfaction was measured as 10 point-Likert scale, and recommendation intention was measured as binomial scale.

Data period is 3 years from January 2013 to December 2015. In the data period, 197 reviews were for the Korean carrier and 214 reviews for Japanese carrier. So, this study collected the review data on 411 people in total during data period and used them for analysis. Among the collected data,

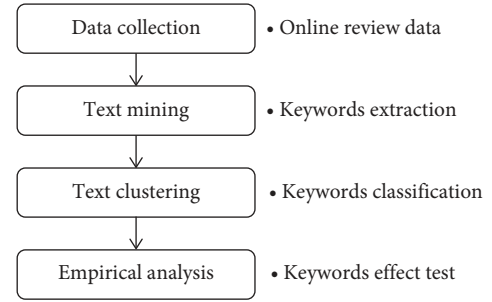


FIGURE 1: Research process.

text data were used for core keyword extraction and, questionnaire data were used for empirical analysis to verify the influence of core keywords.

The number of countries, to which the respondents were affiliated, was 32 in total, and the United States was found to be the most (33.8%), followed by Australia (12.7%), United Kingdom (10.7%), Canada (6.1%), and Singapore (5.1%). Table 1 shows the top 10 countries on the basis of the number of respondents.

**3.3. Analytic Process.** In order to analyze the text data, this study performed two processes. Text mining process was for extracting words, and text clustering process was for analyzing meaning based on extracted words.

The text mining process was a process of extracting words from a text document and creating word data. And this process included a step of extracting meaningful words based on text data. The detailed procedure is as follows [6]:

#### 3.3.1. Text Mining Process

**Step 1 (treat text).** This is the process of creating word data based on words contained in a text documents.

$$\vec{t}_d = (\text{tf}(d, t_1)), (\text{tf}(d, t_2)), \dots, (\text{tf}(d, t_m)), \quad (1)$$

where  $\vec{t}_d$  is the words vector;  $D$  is the documents set  $\{d_1, d_2, \dots, d_n\}$ ; and  $T$  is the words set  $\{t_1, t_2, \dots, t_m\}$ .

**Step 2 (extract words).** This is the process of extracting meaningful words by assigning weights through the appearance frequency of words between documents set.

$$W(t) = 1 + \frac{1}{\log_2 |D|} \sum_{d \in D} P(d, t) \log_2 P(d, t),$$

$$\text{with } P(d, t) = \frac{\text{tf}(d, t)}{\sum_{l=1}^n \text{tf}(d_l, t)}, \quad (2)$$

where  $\text{tf}(d, t)$  is the frequency of word  $t$  in a document  $d$ .

**3.3.2. Text Clustering Process.** On the other hand, text clustering process is a process of clustering through the distance between words. The cluster is set up and adjusted

TABLE 1: Characteristics of country.

No.	Country	Korea		Japan		Total	
		$N$	%	$N$	%	$N$	%
1	United States	58	29.4	81	37.9	139	33.8
2	Australia	50	25.4	2	0.9	52	12.7
3	United Kingdom	25	12.7	19	8.9	44	10.7
4	Canada	4	2.0	21	9.8	25	6.1
5	Singapore	3	1.5	18	8.4	21	5.1
6	Japan	3	1.5	15	7.0	18	4.4
7	South Korea	13	6.6	1	0.5	14	3.4
8	Brazil	1	0.5	12	5.6	13	3.2
9	Germany	7	3.6	4	1.9	11	2.7
10	New Zealand	8	4.1	1	0.5	9	2.2

after the cluster was set. In this method, (1) generate  $k$  initial seed randomly within data domain, (2) create  $k$  clusters by associating every observation with the nearest seed position, and (3) adjust the center position of each of the  $k$  clusters using the average of the observations belonging to the cluster. By repeating this process until all observations are associated, it can make clusters which have similar observations corresponding to the number of  $k$ .

*Step 1* (cluster setting). Set up a cluster with random seeds in the data, and the cluster is grouped by the Euclidean distance between each data and seed based on near distance.

$$\min S = \sum_{i=1}^k \sum_{x \in s_i} \|x - u_i\|^2 \quad (3)$$

where  $x$  is the words set  $\{x_1, x_2, \dots, x_n\}$ ,  $s$  is the clusters set  $\{s_1, s_2, \dots, s_k\}$ , and  $u_i$  is the average of cluster  $s_i$ .

*Step 2* (cluster adjust). Reset the value of cluster and adjust the position of cluster by using the average of the data in the cluster.

## 4. Results

*4.1. Keywords Extraction by Text Mining.* In this study, several preprocessing steps were performed to improve the accuracy of text mining.

First, all uppercase letters in text data are converted to lowercase letters to unify words. Also, we removed unnecessary special characters such as “@,” “\,” and so on. And the definite article (“the”), the indefinite article (“a,” “an”), prepositions (“of,” “in,” “for,” “through,” etc), and pronouns (“it,” “their,” “his,” etc) were also removed. Through this process, a total of 3,774 keywords were extracted. The keywords were rearranged based on the sparsity which represents the ratio of the whitespace in the matrix, because analyzing all keywords makes it difficult to explain meaningful results.

There were 11 keywords based on the sparsity 0.8, 19 keywords based on the sparsity 0.85, 45 keywords based on the sparsity 0.9, and 132 keywords based on the sparsity 0.95. In this study, a total of 45 keywords were extracted by applying the sparsity 0.9, and it is used for analysis. Too few

keywords are insufficient to understand customer behaviour, but too many keywords can distort results with less important words. In this analysis, 45 keywords were selected considering the usefulness of the interpretation while reflecting the customer behaviour.

The keywords with high frequency were “good,” “seats,” “cabin,” “class,” “excellent,” “comfortable,” “staff,” and so on. Figure 2 shows the frequency distribution of these keywords, and Figure 3 shows the frequency of keyword appearance by the word cloud method.

*4.2. Keywords Classification by Text Clustering.* In this study, we performed keyword classification analysis for semantic analysis of extracted 45 keywords. The results are as follows. Keyword classification was based on hierarchical clustering, and the distance between the keywords was measured by the Euclidean method.

As a result of clustering analysis, 45 keywords were classified into two clusters. Cluster 1 consisted of service content such as “seats,” “cabin,” “class,” and “staff” who provide the service. And it has service evaluation-related keywords such as “good,” “excellent,” “comfortable,” etc.

On the other hand, Cluster 2 consisted of more detail service content such as “meal,” “movie,” “drinks,” “check,” “served,” and so on. Also, it has service evaluation-related keywords such as “great,” “nice,” “well,” “best,” “clean,” etc (see Figure 4).

Figure 5 shows the result of keyword visualization with two clusters based on the  $k$ -means method for better insight. The keywords in Cluster 1 are formed to be spaced apart from each other, and the keywords in Cluster 2 are formed to be relatively wide spacing. And the two components explained 61.83% of the point variability.

*4.3. Effect of Core Keywords.* In this study, we performed combining analysis with core keywords data extracted by text mining and questionnaire respondent data to overcome the disadvantages of existing text mining research and to provide practical implications to the aviation industry. In other words, since core keywords appear as a result of text mining and cluster analysis can be considered to represent online reviews of airline customers, we attempted to understand the meaning of representative keywords by examining the influence of each evaluation concept on customer satisfaction and customer recommendation based on the evaluation of these customers.

In the analysis, we analyzed the effect of the four keywords, such as “seats,” “staff,” “cabin,” and “class,” which are the main keywords corresponding to the contents of the Cluster 1 on customer satisfaction and recommendation. The reason for this is that adjective and adverbial keywords among the keywords of Cluster 1 are mainly keywords indicating the result of the service. And the keywords appearing in Cluster 2 are keywords that deal with details of the service contents of Cluster 1. So, we define that these keywords are conceptually included in the above four top keywords.

In the analytical model, the questionnaire items were used as variables. Independent variables were the



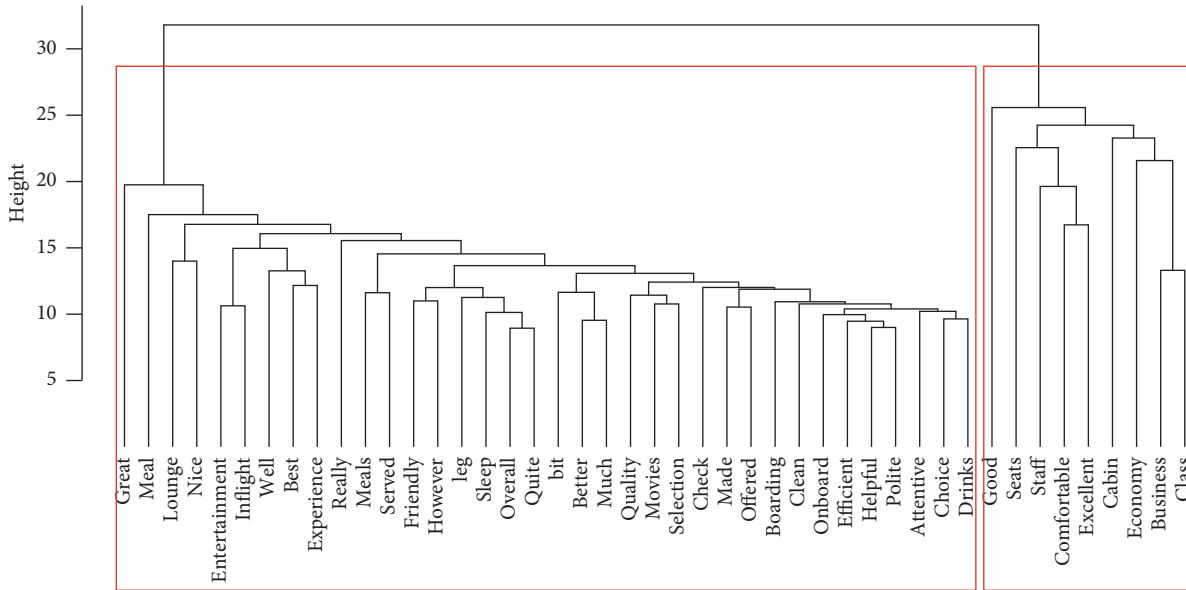


FIGURE 4: Cluster dendrogram.

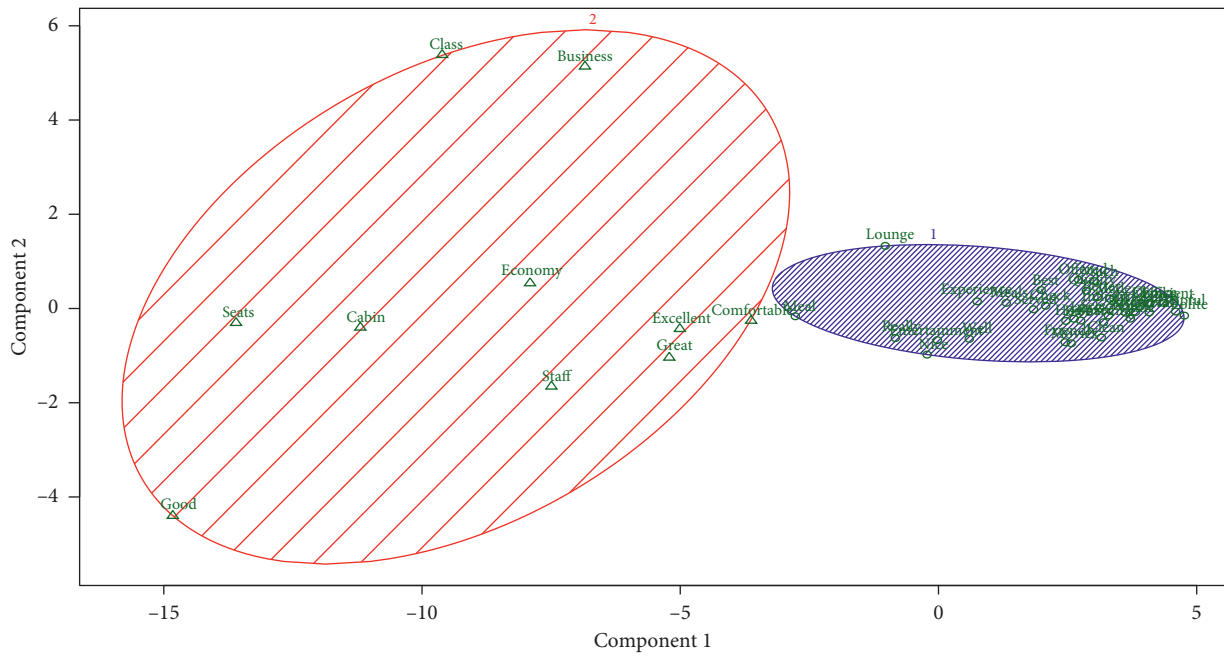


FIGURE 5: Cluster plot. These two components explain 61.83% of the point variability.

TABLE 2: Effects of core keyword on satisfaction.

Variables	<i>B</i>	$\beta$	<i>t</i> -value	<i>p</i> -value
Seats	0.987	0.513	18.6***	0.001
Staff	0.874	0.334	11.2***	0.001
Cabin	0.555	0.235	7.4***	0.001
Class	0.056	0.011	0.5	0.645

\*\*\* $\rho < 0.001$ ; *F* value: 342.4;  $R^2$ : 0.771.

TABLE 3: Effects of core keyword on recommendation.

Variables	<i>B</i>	$\beta$	<i>t</i> -value	<i>p</i> -value
Seats	0.156	0.536	13.5***	0.001
Staff	0.110	0.275	6.4***	0.001
Cabin	0.020	0.055	1.2	0.224
Class	0.039	0.051	1.5	0.139

\*\*\* $\rho < 0.001$ ; *F* value: 111.0;  $R^2$ : 0.518.

study used text mining to develop a unified understanding of keywords in the aviation industry in a data-driven way. Based on the airline review data, we proposed a two-step process of extracting key keywords by text mining and grouping them into cluster analysis. Specifically, we used a combination of metrics and clustering algorithms to preprocess and analyze text data related to keywords extraction method, including text from the scientific literature and news articles. This study seeks firstly for identifying prominent keywords at consumers' side using text mining method on consumers' online review data and then for confirming influences that the keywords affect corporate marketing performance. This study is not only the research that searches for key keywords [16], but also the research that identifies marketing performance in the aviation industry through text mining in service category level. Conclusion and implication are as follows.

First, keywords are shown to have distinctive cluster characteristics. As a result of identifying characteristics of major keywords through clustering, the keywords have been classified into three sections, including service that airlines provides with, details of each service, and assessment on the services. In other words, since the service provided by airline and the details of each service are differentiated at different levels, it indicates that service management should be centered on core service elements that can be clearly recognized in order to improve service evaluation.

Second, the key keywords extracted from text mining were found to have a different relationship with corporate performance. In other words, the service of the seats and the staff was more important to the company performance than the cabin or class. These results show that comfort is the key customer's needs in long-distance air travel and that the service focused on the comfort of the seat or the comfort of the seat is an important factor in corporate performance.

Third, recommendation and satisfaction should be managed distinctively in the service management of the aviation industry. The results show that keywords that affect consumers' satisfaction and keywords that affect consumers' recommendation are found to differ and the degree of impacts is also shown to be different as well. In other words, seats, staff, and cabin are important factors to improve the satisfaction and recommendation of consumers. Seats are the most important factors in consumer satisfaction and recommendation, but relatively more staff and cabin are important for consumer satisfaction and seats are more important for consumer recommendation. These results show that the human service of the crew, which can be relatively subjective, is more influential in satisfaction. On the other hand, seats are more important for relatively objective recommendations.

This study presents academic implication that the study has extended its application area of text mining. It has currently focused on exploratory study, while this study has extended it to study field of cause and effect. Moreover, the research also presents practical implication for corporations to efficiently manage keywords. In spite of the implications and advantage of text mining [17], this study has limitations. First, the research range is limited to two airlines in Korea

and Japan. Secondly, since the review data are produced spontaneously by clients, collected data can be biased and limited to active clients who are willing to express his/her experience. Therefore, in future study, selecting more representative airlines is needed and also minimizing convenience of the respondents is needed to generalize the results of the study in the future.

## Data Availability

The text mining data used to support the findings of this study are available from the authors upon request. The data used in this study are Airline Reviews and Rating data from Skytrax (<http://www.airlinequality.com>).

## Conflicts of Interest

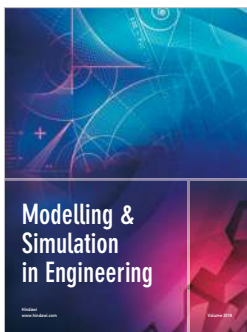
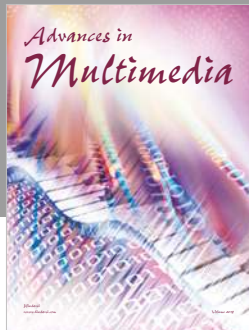
The authors declare that they have no conflicts of interest.

## References

- [1] P. Nayak and A. K. Chatterjee, "Effects of aluminium exposure on brain glutamate and GABA systems: an experimental study in rats," *Food and Chemical Toxicology*, vol. 39, no. 12, pp. 1285–1289, 2001.
- [2] J. A. Chevalier and D. Mayzlin, "The effect of word of mouth on sales: online book reviews," *Journal of Marketing Research*, vol. 43, no. 3, pp. 345–354, 2018.
- [3] W. Fan, L. Wallace, S. Rich, and Z. Zhang, "Tapping the power of text mining," *Communications of the ACM*, vol. 49, no. 9, pp. 76–82, 2006.
- [4] D. Meyer, K. Hornik, and I. Feinerer, "Text mining infrastructure in R," *Journal of Statistical Software*, vol. 25, no. 5, pp. 1–54, 2008.
- [5] R. V. Kozinets, K. de Valck, A. C. Wojnicki, and S. J. S. Wilner, "Networked narratives: understanding word-of-mouth marketing in online communities," *Journal of Marketing*, vol. 74, no. 2, pp. 71–89, 2010.
- [6] A. Hotho, A. Nürnberger, and G. Paaß, "A brief survey of text mining," *Ldv Forum*, vol. 20, no. 1, pp. 19–62, 2005.
- [7] O. Netzer, R. Feldman, J. Goldenberg, and M. Fresko, "Mine your own business: market-structure surveillance through text mining," *Marketing Science*, vol. 31, no. 3, pp. 521–543, 2012.
- [8] M. M. Mostafa, "More than words: social networks' text mining for consumer brand sentiments," *Expert Systems with Applications*, vol. 40, no. 10, pp. 4241–4251, 2013.
- [9] D. Paranyushkin, "Text network analysis (2010)," in *Conference du Performing Arts Forum*, 2011, <http://noduslabs.com/research/pathways-meaning-circulation/>.
- [10] R. Feldman and J. Sanger, *The Text Mining Handbook: Advanced Approaches in Analysing Un-Structured Data*, Cambridge University Press, Cambridge, UK, 2007.
- [11] R. W. Schmenner, "How can service businesses survive and prosper?," *Sloan Management Review 1986-1998*, vol. 27, no. 3, p. 21, 1986.
- [12] J. Fitzsimmons and M. Fitzsimmons, *Service management: Operations, Strategy, Information Technology*, McGraw-Hill Higher Education, New York City, PA, USA, 2013.
- [13] S. Jia, "Behind the ratings: text mining of restaurant customers' online reviews," *International Journal of Market Research*, vol. 60, no. 6, pp. 561–572, 2018.
- [14] J.-W. Hong and S.-B. Park, "Study on the extraction of core keywords and its effects through text mining," *International*

- Journal of Web Science and Engineering for Smart Devices*, vol. 3, no. 2, pp. 7–12, 2016.
- [15] D. Jung and S. B. Kim, “The effects of information quality of mobile on consumer behaviour intension; for airline,” *Asia-pacific Journal of Multimedia Services Convergent with Art, Humanities, and Sociology*, vol. 6, no. 9, pp. 19–26, 2016.
- [16] S.-B. Park and J.-W. Hong, “Relationship of core keywords and marketing performance obtained by using text mining: a comparative study,” *New Physics: Sae Mulli*, vol. 67, no. 5, pp. 562–568, 2017.
- [17] X. Wang and S. Liu, “Analysis and research of enterprise technology competent advantage on text mining and correspondence analysis,” *International Journal of Database Theory and Application*, vol. 6, no. 5, pp. 133–140, 2013.






**Hindawi**

Submit your manuscripts at  
[www.hindawi.com](http://www.hindawi.com)

