International Journal of Engineering, Management and Humanities (IJEMH)
Volume 3, Issue 3, pp: 125-137 www.ijemh.com

A Study on Hotel Review Mining Utilizing Sentiment Analysis and Topic Modelling: With Focus on 5-Star Hotels in Korea

Kim, Dong Won[1], Kim, Yeon Joo[2]
[1]*Professor, Liberal Education Center, Korea Maritime and Ocean University, Korea*
[2]*Professor, Department of English Language and Literature, Korea Maritime and Ocean University, Korea*
Corresponding Author: Prof. Kim,Yeon Joo,

Date of Submission: 24-06-2022 Date of Acceptance: 05-07-2022

ABSTRACT: The purpose of this study is to find a way to improve actual customer satisfaction by utilizing methods such as sentiment analysis and topic modelling based on hotel reviews, which are vast amounts of unstructured data. Although the existing LDA technique extracts the various topics from thousands or tens of thousands of online reviews and can check the key vocabulary composing each topic and what topics each review consists of, there is also a limit not to grasping the relationship between the review score and review content. In this study, we tried to overcome the shortcomings of LDA. In particular, customer data was extracted by utilizing actual customer evaluations in practice, and the DMR (Dirichlet Multinomial Regression) technique is used to identify topic differences between evaluation groups. As a result of analysing 24,123 online review data of 30 five-star hotels in Korea, 8 positive topics, 6 negative topics, and 1 unclassified topic were extracted, and hotel properties that affected the topic score of text content have been identified. The results gave hotel managers the opportunity to identify specific hotel properties that are causing customer dissatisfaction and to discover how to effectively improve customer satisfaction through appropriate feedback.

KEYWORDS: DMR (Dirichlet Multinomial Regression), Online Hotel Review, Sentiment Analysis, Text Mining, Topic Modelling.

I. INTRODUCTION

As hotel reservation platforms become popular, there is currently a very large amount of hotel review data in the form of text on the web. Online platform sites enter the indicators for simple decision-making such as title, score, travel type, etc. together with reviews for potential hotel customers. Potential customers use this optional information to make decisions about hotel visits and can decide more easily [1]. In fact, customers can avoid excessive information and make more efficient and rational purchasing decisions through these indicators [13, 14]. Among these indicators, the customer evaluation score has already been confirmed in several studies as an indicator that affects the purchasing attitude of other customers [3, 4, 8, 26]. From this point of view, if the correlation between the score given by the hotel customer and the review can be confirmed, it can be important information for the hotel and the customer to understand what the score is from each point of view. In line with this recent trend, analysis and research on online review texts to more accurately identify vivid customer needs are continuously expanding not only overseas but also in Korea [12, 18, 19, 25]. The methodology used in these existing studies is a study on attributes that affect review scores through regression, the frequency of appearance of vocabulary related to service key attributes, and identifying important attributes and words using sentiment analysis.

In this study, unlike the methods used by existing studies for text-based unstructured big data analysis, topic modelling is used to perform text mining through topic classification. It is a type of artificial intelligence learning method that divides a large amount of documents into several topics and can effectively analyse and reveal what words the topic consists of. Unlike other text analysis that is based on the frequency of words, topic modelling has the advantage of less over-fitting problems and being able to be analysed even when new data is introduced because it is a probability-based analysis through a variable called ‘topic’. In addition, if the researcher manually specifies the words required for analysis or sets the algorithm plan relatively clearly, as opposed to sentiment analysis, which builds a sentiment lexicon before analysis, most of the pre-processing tasks can be automated and the researcher's manual intervention can be minimized.

The most used technique in topic modelling is LDA (Latent Dirichlet Allocation). LDA is being actively used not only in the engineering field but also in the social science field that analyses various social phenomena [6, 16, 22]. However, LDA topic modelling has several methodological limitations. First, the analysis results give undue value to words with high frequency [6]. Secondly, LDA topic modelling requires setting the number of topics, but there is no criterion for determining the number of topics [9]. Finally, since the LDA does not directly reflect the review score, the relationship between the result and the score derived from the LDA is insufficient [17].

In order to solve these limitations of LDA topic modelling, in this study, 'IDF (Inverse Document Frequency) word weighting', 'topic number setting through concordance score', and 'DMR model application' were attempted. IDF was applied to words as weights to reduce the weight of words that appear excessively in all documents to avoid diluting the meaning of the subject. In addition, the optimal number of subjects was derived using an index indicating semantic similarity called the coherence score. Finally, DMR (Dirichlet Multinomial Regression), one of the derivative models of LDA, was used. DMR is an algorithm created for the purpose of identifying differences in topic by group, and is a model suitable for identifying differences in content composition for each group when divided into positive, neutral, and negative groups according to hotel review scores.

The main purpose of this study is to identify the relationship between hotel attributes and scores through hotel reviews, which is one of the unstructured data that is becoming more and more important in the hotel industry, and to draw implications from this. To this end, we tried to solve the limitations of topic modelling using the existing LDA through DMR and effectively solve the researcher's subjective intervention and over-fitting problems in text analysis. In addition, by minimizing manual tasks such as making a sentiment lexicon, automation is maximized in terms of time efficiency. Finally, using this research method, the hotel attributes that affect customer sensibility were derived and interpreted, and practical implications were also presented.

II. THEORETICAL BACKGROPUND

Recently, when big data is used in various industries, there are many attempts to analyse non-quantitative data such as text and images. Recent representative studies are as follows. Lim et al. (2019) created a sentiment lexicon using word frequency and Harvard IV lexicon for hotel online reviews, and conducted elastic regression between the words in the sentiment lexicon and the review score to derive keywords affecting positive and negative [18]. Kwak et al. (2019) organized important hotel attributes, set 10 categories and 36 detailed attributes, derived keywords for each attribute through qualitative analysis, and measured the positive rate for each attribute[19]. Shim et al. (2018) also identified the individuality items of the tourist destination, and the relationship between the review score and the individuality item of the tourist destination by regression analysis of only sentences containing words and the Word2Vec word embedding technique [25]. In these studies, important attributes and words related to hotels were derived in advance through previous studies, and the frequency of appearance of the word was measured or emotional scores were derived through a sentiment lexicon.

In the existing research through sentiment analysis, a pre-made sentiment lexicon was built with hotel properties selected by the researcher, and the relationship between hotel properties and review scores was identified through the frequency of positive and negative words in the sentiment lexicon [8, 9, 25]. The method of deriving implications through the measured sentiment score was mainly used, and in some cases, the existing sentiment lexicon was modified to suit the research purpose, but the researchers manually judge it [13, 18, 19]. On the other hand, topic modelling is an artificial intelligence learning method that automatically classifies reviews according to topics without word designation in a dictionary such as a sentiment lexicon, and grasps the content of the topic. Unlike sentiment analysis using a sentiment lexicon, introducing topic modelling can minimize word designation through researcher intervention. The purpose of this study is to understand the relationship between hotel properties and scores through hotel review through topic modelling. Topic modelling mainly uses LDA, but LDA classifies topics based only on review content. Therefore, this study used DMR, one of the algorithms derived from LDA, to derive a topic in consideration of review scores. Using DMR, we try to find hotel attributes and meanings that affect customers' evaluation scores based on text analysis of the contents of review data left online by customers using domestic hotels after staying at the hotel.

2.1 LIMITATIONS OF LDA RESEARCH AND IMPROVEMENT PLAN

LDA, which is a representative technique of topic modelling, is based on the probability theory of identifying the structure of a topic hidden from a document based on the Dirichlet distribution. It starts with the assumption that all documents share topics, and that each document has various combinations of probabilities for the topic. LDA uses the Bayesian probability theory to estimate that the prior distribution, which can best explain the posterior distribution, has the maximum log-likelihood when the number of subjects is specified for how many clusters of words are grouped [2]. In this study, each review is set as a document, and when LDA is applied, two probability distributions P(word|topic) and P(topic|review) can be derived. From the study of simply classifying documents into a set number of subjects using LDA [16], the reviews of TripAdvisor was classified through LDA and the relationship between topics and the review scores was also confirmed [6, 22]. Furthermore, when cluster analysis or sentiment analysis was performed on the document, it was also suggested that even if only keywords derived from LDA were used for analysis, the characteristics of the document could be reflected as well using all the words [12, 13, 15]. As such, LDA can be used for various analyses other than frequency analysis by converting unstructured data called words into structured data called a probability distribution. However, studies using existing topic modelling still have several methodological limitations. “High-frequency word processing”, “arbitrarily setting the number of topics”, and “using a limited topic modelling algorithm” are examples. In this study, to overcome the limitations of LDA, weight setting through the IDF, the number of subjects selection, and the score prediction through DMR were used.

2.2 WEIGHT SETTING(IDF, INVERSE DOCUMENT FREQUENCY)

In precedent studies using LDA, analysis was performed with the same weight for all words. This causes a problem in that the tendency to derive topics mainly from words with a high frequency of appearance increases. The probability of the word appearing across all themes increases that much, and there is no choice but to show a limit in acquiring the unique meaning of the subject. Several methods have been devised to solve this problem. In some cases, the researcher arbitrarily removes and analyses some high-ranking words with an overwhelmingly high frequency of occurrence compared to other words, and applies IDF to topic modelling to reduce the weight of words that appear evenly across multiple documents. A method of carrying out this has also been suggested. IDF is calculated as “$\log(N/n) + 1$”, where ‘N’ is the total number of documents and ‘n’ is the number of documents containing a specific word. If a specific word appears in all documents, the IDF value becomes 1, and the value increases as the number of documents including the word decreases [5]. This method can be seen as more objective than the method of excluding words with a high frequency of occurrence arbitrarily, and it can effectively prevent the occurrence of specific words across multiple topics. Huang et al. conducted a topic modelling analysis comparing the LDA model reflecting the IDF with the basic LDA and K-cluster models [7]. As a result of checking the accuracy of the three models, the LDA model to which the IDF weight was applied had a higher F score, indicating accuracy, than the other two models, and showed higher performance than the LDA in calculation speed.

2.3 SELECTING THE NUMBER OF TOPICS

In topic modelling, several methods have been devised to evaluate the appropriate number of topics. The researcher arbitrarily selected an appropriate number of topics, checked out the keywords that compose the topic, and left only the topics that were easy to interpret, but also removed the rest [8, 16, 22]. However, since this method involves the subjectivity of the researcher, in other studies, when selecting an appropriate number of topics, topic modelling is first performed with the number of multiple topics, and the perplexity is calculated to determine the number of topics with the lowest perplexity for the optimal topic [6].

The perplexity is calculated as $2^{-\sum LL(\omega)/N}$ where *LL(ω)* is the log-likelihood. It is a value obtained by taking the logarithm of the probability that a specific word will be assigned to a specific topic within topic modelling, and after adding up all these values, dividing by N, the total number of words, is averaged and indexed. This method takes advantage of the fact that when the number of topics is less than or more than the proper number, the topic assignment of words is not probabilistically consistent, but when the number of topics is the proper number, the words are probabilistically consistently assigned to the topic. However, as this value is smaller, the result of topic modelling is probabilistically well-learned but does not mean that the result is suitable for researcher interpretation. Newman et al. collected news and book data and topic-modelled them to create a rating scale suitable for researchers to interpret, and confirmed how many topic keywords appear together in this data [20]. The topic coherence was established. In other words, it was confirmed that the more semantically similar words are gathered, the higher the degree of coherence. Roder et al. measured the similarity between words based on the Pointwise Mutual

International Journal of Engineering, Management and Humanities (IJEMH)
Volume 3, Issue 3, pp: 125-137 www.ijemh.com

Information (PMI), a measure used as a useful index for language extraction in linguistics in a study on topic coherence measurement [24]. Normalized PMI can be calculated with the following formula.

$$NPMI(\omega_i,\omega_j)=log(P(\omega_i,\omega_j)+\epsilon/P(\omega_i)\cdot P(\omega_j))/-log(P(\omega_i,\omega_j)+\epsilon)$$

Here, $P(\omega)$ is the probability of occurrence of a word, and $P(\omega_i, \omega_j)$ is the probability of simultaneous occurrence between tokens on both sides of the word reference. There are various ways to measure the similarity between words based on the above formula. For example, words such as one-all, one-any, any-any, one-one, etc., depending on whether words appearing simultaneously are compared with one or several words. It is possible to calculate the probability of occurrence in various ways depending on whether it is counted within the directly set token range, counted in one document, or how the scores are summed. The $\mathcal{C}_{\upsilon}$ method using -set, ±110 word range, cosine similarity measurement, and arithmetic score summing method showed the most similarity. The concordance score has a value between [0, 1], and if the concordance score is 1, it means that the word group always appears together in all documents.

2.4 SCORE PREDICTION WITH DMR

When conducting text analysis-related research in the field of hotel tourism, the main topic of interest is the factor affecting the review score can be identified. However, LDA, which can be considered the most representative model among topic modelling, does not consider the review score, so the topic derived through analysis is actually an independent value that has no correlation with the review score. Therefore, the topics derived through LDA only have a relatively insignificant effect on the review score compared to other attributes or have no choice but to use the sentiment lexicon together [6]. Contrary to this, Dirichlet Multinomial Regression (DMR) is a model that assumes that metadata (author, year, etc.) affecting a sentence will affect this Dirichlet prior coefficient α. This study used this DMR analysis method. DMR is an algorithm that divides text groups based on a nominal variable called meta data and simultaneously considers differences in the subject structure of documents in each group. In DMR, not only P(topic|document) and P(word|topic), but also the λ value indicating how much each topic is reflected according to the group is also derived. Here, e^{λ} is equal to α, which is the prior coefficient of LDA, and the probability that the corresponding group includes a specific topic increases in proportion to this value.

In this study, the topics were identified relatively more frequently according to each score group by performing DMR by dividing them into 5 groups related to the review scores (1 to 5). As such, DMR is an effective methodology to identify the differences in topics by a group but has not yet been applied to research in the social sciences.

III. STUDY DESIGN

3.1 RESEARCH PHASE

The research tasks of this study are as follows.

i. Research task 1: Through DMR, classify the topics into positive and negative topics by identifying the weight of topics by the group with evaluation scores (1 to 5) left by customers.
ii. Research Task 2: For topics classified as positive and negative, track reviews related to each topic, and determine how often the review actually appeared in positive (4-5 points) or negative (1-2 points) groups.
iii. Research Task 3: Review the validity of this research method by comparing the frequency of appearance by the score with positive and negative topics through DMR.
iv. Research Task 4: Identify hotel attributes that affect customer sensibility on the verified topics and derive practical implications.

In order to achieve this research task, in this study, as shown in Fig. 1, online review data of tourists who used five-star hotels in Korea were first collected, and after refining this text data by natural language processing, each Calculate the IDF weight for each word. Second, by applying the IDF weight to the words, the reviews were divided into five groups by score and DMR topic modelling analysis was performed. The number of topics was set in units of 5 from 5 to 50, and the number of topics showing the highest concordance score was finally used. Third, emotional scores were calculated through the topic weight α for each score group obtained from DMR, and topics related to positive or negative were classified based on 3 points, which is a median value of 1 to 5 points. Fourth, the validity of the positive/negative topics classified as DMR was verified by tracking the reviews on the positive/negative topics derived in this way and measuring the frequency by scores of these reviews. Lastly, practical implications were presented on the effects of 'what' in customer reviews on verified positive and negative topics.

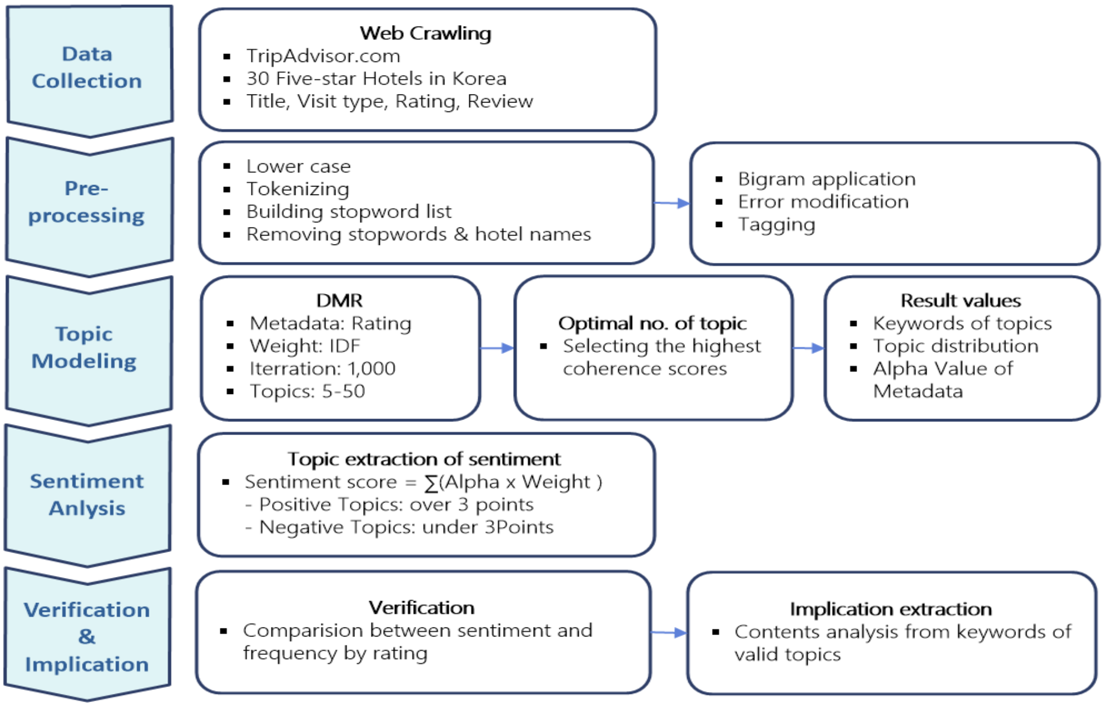

Figure1 Research flow

3.2 DATA COLLECTION

The subjects of this study are customers who use five-star hotels in Korea. In order to collect data, on December 31, 2021, 30 hotels registered as five-star hotels in Korea on TripAdvisor (www.tripadvisor.com) All English reviews registered up to now were searched. The reason for limiting the study to domestic five-star hotels is that the topics that affect customer sentiment may differ depending on the hotel rating and purpose [18]. This is because positive or negative tendencies may be diluted depending on the rating of the hotel. Web crawling was made directly in Python, and the 'selenium' library was used. Collection items such as review title, review content, and evaluation score and a total of 24,123 English reviews were obtained as data for analysis.

3.3 DMR(DIRICHLET MULTINOMIAL REGRESSION) LEARNING

When performing the LDA technique, α and β values called prior variables are set as essential input variables prior to analysis. α is the topic distribution P(topic|review) Dirichlet coefficient, and the smaller the α value, the more reflected the posterior distribution in the learning process [21]. β is the topic distribution P(word|topic) Dirichlet coefficient, and the smaller the β value, the more reflected the posterior distribution in the learning process. Usually, α is set to 0.1, and β is set to 0.01 and analysed. DMR also requires the input of α and β variables like LDA, but α converges to an optimal solution in the learning process, and the weight of topics according to score groups can be grasped according to the α value. In addition, the number of topics, minimum word frequency, maximum word frequency, word weight, and variance were set as input variables. In this study, α was initially input as 0.1, and β was set as 0.01. The number of topics was increased in units of 5 from 5 to 50. The minimum word appearance frequency was 10, and the maximum word appearance frequency was not specified. The word weight was set to IDF, and the variance was set to 2.0. The larger the variance value, the larger the difference in t value of the topic for each score group. A total of 1,000 iterations were performed, and the contents of these input variables are summarized in Table 1.

Table 1 Input Variables

Variable	Values
α	0.1
β	0.01
The number of topics	5,10,15,20,25,30,35,40,45,50
Minimum word frequency	10
Maximum word frequency	None
Word weight	IDF
Variance	2.0
The number of iteration	1000

3.4 SELECTING OPTIMAL NUMBER OF TOPICS

In this study, the coherence score was calculated using the C_υ method, which showed the highest correlation score compared to similarity score that researcher directly input [24]. After topic modelling is performed with the number of topics from 5 to 50, the coherence score of the C_υ method is calculated for the keywords of each derived topic. After that, the number of topics with the highest coherence score was set as the optimal number of topics. Calculation of coherence score was performed through Python gensim.

3.5 CALCULATION OF SENTIMENT SCORES

By executing DMR with the optimal number of topics, it is possible to derive the topic word probability distribution P(word|topic), the review topic probability distribution P(topic|review), and α value representing the weight of the topic for each score group. After finding keywords for each topic through topic word probability distribution, and guessing the meaning and contents of the topic, we calculate the weight by dividing the α values from 1 to 5 in each topic by the sum of the α values of the topic, and then the weight is multiplied by the score to calculate the sentiment score. The formula for calculating the sentiment score is as follows.

$$p \text{ topic sentiment score} = \sum_{q=1}^{5} (\alpha_{p,q} / \textstyle\sum_q \alpha_{p,q} \times q)$$

p: topic number, q: score, $\alpha_{p,q}$: α value of the q score group in the p topic, “$\alpha_{p,q}/\sum_q \alpha_{p,q}$” is the weight of the topic's α value, and the higher the α weight of a specific score, the more similar the sentiment score to that score.

3.6 VALIDATION

The validity of the research results was verified by comparing the positive and negative of the topic through the sentiment score with the frequency of review of the topic by actual score. First, for verification, a topic with a probability of P(topic|review) of 50% or higher is judged as a review of the topic. After finding the actual frequency and expected frequency for each topic and score, the difference between the two was divided by the expected frequency to calculate the frequency ratio.

Frequency Ratio = (Actual Frequency – Expected Frequency)/Expected Frequency

If the Frequency Ratio is negative, it means that the number of actual reviews is smaller than the stochastic expected value. Conversely, if the Frequency Ratio is positive, it means that there are more actual reviews than the stochastic expected value. If it is exactly '0', it means that there are as many reviews as the probabilistic expected value. For reviews on negative topics, the frequency ratio must be positive because the actual frequency must be higher than the expected frequency at points 1 and 2, and the frequency ratio is negative because the actual frequency is less than the expected frequency at points 4 and 5. Conversely, reviews on positive topics should yield negative values in the 1 and 2 point intervals and positive values in 4 and 5 points. It is verified by comparing positive and negative through frequency ratio and positive and negative through sentiment score.

IV. RESEARCH RESULTS

4.1 SPECIMEN CHARACTERISTICS

In order to collect the analysis data for this study, a total of 24,123 customer review scores of 30 domestic hotels are shown in Table 2 below.

Table 2 Customer Review Scores

Review Points	1	2	3	4	5	Total
Number of Cases	499	652	1,890	6,531	14,551	24,123
Ratio	2%	2.7%	7.8%	27%	60.3%	100%

Most of the number of cases was 4 points, 6,531 cases (27.0%) and 5 points, 14,551 cases (60.3%), accounting for 87.3%. As the main purpose of the study is to understand the topics of customer reviews that affect the customer's hotel evaluation score in all five-star domestic hotels, we conducted data research under the judgment that reviews of hotels with few reviews would not cause errors in the research results. As a result of the natural language pre-processing for the entire review data, a total of 1,465,626 words were input to the DMR analysis, and 6,132 words were used.

4.2 RESULTS OF DMR LEARNING

As a result of performing DMR with the number of topics in 5 units from 5 to 50, the agreement score was the highest at 0.49 when the number of topics was 20 as shown in Fig. 2. This result is interpreted as saying that the words grouped by each topic for 20 topics always come out together with a probability of about 50% in the overall review

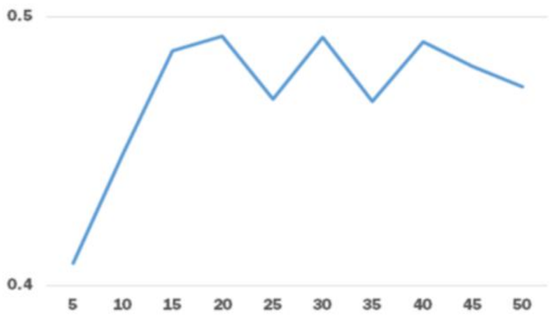

Figure 2 Coherence Score of the Number of Topics

As a result of the DMR analysis, α is derived according to topics (20) × scores (5, 1 to 5). Based on the score, it is interpreted that a topic with a larger α value contains more words of that topic, and is judged which topic appeared remarkably in a specific score group as finding a probabilistically exceptionally higher α value. For a clear judgment standard, the sentiment score was calculated by multiplying the weight and the score of each topic's α value, and based on this score judgment is made on whether the topic is positive or negative. Specifically, based on the sentiment score of 3, topics higher than 3 are judged as positive and lower as negative and the intensity of positive and negative can be determined by the difference in score from 3 points. In addition, based on the top 10 keywords for each topic, it is possible to determine what the topic corresponds to among hotel properties and specifically what content contains in relation to hotel properties. For instance, topic 1 has a sentiment score of 1.77, and words such as 'smoke', 'construction', 'noise', 'floor', 'elevator', and 'smell' appear as keywords causing customer dissatisfaction. It is predicted that the cause is the smell of cigarettes, the structure of the room, or noise caused by other causes, and it can be seen that the review consisting of topic 1 results in a low score. Topic 2 has an sentiment score of 3.81 and is composed of positive words such as 'great', 'perfect', 'thank', 'visit', 'staff', 'best', 'recommend', 'time', and 'stay'. These words come out when customers are overall satisfied with the hotel's accommodation and service, and the word composition of the topic and the sentiment score are almost identical. Through this process, a title can be assigned to each topic and organized as shown in Table 3. Table 3 summarizes the sentiment scores for each topic and the top 10 keywords

Table 3 Results of Topic Extraction

Topic number	Sentiment scores	Topic title	Top 10 Keywords				
1	1.78 (Negative)	Noise/Smell	smell	smoking	floor	elevator	noise
			smoke	door	construction	bathroom	hear
2	3.8 (Positive)	Accommodation/Staff	recommend	visit	make	stay	staff
			best	perfect	thank	great	time
3	3.2 (Positive)	Pool/Gym	hot	use	swim	water	gym
			pool	sauna	kid	child	swimming
4	2.44 (Negative)	Membership	member	spg	upgrade	platinum	check
			lounge	pay	tell	book	star
5	3.15 (Positive)	Location/View	mall	river	view	good	business
			nice	restaurant	ifc	pool	great
6	2.48 (Negative)	Breakfast	egg	fruit	juice	fresh	bread
			dish	breakfast	buffet	korean	western
7	2.77 (Negative)	Room facilities	shower	bathroom	window	light	control
			floor	curtain	ceiling	tub	glass
8	2.92 (Negative)	Reception	wait	guest	get	desk	manager
			check	tell	ask	call	say
9	3.23 (Positive)	Staff service	help	concierge	taxi	thank	driver
			make	check	desk	name	front
10	3.86 (Positive)	Experience	property	season	best	experience	service
			lounge	stay	top	staff	make
11	4 (Positive)	Ancillary facilities	gym	pool	great	good	lounge
			breakfast	excellent	restaurant	nice	view
12	2.78 (Negative)	Club lounge	club	lounge	drink	floor	wine
			breakfast	food	evening	snack	buffet
13	3.66 (Positive)	Location/Tourism	wing	good	location	business	great
			shopping	new	mall	place	kid
14	2.15 (Negative)	Food/Beverage	coffee	order	tea	eat	food
			drink	menu	chicken	ask	get

International Journal of Engineering, Management and Humanities (IJEMH)
Volume 3, Issue 3, pp: 125-137 www.ijemh.com

15	3.01	Anniversary visit	birthday	thank	special	make	lounge
			cake	team	amazing	stay	wonderful
16	3.77 (Positive)	Room facilities	toothbrush	phone	charge	provide	bed
			water	plug	power	free	wifi
17	1.72 (Negative)	Room environment	air	temperature	conditioning	hot	sleep
			degree	window	turn	control	cold
18	3.63 (Positive)	Transportation/ Accessibility	bus	airport	station	walk	subway
			shopping	store	take	shuttle	city
19	3.63 (Positive)	Price	business	good	price	internet	restaurant
			area	nice	expensive	location	bit
20	2.19 (Negative)	Residence	bedroom	kitchen	suite	apartment	living
			bed	dryer	machine	bathroom	washing

Keywords calculated by word weight of the topic.

4.3 VALIDITY VERIFICATION

As shown in Table 4, it is possible to confirm whether sentiment scores positively or negatively impact actual customer reviews by comparing positive and negative topics determined by sentiment scores and the frequency rate of reviews by the scores of each topic. Prior to this, the frequency ratio is small because the expected frequency of the groups of 4 and 5-point is relatively high, since the 4 and 5-point reviews are 21,082 (87.3%), and the 1 and 2-point reviews are only 1,511 (4.7%). Therefore, the sign should be considered more carefully than the size of the value in the interpretation of the results.

Table 4 Frequency Ratio

Topic number	1 point group	2 point group	3 point group	4 point group	5 point group	Sentiment scores
1(Negative)	8.15	8.43	4.06	0	-0.8	1.78(Negative)
2(Positive)	-0.72	-0.92	-0.88	-0.74	0.42	3.80(Positive)
3(Not defined)	-0.26	0.35	0.38	0.34	-0.18	3.20(Positive)
4(Negative)	7.39	6.97	4.92	0.15	-0.89	2.44(Negative)
5(Positive)	-0.92	-0.85	-0.55	0.51	-0.11	3.15(Positive)
6(Negative)	-1	1.74	-0.2	0.09	-0.04	2.48(Negative)
7(Positive)	-0.77	-0.17	-0.51	-0.04	0.09	2.77(Negative)
8(Negative)	17.02	9.6	2.25	-0.43	-0.71	2.92(Negative)
9(Positive)	-0.6	-0.76	-0.75	-0.56	0.33	3.23(Positive)
10(Positive)	-0.86	-0.75	-0.72	-0.56	0.34	3.86(Positive)
11(Positive)	-0.95	-0.82	-0.56	-0.08	0.13	4.00(Positive)
12(Positive)	-0.55	-0.8	-0.11	0.25	-0.06	2.78(Negative)
13(Positive)	-1	-0.67	-0.52	0.02	0.08	3.66(Positive)
14(Negative)	2.99	7	1.98	0.16	-0.51	2.15(Negative)
15(Positive)	-0.81	-0.65	-1	-0.81	0.46	3.01(Not defined)
16(Not defined)	-0.41	1.69	2.92	0.46	-0.49	3.77(Positive)
17(Negative)	5.1	7.03	3.15	0.35	-0.75	1.72(Negative)
18(Positive)	-0.93	-0.86	-0.5	0.23	0	3.63(Positive)
19(Positive)	-0.71	-0.07	1.59	0.58	-0.37	3.63(Positive)
20(Positive)	-1	-1	-0.43	-0.01	0.09	2.19(Negative)

Reviews are assigned topic that likelihood is over 50%.
Expected frequency of topic , score j = P(topic 姹)×P(score 蒴)×the number of all reviews

International Journal of Engineering, Management and Humanities (IJEMH)
Volume 3, Issue 3, pp: 125-137 www.ijemh.com

Topic 1, Topic 4, Topic 6, Topic 8, Topic 14, and Topic 17 are all negative topics lower than 3 points, and 1 and 2-point groups are positive numbers, and 4 and 5-point groups are negative values. On the other hand, topic 2, topic 5, topic 9, topic 10, topic 11, topic 13, topic 18, and topic 19 were positive topics higher than 3 points, and negative numbers were found in 1 and 2-point groups and positive numbers in 4 and 5-point groups. In this topic, the emotional score of the topic is confirmed to match the score distribution and direction of the actual review. However, when topic 3, topic 7, topic 12, topic 15, topic 16, and topic 20 are based on emotional score 3, the direction of the score does not match the score distribution. For example, topic 3 and topic 16 should have negative numbers for the 1 and 2-point groups, and positive numbers for the 4 and 5-point groups, but their direction is unclear. In addition, the direction of positive and negative reviews by score was opposite in topic 7, topic 12, and topic 20, and the results were summarized as Table 5.

Table 5 Positive/Negative Topics of Sentiment Score and Review Frequency Ratio

Topic number		Sentiment Scores		
		Positive (Over 3 point)	Not defined (Too closed to 3point)	Negative (Under 3 point)
Frequency ratio	Positive 1, 2 point group(-) 4, 5 point group(+)	Group 1 Topic2(Accommodation/Staff), Topic5(Location/View), Topic9(Staff service), Topic10(Experience), Topic11(Ancillary facilities), Topic13(Location/Tourism), Topic18(Transportation/Accessibility), Topic19(Price)	Group2 Topic15 (Anniversary visit)	Group3 Topic7(Room facilities), Topic12(Club lounge), Topic20(Residence)
	Not defined (Irregular)	Group4 Topic3(Pool/Gym), Topic16(Room facilities)		
	Negative 1, 2 point group(+) 4, 5 point group(-)			Group5 Topic1(Noise/Smell), Topic4(Membership), Topic6(Breakfast), Topic8(Reception), Topic14(Food & Beverage), Topic17(Room environment)

Groups 1 and 5 are the groups in which the distribution of reviews according to the score and the direction of the score match. Group 1 is composed of positive topics that express satisfaction with accommodation/staff, location/landscape, staff service, experience, ancillary facilities, location/neighborhood, transportation/accessibility, and price. Group 5 is composed of topics that show dissatisfaction in terms of noise/smell, benefits, breakfast, customer service, food and beverage, and temperature/sleep. In the remaining group 2, group 3, and group 4, the positive/negative direction of topics are unclear and include guest room facilities and club lounges, residence, and swimming pool/gym.

The groups that did not match the review frequency and direction according to the sentiment scores and scores is divided into three groups. Group 4 is a case in which it is difficult to measure the sensitivity by appearing evenly in the high and low scores without a specific direction. Looking at the representative words of Topic 3 and Topic 16 in Group 4, it is difficult to see the bias phenomenon according to the scores as it mainly consists of the names of hotel facilities. In addition, Group 2 is composed of topics to be difficult to grasp positively or negatively with sentiment scores. In this case, the direction of the review distribution according to the score range is unclear compared to other topics, so the results from the DMR seem reasonable. Finally, Group 3 consists of topics that frequently come up with high scores in reality, however, the sentiment score is not determined to be positive. To infer the cause of these results, reviews usually consist of various topics. If there are many cases where reviews related to Group 1 given a high score have the same content as the topic in Group 3, there may be cases where it cannot be judged. In fact, Topic 7, Topic 12, and Topic 20 are composed of keywords that simply refer to facilities, and are highly likely to be reflected in the reviews related to the positive topics of Group 1. Therefore, the results for Group 3 and Group 4 suggest that it is feasible to mark down the reliability of the sentiment analysis results through DMR.

This study tried to increase the credibility of the research by using both the sentiment scores and the frequency of topic-related reviews to designate only the intersection of topics with a clear sentiment direction as sentiment topics. In other words, for the hotel properties of Group 2, Group 3, and Group 4, it is possible to be determined that the results of DMR are not valid, whereas sentiment scores from DMR and actual review frequency are matched in the hotel properties of Group 1 (positive) and Group 5 (negative), and it is possible to identify the causes of the satisfaction and dissatisfaction of tourists who have stayed in 5-star hotels in Korea.

4.4 DERIVATION OF PRACTICAL IMPLICATIONS

Implications for 14 topics whose validity has been verified are as follows. First, it is interesting to note that there are several detailed topics for one hotel property. For instance, topics related to 'Staff Service' are Topic 2 and Topic 9. However, Topic 2 (3.80 points) is satisfaction with the overall staff service, and Topic 9 (3.23 points) is the staff's coping in specific situations. The degree of affirmation of these two topics is different. Kindness in a specific situation is important, but consistent and overall service is more important than a customer's kindness in a specific situation. In this context, with regard to 'Location', Topic 1 (3.66 points) is about a convenient location close to the surrounding commercial area, and Topic 5 (3.15 points) is about a location with a good view. A location close to the surrounding commercial district shows a higher degree of positivity. In general, when the city view is good, it is close to the surrounding commercial area and transportation is convenient, but when the natural scenery is good, it is relatively far from the surrounding commercial area. In the case of hotels pursuing natural scenery, actions to increase accessibility (Topic 18, 3.63 points) through transportation such as shuttle service sufficiently overcome these shortcomings.

In addition, ancillary facilities (Topic 11, 4.00 points) are good hotel resources to bring out customer satisfaction. It is important to link this resource to provide customers with special experiences through the hotel's promotions or concepts (Topic 10, 3.86 points). Hotels need to continuously develop promotions to make the most of these resources. Satisfaction with the price compared to facility/location (Topic 19, 3.63 points) is seen a lot in business customers. A large number of mentions of price can make price an important attribute in hotel selection, and it is important to appeal to and provide services that make customers feel reasonable.

Regarding factors influencing negative customer evaluation, for Topic 1 (1.78) and Topic 17 (1.72) on 'Room condition', noise from inner floors and doors, smoke/smell from toilets or elevators, inconveniences in sleeping related to the temperature and air in the room are all caused when careful observation and thorough management are not carried out in the management of room facilities. The facility/room management department should prevent such inconvenience by checking out the conditions in advance that may be basic and sensitive to customers other than these sounds, smells, and temperatures.

Topic 6 (2.48) of 'Food and Beverage' is different from the service area of food and beverage in Topic 14 (2.15) because customers were disappointed with the food itself. Regarding Topic 6, breakfast-related menus (American/Korean) are structured to satisfy a variety of customers, and strict quality control is required for ingredients that are important for freshness, such as juice, salad, and eggs. Regarding Topic 14, the failure to respond to customers at the point of service provision rather than the quality of the food itself and dissatisfaction with the food is complex. In food and beverage establishments, services are provided in an environment that is closer to the customer than the guest room. That is, while the time spent in the room is long, service requests are relatively rare, and the food and beverage establishment is a place where service requests are frequently generated while the time in which food service is provided is short. Given that these places are hotel properties that strongly affect customer dissatisfaction, they need to respond more sensitively to customer requests and respond appropriately to customer requests. The quality must be kept good.

In the case of complaint topic 4 (2.44) related to member benefits, dissatisfaction occurs when the service expected by the customer is not provided due to the circumstances of the hotel. Therefore, it is necessary for member customers to clearly recognize the member benefits provided by the hotel and the scope of the services provided. Topic 8(2.92) affects relatively weak negative evaluation. It is related to complaints that occurred because a request was made to the manager or the desk at check-in or other circumstances due to inappropriate response. It is more like a complaint in a specific situation rather than an overall service failure. It is judged that this had a relatively insignificant effect on the negative evaluation. What is interesting about hotel quick evaluation that affects positive/negative evaluations is that topics affecting negative evaluations have stronger and clearer directions than topics related to positive evaluations. Hotel attributes appearing in negative evaluations appear more clearly in 1 and 2-point reviews. They are elements that can be removed more clearly when improvements are made through feedback, compared to attributes that appear evenly from 1 to 5. In addition, it was found that there was no effect on specific scores with respect to swimming pool/gym (Topic 3), in-room facilities/tools (Topic 7), and anniversary visit (Topic 15). These are factors that do not strongly influence positive or negative evaluations, and when reflecting on customer reviews from the point of view of hotel workers, they should give priority to opinions on topics that strongly influence other positive and negative opinions.

V. CONCLUSION

In this study, topic modelling, a type of text analysis, was used to identify hotel attributes that affect scores from online reviews of 5-star hotel guests in Korea. In addition, in order to overcome the limitations of topic modelling studies using existing LDA, which are 'high-frequency word processing' and 'subjective topic number setting', 'IDF word weight application', 'identification of the optimal number of topics through concordance scores', 'DMR topic modelling' are utilized and achievements have been made.

The academic implication of this study is that the relationship between the existing hotels attributes and customer sensibility, as presented above, could be identified based on hotel reviews and review scores. In addition, it was confirmed that by specifying only initial setting values such as t and p during the analysis process, it was possible to automate the process from collecting reviews to the derivation of results.

As a result of the analysis, a total of 20 topics were derived, and the sentiment scores were calculated based on the weight of each topic's score. Those with more than 3 points were classified as positive topics and those with less than 3 points as negative topics, so there were 10 positive topics, 9 negative topics, and 1 unclassified topic. Afterward, by tracking the frequency of reviews related to each topic, it was examined whether the reviews are intensively distributed in the appropriate score range (negative: 1-2 points, positive: 4-5 points). 8 positive topics and 6 negative topics, as the distribution of reviews by score and the result of sentiment score match, were identified, and practical implications were presented only for 14 topics with the same emotional score and review distribution by the score for hotel attributes and contents that affect customer sensibility. In addition to methodological improvements, this study also yielded other advantages. Among the various hotel evaluation attributes (service, location, price, facilities, etc.), only those that affect the review score are derived, and that one attribute appears as a detailed character. These results help hotel practitioners to set priorities for intensive management of hotel attributes and to pursue directions for attributes.

Based on the DMR topic modelling of online customer reviews for 5-star hotels in Korea, sentiment analysis that divided customers' opinions by major topics was attempted. We try to suggest a direction for future research, focusing on some limitations drawn during the conduct of this study.

First, the score is a sequence measure with continuity, but DMR topic modelling sees the score as a nominal measure. For this reason, it is judged that both the positive and negative topics identified in the DMR did not lead to valid conclusions. Therefore, it will be possible to provide better results if the topic affecting the score is identified by applying it to topic modelling that can later be input as a ranking scale. Second, this study was analysed based on English reviews of foreign customers, which are easy to analyse morphemes. In the future, to conduct research on domestic customers of domestic hotels, it is necessary to collect reviews from various hotel reservation platform sites in addition to TripAdvisor to secure a sufficient number of reviews, and it will be necessary to secure a Hangul morpheme analyser with excellent performance. Third, although this study analysed travel types without distinction, it would be an interesting study to analyse the differences in sensibility for each travel type through the segmentation of travel types. Finally, it is an important task for hotel executives to identify which hotel properties have a competitive advantage over other competitive hotels in the same market and which ones do not, rather than the entire domestic five-star hotel. Therefore, it will be necessary to study a specific hotel compared to other hotels to find out where the hotel's strengths (positive topics) come from and its weaknesses (negative topics) that are lagging behind its competitors.

REFERENCES.

[1]. Baek, H., Ahn, J., & Choi, Y. (2012) Helpfulness of online consumer reviews, Readers' objectives and review cues, *International Journal of Electronic Commerce*, Vol.17, No.2, pp. 99-126

[2]. Blei, D. M., Andrew, Y., & Michael, I. (2003) Latent dirichlet allocation, *The Journal of Machine Learning Research*, Vol. 3, pp. 993-1022.

[3]. Cao, Q., Duan, W., & Gan, Q. (2011) Exploring determinants of voting for the "helpfulness" of online user reviews, *A text mining approach, Decision Support Systems*, Vol. 50, No. 2, pp. 511-521.

[4]. Chung Hun Lee. (2014).The Impact of Negative Online Review Consensus and Organizational Response Strategy on Potential Consumers` Attribution Process: The Case of Online Restaurant Review Websites. *Korean Journal of Hospitality & Tourism*, Vol.23, No.1, pp. 235-247.

[5]. Deerwester, S., Dumais, S. T., Furnas, G. W., Landauer, T. K., Harshman, R. (1990) Indexing by latent semantic analysis, *Journal of the American Society for Information Science*, Vol.41, No.6, pp391-407.

[6]. Han, H. J., Mankad, S., Gavirneni, N., & Verma, R.(2016), What guests really think of your hotel: Text analytics of online customer reviews, *Cornell Hospitality Report*, Vol.16, No.2, pp.3-17.

[7]. Huang, L., Ma, J., & Chen, C.(2017), Topic detection from micro-blogs using T-LDA and perplexity, 24th Asia-Pacific Software Engineering Conference Workshops (APSECW), pp.71-77.

[8]. Hyung Joon Kim, Yong Il Kim. (2014). The influence of building trust by characteristics of e-WOM based on application of food information service on behaviour intention. *Korean Journal of Hotel Administration*, Vol.23, No.1, pp. 45-64.

[9]. Hye Min Yun, Eun Jung Choi. (2019). A Study on Sentiment Analysis of Media and SNS response to National Policy : focusing on policy of Child allowance, Childbirth grant . *Journal of Digital Convergence*, Vol.17, No.2, pp.195-200.

[10]. Jun-Sik Sim, Hyoung-Joong Kim.(2017).A Searching Method for Legal Case Using LDA Topic Modelling. *J. the Institute of Electronics and Information Engineers*, Vol.54, No.9, pp.67-75.

[11]. Ki Young Min, Hoon Tae Kim, Yong Gu Ji.(2014).A Pilot Study on Applying Text Mining Tools to Analyzing Steel Industry Trends : A Case Study of the Steel Industry for the Company "P". *The Jounal of Society for e-Business Studies*, Vol.19, No.3, pp.51-64.

[12]. Kim, D., Yu, S. (2017) Hotel review mining for targeting strategy: Focusing on Chinese free independent traveler. *Journal of Theoretical and Applied Information Technology*, Vol.95, No.18, pp.4436-4445.

[13]. Kim, N., Kim, M. (2010) The effects of word-of-mouth information quality on word-of-mouth effects in online tourism context, *Journal of Tourism Management Research*, Vol. 19,No.4, pp. 59-79.

[14]. Kim, Y, S., Jung, K. Y. (2009) Hotel brand blog characteristics influencing purchase intention and word-of-mouse communication : blog attitude as an intervening variable - Focused on blog users in their 20's and 30's, *Journal of Tourism Management Research*, Vol.18, No.3, pp.1-22.

[15]. Lee, Hye Jin, Choi Jayoung. (2019) Sentiment Analysis of Twitter Reviews toward Convenience Stores Customer in Korea, *Academic Society of Global Business Administration*, Vol.16, No.4, pp.143-164.

[16]. Lee, Jong-Ho, Lee, In-Soo, Jung, Kyeong-Soo, Chae, Byeong-Hoon, and Lee, Joo-yeoun. (2017) Patents and Papers Trends of Solar-Photovoltaic(PV) Technology using LDA Algorithm, *Journal of Digital Convergence*, Vol.15, No.9, pp.231–239..

[17]. Li, A., Huang, Y., & Axhausen, K. W. (2018) Bicycling accessibility based on dock-less bicycle-sharing data, Arbeitsberichte Verkehrs- und Raumplanung, pp.1377.

[18]. Lim, Young-hee, Kim, Hong-bumm.(2019) A study on the sentiment analysis using big data of hotels online review. *Korean Journal of Hospitality & Tourism*, Vol.28, No.7, pp.105-123.

[19]. Minjung Kwak, Jiyu Choi, Sohyun Park. (2010) Sentiment analysis of online reviews for evaluating customer satisfaction on hotel service attributes. *Journal of Tourism Management Research*, Vol.23, No.4, pp.1-25.

[20]. Newman, D., Lau, J. H., Grieser, K., & Baldwin, T. (2010). Automatic evaluation of topic coherence, HLT'10: Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics, pp.100-108.

[21]. Olkin, I., & Rubin, H. (1964) Multivariate beta distributions and independence properties of the wishart distribution, *The Annals of Mathematical Statistics*, Vol.35, No.1, pp.261-269.

[22]. Park, Seung Jeong, John, Jino, Seon Woo Kim, Kim, Sung Tae. (2017).A Big-Data Analysis of Presidential Issue Ownership of two Prestigious Korean Newspapers : Focusing on LDA(latent Dirichlet allocation) Topic Modelling. *The Journal of Political Science & Communication*, Vol.20, No.3, pp.25-55.

[23]. Ren Gang, Hong Tae-ho, Niu Hanying, Park Ji-young. (2018) Multi-Topic Sentiment Analysis using LDA for Online Review. *The Journal of Information Systems*, Vol.27, No.1, pp.89-110.

[24]. Roder, M., Both, A., & Hinneberg, A. (2015) Exploring the space of topic coherence measures, WSDM'15: Proceedings of the 8th ACM International Conference on Web Search and Data Mining, pp. 399-408.

[25]. Sim Young-seok, Kim Hong-bum. (2018) An Analysis on the Structural Relationship between Destination Personality and Traveler Rating with Word2Vec Technique based on Online Review Big Data. *Journal of Tourism Sciences*, Vol.42, No.8, pp.165-189.

[26]. Sung Taek Lim, Won Seb Cho. (2011) The Effects of Business Blog Information Characteristics Influencing on Electronic Word-of-Mouth In the Food Service Industry: Emphasis on Trust Transference. *Korean Journal of Hospitality & Tourism*, Vol.20, No.5, pp.165-180.

[27]. Tian, X., He, W., Tao, R. & Akula, V. (2016) Mining online hotel reviews: A case study from hotels in China, In Proceeding of 22nd Americas Conference on Information Systems.