Research Trend Analysis on Customer Satisfaction

in Service Field Using BERTopic and LDA*

Woo-Ryeong YANG[1], Hoe-Chang YANG[2]

Abstract

Purpose: The purpose of this study is to derive various ways to realize customer satisfaction for the development of the service industry by exploring research trends related to customer satisfaction, which is presented as an important goal in the service industry. **Research design, data and methodology:** To this end, 1,456 papers with English abstracts using scienceON were used for analysis. Using Python 3.7, word frequency and co-occurrence analysis were confirmed, and topics related to research trends were classified through BERTopic and LDA. **Results:** As a result of word frequency and co-occurrence frequency analysis, words such as quality, intention, and loyalty appeared frequently. As a result of BERTopic and LDA, 11 topics such as 'catering service' and 'brand justice' were derived. As a result of trend analysis, it was confirmed that 'brand justice' and 'internet shopping' are emerging as relatively important research topics, but CRM is less interested. **Conclusions:** The results of this study showed that the 7P marketing strategy is working to some extent. Therefore, it is proposed to conduct research related to acquisition of good customers through service price, customer lifetime value application, and customer segmentation that are expected to be needed for the development of the service industry.

Keywords : Research Trend, Service Field, Customer Satisfaction, BERTopic, LDA.

JEL Classification Code : J50, L10, L50, L80, M30.

1. Introduction

The service industry is the industry that people face the most in their daily lives. The service industry is a general term for industries that produce services based on the development of primary and secondary industries, and provides services in various industries such as commerce, finance, insurance, transportation, telecommunications, tourism, and advertising as well as medical, legal, and education.

The added value and job creation effect of the service industry accounts for 60% of the added value of the Korean

economy and 70% of jobs, and is known to greatly exceed the added value and job creation effect of other industries such as manufacturing and construction (MOEF, 2021). For this reason, if the industries that have driven the country's economic growth have been manufacturing and exports, the service industry is emerging as a new growth engine.

However, according to a report by KPC (2018), the

value-added share of the service industry in Korea has been stagnant at 60% for the past 10 years, and the gap with advanced countries such as the US 79.8% and Japan 69.6% continues at about 10-20%p. As shown in <Table 1>, the share of added value across all sectors by industry was found to be lower than the OECD average.

Table 1: International Comparison of Value Added Ratio by Industry

	Service Industry	Comparison by industry				
		wholesale and retail, food	information communication	finance, insurance	professional science	public health
Korea	60.9	13.7	4.6	6.0	9.5	16.5
G7	75.1	18.4	5.4	5.3	11.3	18.8
OECD	70.9	20.0	5.0	5.6	9.9	17.7

Source: KPC(2018)

At the same time, the share of employment in the service industry was also confirmed to be 70.8% as of 2019, showing a gap of 5-10%p compared to major countries such as the United States (79.9%) and Japan (72.8%). By industry, the proportion of wholesale and retail, food and lodging, etc.,

which are closely related to life was high. However, it was confirmed that the share of employment in high value-added industries such as professional, scientific, and health care was low (refer to <Table 2>).

Table 2: International comparison of employment share by industry

	Service Industry	Comparison by industry				
		wholesale and retail, food	information communication	finance, insurance	professional science	public health
Korea	69.8	27.5	3.1	3.1	9.0	18.7
G7	76.7	25.2	3.2	3.2	12.7	23.6
OECD	73.8	25.4	3.3	2.8	11.7	24.3

Source: KPC(2018)

Due to the high proportion of industries close to life, the labor productivity of the service industry in Korea is only 50.3% of that of the manufacturing industry as of 2018. Even in international comparison, it is known to be at the bottom of the list, ranking 28th out of 33 OECD countries (MOEF, 2021).

In particular, service consumption which has a large proportion of face-to-face contact has greatly reduced the productivity of the service industry due to the bad news of Corona 19. On the other hand, it has been confirmed that food and lodging, wholesale and retail, art, sports, leisure, and transportation, which are relatively greatly affected by social distancing, suffer enormous damage.

The reason this study focuses on the service industry is as follows. First, it is because the service industry affects the survival of each industry according to the degree of adaptation capacity of the corona pandemic. Second, the development of new technologies such as big data and AI (artificial intelligence) is changing the status of the traditional service industry. Third, demand for personalization is expanding due to low birth rate, expansion

of single-person households, and aging population. Fourth, as service trade is activated, global competition in the service industry is intensifying.

In particular, the concept of service, defined by the American Marketing Association(AMA) as an activity provided for sale in the past or provided along with product sales, is an activity in which services are intangible and relational goods and activities that create something out of the absence of exercising human capital (Kim, 2018), it is necessary to consider the nature of the service because the boundary between service and product is breaking down as well as industry boundary.

Therefore, it is judged that the core keyword that the service industry has dealt with so far is customer satisfaction, and this study tried to confirm the research trend related to customer satisfaction in the service industry. This is because the services used as supplementary products by companies producing products in the past have increased interest in the role of customers due to the maturation of the industry. In addition, as the company's service proportion increases, customer satisfaction which is the result of customer

response cannot help but be treated as important.

The purpose of this study was to redefine the perspective on services that have passed the maturity period through the analysis of research trends on customer satisfaction in the service industry.

To this end, in this study, after searching for the keyword 'customer satisfaction' through scienceON as of October 7, 2022, and applying 'service' as the detailed search keyword, a total of 7,252 papers were searched.

Among them, duplicate data were removed and 1,456 papers with English abstracts were used for analysis. Python 3.7 was used as an analysis tool, and word frequency analysis, word clouding, and word co-occurrence frequency analysis were performed. In addition, after deriving major topics through BERTopic and LDA which is one of unsupervised learning among machine learning, OLS regression analysis was conducted to analyze topic trends by year.

The results of this study are expected to provide insight into which areas of research should be added for the development of the service industry in the future by exploring which areas have been mainly conducted in research related to customer satisfaction in the service industry so far. In addition, it was expected to provide various clues about the role of the service industry and the direction beyond customer satisfaction under the service dominant logic.

2. Literature Review

2.1. Service and Service Industry

According to Krajewski et al. (1999), the manufacturing industry has the characteristics of a capital-intensive industry in which output is tangible and durable, inventory accumulation is possible, customer contact is relatively small, large-scale facilities are required, and quality measurement is relatively easy. However, since the service industry is shapeless and storage is impossible, inventory cannot be accumulated, and it is labor-intensive and difficult to measure quality in small-scale facilities (Krajewski et al., 1999).

Looking at the definition of traditional service in previous studies, it is as follows. Grönroos (1990) defined service as a series of essential activities, generally intangible, but the interaction between a customer and a service provider, or between physical resources and a service provider to solve a customer's problem (Grönroos, 1990). On the other hand, Kotler (1991) defined a service as any performance or action that one party can provide to the other party which is intangible and not linked to ownership, and that the production of a service may or may not be linked to

a physical product (Kotler, 1991). However, in the traditional definition of service, it can be seen that a service is defined as an element added to the characteristics of a product.

Compared to the traditional definition, Katzan (2008) defined service as providing support and expertise through interactions between suppliers and customers to create value in business, education, individual and government endeavors (Katzan, 2008).

In addition, Nam et al. (2008) also defined service as creating new value through interactions between service providers and customers. In other words, service is a coordinated joint effort in a specific, time-space environment to maximize customer experience (Nam et al., 2008). This shows that service as an element added to product characteristics in the past is expanding into a major factor creating value through customer experience.

According to the data provided by the Korean Standard Service Quality Index (KS-SQI), the service industry is divided into four areas industries according to the degree of labor intensity and customer interaction and customerization such as service factory, service shop, mass service, and professional service (Schmenner, 1996).

Among the four areas classified by Schmenner (1996), a service factory has low labor intensity and low interaction with customers, so it is an airline, cargo trucks, hotels and recreational facilities that supplies standardized services in large quantities that can produce a flow similar to that of a factory. Service shop has low labor intensive but have high interaction with customers, providing highly customized services that require a lot of capital investment such as hospitals, auto repair shops, and other maintenance companies. Mass service has low customer interaction but high labor intensity. Therefore, when providing services, standardized services are provided which require a high proportion of the labor force required for this, such as wholesale and retail businesses, and include large-scale lectures among educational services. Finally, professional service is an industry in which both interaction with customers and labor intensity are high. Therefore, it means that professionally trained professionals such as doctors, lawyers, accountants, and architects provide various services for customerization to customers.

According to the '2019 provisional results of the service industry survey' provided by KOSIS, the number of service businesses in 2019 was 2.977 million, the number of employees was 12.53 million, and sales were 2,188 trillion won (KOSIS, 2020). Compared to 2006, the number of businesses increased by 750,000, the number of employees increased by 5.1 million, and sales increased by KRW 1,371 trillion, indicating that it is a major industry that plays a pivotal role in the Korean economy (See Appendix 1).

2.2. Customer Satisfaction

Satisfaction is defined as the extent to which a consumer's emotion or feeling arises from an evaluation of the discrepancy between the consumer's expectations and the service provider's performance (Sumaedi et al., 2016). Service quality and customer satisfaction are judged as very similar concepts in the minds of customers, but they have quite different characteristics (Cronin & Taylor, 1992). Because service quality is a specific measure of the extent to which a provided service meets customer expectations (Joewono & Kubota, 2007), satisfaction is an overall emotional response to the perceived discrepancy between previous expectations and perceived performance (Oliver, 1980, 1999). In other words, customer satisfaction is different from perceived service quality because it is the result of a customer's overall and cumulative evaluation of not only the intrinsic aspects of the service but also the external aspects of the service such as image and price (Bakti et al., 2020). In previous studies (e.g. de Oña et al., 2015; Oliver, 1997), the perception of service quality has a positive effect on both customer satisfaction and behavioral intention, and It was said to have a positive effect on satisfaction as a post-cognition has the same behavioral criteria as intention and actual behavior (Fornell et al., 1996; Wen et al., 2005). Customer satisfaction is one of the positive goals pursued by the service industry due to its long-term benefits such as positive word of mouth, customer loyalty, and sustainable profitability (Greenwell et al., 2002; Liu & Jang, 2009).

Measurement of customer satisfaction in the service industry is composed of a one-dimensional measure of overall satisfaction with the result of comprehensively judging all interactions and contact points with service organizations (Chen & Tsai, 2008; Yang & Peterson, 2004). However, in order to accurately measure customer satisfaction, the following situations should be considered as Oliver (1999) pointed out. First, since there may be individuals who pay for a product or service but do not actually use it, it should be limited to measuring the satisfaction of actual users rather than buyer satisfaction which may include non-users. Second, because satisfaction is a feeling and is a short-term attitude that can easily change depending on the situation, it is different from observable behaviors such as product selection, dissatisfaction, and repurchase that are settled in the user's mind. Third, satisfaction generally has thresholds at both lower levels such as insufficient or unsatisfied, and higher levels such as over-satisfaction. This means that consumers may be less satisfied if they get too much of a good thing. However, there are cases in which the meaning of a high threshold value is ignored while focusing only on a low threshold value in normal cases. In this way, it is necessary to consider many things in order to understand the relationship between the level of customer satisfaction and its influence.

In this study, we determined that customer satisfaction is the most important factor in securing competitiveness in the service industry, and tried to identify research trends related to customer satisfaction in the service industry so far.

2.3. Topic Modeling

Topic modeling is one of the most powerful tools of text mining for data mining, latent data retrieval and finding relationships between data and text texts, and is applied to various fields such as software engineering, political science, medicine and linguistics (Jelodar et al., 2019; Yang, 2022).

NMF (Non-Negative Matrix Factorization) (Févotte & Idier, 2011) or LDA (Blei et al., 2003) have various advantages in text mining, but they require the number of topics, stopword list, and morpheme analysis to obtain optimal results. In addition, since it has a weakness that the semantic relationship between words is ignored, text embedding technology is popular in the field of natural language processing (Yang, 2022). Among them, BERTopic is to generate word and sentence vector representations suitable for context (Yang, 2022). The semantic properties of these vector representations allow similar texts to encode the meaning of text in an approximate way in vector space (Grootendorst, 2022). BERTopic, which utilizes clustering techniques and class-based transformation of TF-IDF to generate consistent topic representations, utilizes BERT-based embedding and c-TF-IDF word weights in the text embedding step. And it is a topic modeling technique that finds latent meaningful themes in text by text clustering suitable for each domain (Ko et al., 2022; Yang, 2022).

On the other hand, LDA (Latent Dirichlet allocation) is the most used topic modeling technique in studies using topic models so far because it is highly applicable not only to structured data but also to unstructured data (Blei et al., 2003). LDA-based topic modeling has been applied to natural language processing, text mining, social media analysis, and information retrieval, and has characteristics that not only excel at demonstrating discrete data, but also provide effective access to find hidden structures and meanings in enormous information (Jelodar et al., 2019; Yang, 2022).

In this study, we tried to derive topics related to customer satisfaction research in the service industry through BERTopic. However, due to the nature of BERTopic, when there are many outlier documents that are not classified as topics, excluding them may have limitations in confirming the overall research trend. Therefore, in this study, as used in the research methods of Yang (2022) and Yang and Yang (2022), outlier documents excluded from the BERTopic result topic were tried to derive the topic again through LDA.

3. Research Procedure

In order to confirm research trends related to customer satisfaction in the service industry, academic papers published so far were extracted and analyzed through the research procedure shown in <Figure 1>. For analysis, word frequency analysis, word clouding, and word co-occurrence frequency analysis were used to confirm which words were frequently used in the study, and topics were derived through BERTopic, and LDA was performed for outlier documents using Python 3.7. Afterwards, OLS regression analysis was conducted to analyze yearly trends in total topics.
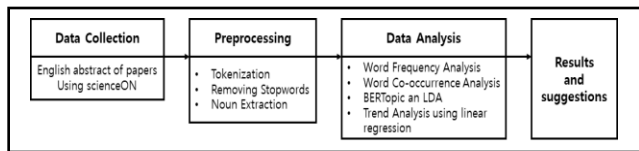

Figure 1: Research Procedure

3.1. Data Collection

As of October 7, 2022, a total of 7,252 papers were searched as a result of searching scienceON (https://scienceon.kisti.re.kr) with the keyword 'customer satisfaction'. As a result of conducting a detailed search with the keyword 'service' and removing duplicate data, 2,960 papers were derived, and 1,456 papers with English abstracts were extracted and used for analysis. When looking at the total number of articles or papers to be analyzed by year group, it was found that the increase after 2000 decreased from 2015, it suggesting that interest in customer satisfaction which has been dealt with heavily in the service industry is changing. The number of publications by year and group is presented in Appendix 2.

3.2. Data Preprocessing

Data preprocessing which is a process of data refinement has the advantage of minimizing noise, distortion, and deviation as well as reducing data measurement errors. In this study, data preprocessing was performed to derive topics more elaborately by deleting meaningless words or stopwords included in the abstract. Specifically, words such as purpose, methodology, aim, conclusion, and results that frequently appear in abstracts, and words such as correlation, regression, SPSS, AMOS, and coefficients presented in statistical analysis were removed as stopwords, and a stopword dictionary was used. Therefore, words such as special symbols and numbers were removed, and customer and satisfaction were also set as stop words.

4. Empirical Analysis Results

4.1. Results of Word Frequency Analysis

As a result of word frequency analysis, it was confirmed that words such as quality (3,205), intention (1,087), relationship (1,072), and loyalty (962) were frequently used. These results indicate that service quality, customer relationship, and loyalty are keywords of interest in research so far. As a result of the frequency analysis, the top 20 keywords are presented in Appendix 3, and the word clouding results, which are expressed differently in size by word frequency, are presented in <Figure 2>.

Figure 2: The result of wordclouding

4.2. Results of Co-occurrence Frequency Analysis

In order to confirm the relationship between words, after constructing a dictionary that counts the frequency of co-occurring words, an analysis was conducted in consideration of connection centrality. As a result of co-occurrence frequency analysis, it appeared to be words such as quality, relationship, intention, influence, loyalty, etc. are related to words related to service workers, environment, system, performance, etc. The top 10 pairs of word co-occurrence frequency are presented in Appendix 4, and the visualization results are presented in <Figure 3>.
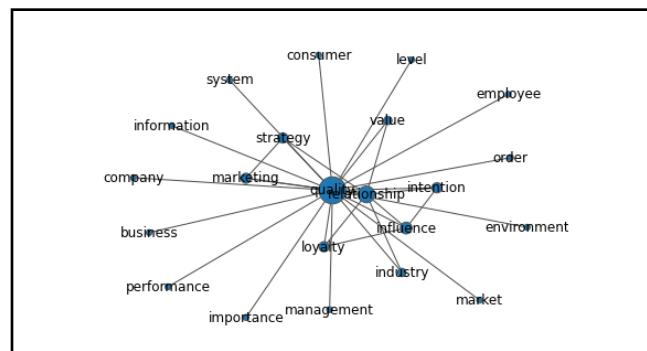

Figure 3: Co-appearance frequency visualization result (connection frequency = 100 times)

4.3. Results of Topic Modeling

As a result of conducting BERTopic, a total of 15 topics were derived. However, as a result of checking the derived topics, many topics consisted of about 10 documents, and as a result of confirming the similarity between topics, many similar topics occurred. Therefore, the number of papers composed of each topic was limited to 50, and as a result of conducting BERTopic again, 7 topics and 736 outlier documents were derived.

<Topic 1> derived through BERTopic was composed of keywords such as restaurant, food, foodservice, and menu, and was judged to be related to food such as restaurants, and was named 'Catering service'. <Topic 2> consisted of keywords such as loyalty, brand, justice, and trust. This result was judged as the degree of fairness perceived by customers in relation to the product or service and named it 'Brand justice'. <Topic 3> consisted of keywords such as mall, shopping, internet, site, online, etc., and was judged to be related to online shopping, and was named 'Online shopping'. <Topic 4> is composed of keywords such as hotel, job, employee, and labor, and it was judged to be related to hoteliers and other workers in the lodging industry, and it was named 'Accommodation service'. <Topic 5> is composed of keywords such as system, center, call, evaluation, and network, and it is judged to be related to inbound or outbound call centers, and it is named 'Call center service'. <Topic 6> is composed of keywords such as logistics, port, transportation, and delivery, and was judged to be related to the delivery of products or services, and was named 'Logistics delivery service'. <Topic 7> is composed of keywords such as airline, passenger, cabin, airport, etc., and was determined to be related to airlines, and was named 'Air transportation service'.

As a result of conducting LDA on 736 outlier documents, the highest coherence score was 0.2408 when there were 4 topics, so 4 topics were added. <Topic 8> is composed of keywords such as job, employee, company, industry, etc., and it was judged to be related to general service workers, so it was named 'Service workers'. <Topic 9> is composed of keywords such as consumer, loyalty, trust, and value, and it was judged to be related to the relationship between the service company and the customers who use it, and named it 'CRM (Consumer Relationship Management)'. <Topic 10> is composed of keywords such as loyalty, value, brand, and marketing, and it was judged as brand performance due to service company marketing and named it as 'brand value' recognized by customers. Lastly, <Topic 11> is composed of keywords such as user, facility, system, and performance, and was named 'facility equipment' because it was judged to be an evaluation of physical evidence provided by the service industry recognized by users.

Each topic name and the top 10 keywords for each topic

are presented in Appendix 5.

4.4. Trend Analysis Results of Each Topic

By calculating the ratio of each topic assigned to a specific paper through topic modeling and calculating the average ratio by year of each topic using the calculated ratio and the year of publication of each paper, it is called dynamic topic modeling that the trend of the topic can be identified (Yang, 2022; Yang & Yang, 2022). Dynamic topic modeling has the advantage of being able to track the appearance of each topic over time because the topic ratio is flexible over time (Blei & Laffety, 2006; Yang, 2022). The trend analysis results for each topic are visualized in <Figure 4>.
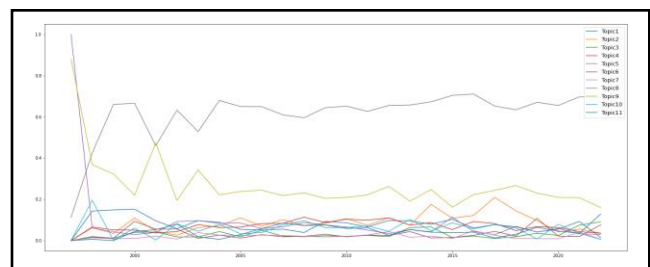

Figure 4: Visualization results of Trend Analysis

The result of the trend analysis in <Figure 3> has the disadvantage that it is difficult to clearly predict which topics are receiving more attention, although it is possible to confirm the degree of liquidity in which topics are attracting interest by year. Therefore, as in the studies of Yang (2022) and Yang and Yang (2022), the independent variable was set to the year of publication of the thesis, and the dependent variable was set to the average weight of the topics in that year, and then OLS regression analysis was performed (Yang, 2022; Yang & Yang, 2022). If the regression analysis result is a statistically significant positive (+) coefficient, it can be interpreted as a hot topic, and a negative (-) coefficient can be interpreted as a cold topic. If not, it can be concluded as a neutral topic (Griffiths & Steyvers, 2004; Yang, 2022; Yang & Yang, 2022).

The results of the regression analysis are as follows. It was confirmed that the researchers' interest in each was relatively focused such as <Topic 2> 'Brand justice' (coeff. = 0.0024, $p<.05$), <Topic 3> 'Online Shopping' (coeff. = 0.0012, $p<.05$) and <Topic 8> 'Service Workers' (coeff. = 0.0093, $p<.01$) at the 95% confidence level. However, 'CRM' (coeff. = -0.0099, $p<.01$) in <Topic 9> appeared to be a topic with low interest. On the other hand, 'facility equipment' (coeff. = 0.0014, $p=.063$) in <Topic 11> classified as physical evidence of the service industry was identified as a topic of attention at the 90% confidence level.

These results can be interpreted as the researcher's interest in the justice of service industry brands and online shopping perceived by consumers at a time when non-face-to-face transactions are recently activated due to COVID-19, etc., and interest in facilities and equipment is also receiving some attention. However, the relatively declining interest of researchers in CRM is worrisome and requires close examination. This is because it contradicts previous studies (e.g. Ishaq et al., 2014; McDougall & Levesque, 2000; Ryu et al., 2012) that CRM is an important predictor variable that affects customer satisfaction in relation to customer value. The results of regression analysis on topics by year are presented in <Table 3>.

Table 3: Results of regression analysis on topics by year

	Topic Name	Coeff.	t-value	p-value	Trend
1	Catering service	-0.0016	-1.680	0.106	-
2	Brand justice	0.0024	2.097	0.047	Hot
3	Online shopping	0.0012	2.258	0.033	Hot
4	Accommodation service	0.0007	1.018	0.319	-
5	Call center service	-0.0088	-1.909	0.068	-
6	Logistics delivery service	0.0007	0.967	0.343	-
7	Air transportation service	0.0003	0.851	0.403	-
8	Service workers	0.0093	3.492	0.002	Hot
9	CRM	-0.0099	-3.106	0.005	Cold
10	Brand value	-0.0005	-0.456	0.652	-
11	Facility equipment	0.0014	1.948	0.063	Hot (mar.)

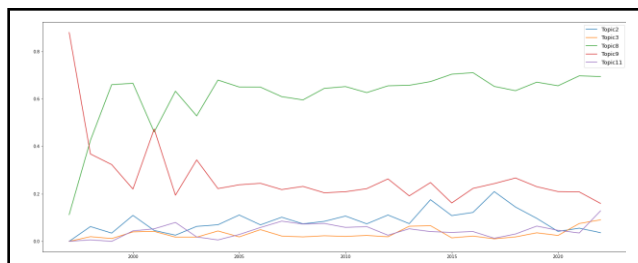Note) Coeff. means Coefficient, Mar. means Marginally

Figure 5: Visualization results of hot & cold topic

5. Conclusions

The purpose of this study is to redefine the perspective on the service industry that has passed its maturity through the analysis of research trends on customer satisfaction in the service industry, and to derive various ways to realize customer satisfaction for the development of the service industry. To this end, the results of analysis using a total of 1,456 papers extracted from scienceON are as follows.

First, as a result of word frequency analysis, words such as quality, intention, relationship, and loyalty appeared frequently. Also, in the co-occurrence frequency analysis, it was confirmed that words such as strategy, marketing, value, and industry were related to words such as quality, relationship, intention, influence, and loyalty. These results indicate that service quality, customer relationships, and loyalty have been studied in relation to customer satisfaction.

Second, as a result of BERTopic and LDA, a total of 11 topics named such as 'Catering service' and 'Brand justice' were derived. It can be judged that various results of researchers on 7Ps which are service industry marketing mix factors have been published such as product (catering service, accommodation service, call center service), promotion (brand justice, online shopping), place (logistics delivery service, air transportation service), people (service workers), process (customer value, brand value), and physical evidence (facility equipment).

Third, as a result of trend analysis through dynamic topic modeling and regression analysis, it was confirmed that researchers paid attention to 'brand justice', 'online shopping', 'service workers' and 'facility equipment'. However, it is noteworthy that researchers' interest in 'CRM' is relatively low. At a time when non-face-to-face transactions are activated due to the pandemic, it is judged appropriate to expand interest in 'brand justice' and 'online shopping'. In addition, when researchers apply the importance of 'people' as an important factor that determines customer satisfaction in the service industry, and 'space' which is being reinterpreted in the service industry, the increased interest in 'facility equipment' is considered natural. However, it is necessary to closely examine the declining interest of researchers in CRM, an important predictor of customer satisfaction.

Therefore, we would like to suggest the following research directions that researchers will be interested in in relation to customer satisfaction for the development of the service industry in the future.

First, it is necessary to pay attention to research related to service prices among the 7Ps which are relatively excluded from topic modeling. As shown in Lu and Kim's (2021) study that the price discount rate and involvement affect the effectiveness of advertising messages, service price is also related to involvement and is a major variable that affects customer satisfaction. Therefore, it is judged that research related to the degree of consumer awareness of the price provided by the service industry can provide various implications for practicing customer satisfaction.

Second, it is necessary to conduct various studies on the

emotional labor one of the 7Ps of 'people', especially workers. Workers in the face-to-face industry are often exposed to job burnout and stress because they are forced to make superficial emotions. Therefore, research from various perspectives that can control and overcome the negative emotions of workers in each service field should be continued.

Third, it is necessary to study customer satisfaction in the service industry by applying customer value, especially customer lifetime value (CLV) for the development of the service industry, although this study showed that researchers' interest in CRM is declining. Customer lifetime value (Berger & Nasr, 1998; Blattberg & Deighton 1996; Gupta et al., 2006) refers to the flow of revenue generated by a customer over the entire customer lifecycle (Chang et al., 2012). Therefore, by deriving a plan to develop various marketing activities from the perspective of customer lifetime value that applies customer value to customer relationship management (CRM), service companies will get various clues not only increase the level of customer satisfaction, but also accumulate customers as assets.

Fourth, there is a need for research on how to predict customer characteristics along with customer segmentation and develop appropriate marketing for service companies. In the past, customer segmentation was simply based on demographic criteria, but recently, various attempts have been made to classify customers into product-based, brand-based, and behavior-based groups. The reason for this is that segmentation is possible using big data. Therefore, when various segmentation criteria are applied and factors affecting customer satisfaction that reflect customer characteristics are derived, service companies are highly likely to secure competitiveness by practicing customer satisfaction through the provision of more sophisticated customized services.

Although this study presented various results and suggestions through the exploration of research trends related to customer satisfaction in service companies, there are some limitations of research trend studies, which need to be supplemented in future studies. First, it can be cited that the limitation of topic name determination which is a common limitation suggested by many studies such as Yang (2022) and Yang and Yang (2022) that performed topic modeling. This is because although experts in the field are involved in determining the topic name, it is not possible to reflect all the specific contents of the classified topic. Therefore, future research will need to study the standards and methods for determining topic names. Second, it can be pointed out that it is still experimental to extract additional topics through LDA for outlier documents other than the topic derived from BERTopic. As mentioned in this study, it is judged that reviewing outlier data close to 50% is a necessary procedure for identifying research trends, but

future studies need to confirm the effectiveness of these methods and the difference in effectiveness through a single analysis. Lastly, as a limitation of research trend analysis, although it suggests a future direction through trend identification, there is a problem that it is not specific. Therefore, future research will need an approach to provide more specific clues such as identifying trends and building a causal model.
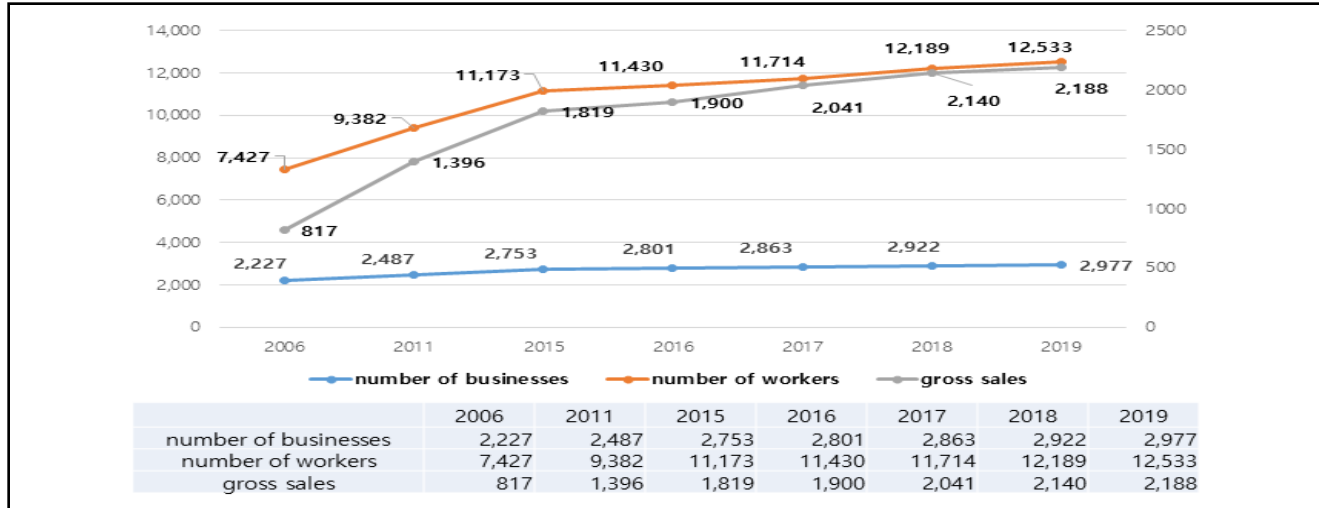
References

Bakti, I. G. M. Y., Rakhmawati, T., Sumaedi, S., Widianti, T., Yarmen, M., & Astrini, N. J. (2020). Public transport users' WOM: an integration model of the theory of planned behavior, customer satisfaction theory, and personal norm theory. *Transportation Research Procedia, 48*, 3365-3379.

Berger, P. D., & Nasr, N. I. (1998). Customer lifetime value: Marketing models and applications. *Journal of interactive marketing, 12*(1), 17-30.

Blattberg, R. C., J. Deighton. 1996. Manage marketing by the customer equity test. *Harvard Business Review, (July− August)* 136− 44.

Blei, D. M., & Lafferty, J. D. (2006, June). *Dynamic topic models.* In Proceedings of the 23rd international conference on Machine learning (pp. 113-120).

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of machine Learning research, 3*(Jan), 993-1022.

Chang, W., Chang, C., & Li, Q. (2012). Customer lifetime value: A review. *Social Behavior and Personality: an international journal, 40*(7), 1057-1064.

Chen, C. F., & Tsai, M. H. (2008). Perceived value, satisfaction, and loyalty of TV travel product shopping: Involvement as a moderator. *Tourism management, 29*(6), 1166-1171.

Cronin Jr, J. J., & Taylor, S. A. (1992). Measuring service quality: a reexamination and extension. *Journal of marketing, 56*(3), 55-68.

de Oña, R., Machado, J. L., & de Oña, J. (2015). Perceived service quality, customer satisfaction, and behavioral intentions: Structural equation model for the Metro of Seville, Spain. *Transportation Research Record, 2538*(1), 76-85.

Févotte, C., & Idier, J. (2011). Algorithms for nonnegative matrix factorization with the β-divergence. *Neural computation, 23*(9), 2421-2456.

Fornell, C., Johnson, M. D., Anderson, E. W., Cha, J., & Bryant, B. E. (1996). The American customer satisfaction index: nature, purpose, and findings. *Journal of marketing, 60*(4), 7-18.

Greenwell, T. C., Fink, J. S., & Pastore, D. L. (2002). Assessing the influence of the physical sports facility on customer satisfaction within the context of the service experience. *Sport Management Review, 5*(2), 129-148.

Griffiths, T, L., & Steyvers, M. (2004). *Finding scientific topics,* in Proceedings of the National Academy of Sciences, 101 (Supplement 1), 5228− 5235. https://doi.org/10.1073/pnas.0307752101.

Grootendorst, M. (2022). *BERTopic: Neural topic modeling with a*

class-based TF-IDF procedure. arXiv preprint arXiv:2203.05794.

Gronroos, C. (1990). Relationship approach to marketing in service contexts: The marketing and organizational behavior interface. *Journal of business research*, *20*(1), 3-11.

Gupta, S., Hanssens, D., Hardie, B., Kahn, W., Kumar, V., Lin, N., Ravishanker, N., & Sriram, S. (2006). Modeling customer lifetime value. *Journal of service research, 9*(2), 139-155.

Ishaq, M. I., Hussain, N., Asim, A. I., & Cheema, L. J. (2014). Brand equity in the Pakistani hotel industry. *Revista de Administração de Empresas, 54*(3), 284-295.

Jelodar, H., Wang, Y., Yuan, C., Feng, X., Jiang, X., Li, Y., & Zhao, L. (2019). Latent Dirichlet allocation (LDA) and topic modeling: models, applications, a survey. *Multimedia Tools and Applications, 78*(11), 15169-15211.

Joewono, T. B., & Kubota, H. (2007). User satisfaction with paratransit in competition with motorization in Indonesia: anticipation of future implications. *Transportation, 34*(3), 337-354.

Katzan H. (2008). Foundations of service science concepts and facilities. *Journal of Service Science, 1*(1), 1-22,

Kim, H. S.(2018). A Wisdom asset service design proposal by service nature. *Journal of Service Research and Studies, 8*(1), 1-12.

KOSIS (2020). *2019 Service Industry Survey Results (Provisional)*. Daejeon: Statistics Korea.

Ko, Y. S., Lee, S. B., Cha, M. J., Kim, S. D., Lee, J. H., Ham, J. Y., & Song, M. (2022). Topic modeling insomnia social media corpus using BERTopic and building automatic deep learning classification model. *Korean Society for Information Management, 39*(2), 111-129.

KPC (2018). *Productivity Statistics DB*. Seoul: Korea Productivity Center.

Krajewski, L. J., & Ritzman, L. P. (1999). *Operations management: Strategy and analysis*. Massachusetts: Addison Wesley.

Liu, Y., & Jang, S. S. (2009). Perceptions of Chinese restaurants in the US: what affects customer satisfaction and behavioral intentions?. *International journal of hospitality management, 28*(3), 338-348.

Lu, S. Y., & Kim, H. G. (2021). A study on the effect of advertising messages according to price discount rate and scarcity message: Focusing on anxiety and situational involvement in COVID-19 era. *The Korean Journal of Advertising, 32*(7), 63-93.

McDougall, G.H., & Levesque, T. (2000). Customer satisfaction with services: Putting perceived value into the equation. *Journal of services marketing, 14*(5), 392-410.

MOEF (2021). *Service industry COVID-19 response and development strategies*. Sejong: Ministry of Economy and Finance. (https://www.moef.go.kr › com › synap › synapView)

Nam, K. C., Kim, Y. J., Nam, J. T., Bae, W. W., Byun, H. S., & Lee, N. H. (2008). Service science: Theory review and development of Analytical framework. *Information Systems Review, 10*(1), 213-235.

Oliver, R. L. (1980). A cognitive model of the antecedents and consequences of satisfaction decisions. *Journal of marketing research, 17*(4), 460-469.

Oliver, R. L. (1997). *Satisfaction: A Behavioral Perspective on the Consumer*. New York: McGraw-Hill Book Company.

Oliver, R. L. (1999). Value as excellence in the consumption experience. In M. Holbrook (Ed.), *Consumer value: A framework for analysis and research*, 43-62, New York: Rout ledge.

Oliver, R. L. (1999). Whence consumer loyalty?. *Journal of marketing, 63*(4_suppl1), 33-44.

Ryu, K., Lee, H. R., & Kim, W. G. (2012). The influence of the quality of the physical environment, food, and service on restaurant image, customer perceived value, customer satisfaction, and behavioral intentions. *International Journal of Contemporary Hospitality Management, 24*(2), 200-223

Sumaedi, S., Bakti, I. G. M. Y., Rakhmawati, T., Astrini, N. J., Widianti, T., & Yarmen, M. (2016). Factors influencing public transport passengers' satisfaction: a new model. *Management of Environmental Quality: An International Journal, 27*(5), 585-597.

Schmenner, R. W. (1986). How can service businesses survive and prosper?. *Sloan Management Review (1986-1998), 27*(3), 21-32.

Wen, C. H., Lan, L. W., & Cheng, H. L. (2005). Structural equation modeling to determine passenger loyalty toward intercity bus services. *Transportation Research Record, 1927*(1), 249-255.

Yang, H. C. (2022). Analysis of distribution industry research trends using BERTopic and LDA. *Journal of Creativity and Innovation (JCI),* In pressing.

Yang, W. R., & Yang, H. C. (2022). Topic modeling analysis of social media marketing using BERTopic and LDA. *Journal of Industrial Distribution & Business, 13*(9), 39-52.

Yang, Z., & Peterson, R. T. (2004). Customer perceived value, satisfaction, and loyalty: The role of switching costs. *Psychology & marketing, 21*(10), 799-822.

Appendixes

Appendix 1: 2006~2019 Service Industry Survey Provisional Results

Note) The results of the 2016-2019 service industry survey by KOSIS are charted by the researcher

Appendix 2: Number of publications by year group

Division	1988~1999	2000~2004	2005~2009	2010~2014	2015~2019	2020~Present	Sum
Total number of papers	45	336	800	909	610	260	2960
Number of analysis papers	7	145	451	508	255	90	1,456

Appendix 3: Top 20 keywords as a result of frequency analysis

No	Word	Freq.	No	Word	Freq.	No	Word	Freq.	No	Word	Freq.
1	quality	3205	6	value	747	11	performance	517	16	employee	461
2	intention	1087	7	system	651	12	information	506	17	job	452
3	relationship	1072	8	company	561	13	consumer	477	18	user	419
4	loyalty	962	9	management	524	14	marketing	471	19	food	398
5	influence	772	10	restaurant	522	15	industry	461	20	environment	397

Appendix 4: Results of Word Co-occurrence Frequency Analysis

Rank	Word	Freq.	Rank	Word	Freq.
1	quality	344	6	intention	171
	relationship			relationship	
2	influence	246	7	quality	160
	quality			strategy	
3	intention	236	8	management	159
	quality			quality	
4	loyalty	216	9	loyalty	156

	quality			relationship	
5	influence	173	10	industry	150
	relationship			quality	

Appendix 5: Results of Topic modeling using BERTopic and LDA

Topic	Topic Name	Top 10 Keywords by Topic
1	Catering service	**restaurant, food,** quality, foodservice, menu, family, intention, student, university, value
2	Brand justice	**loyalty,** recovery, relationship, **brand, justice,** coffee, store, **trust,** influence, quality
3	Online shopping	**mall, shopping, internet,** consumer, intention, **site, online**, commerce, **quality**, product
4	Accommodation service	**hotel, job, employee, labor,** stress, quality, influence, relationship, **support**, industry
5	Call center service	**system, center,** performance, management, **call,** information, process, **evaluation, network**, application
6	Logistics delivery service	**logistics, port,** company, quality, **transportation, delivery, supply, terminal, chain, container**
7	Air transportation service	**airline, passenger, cabin, airport,** quality, **air,** loyalty, **flight,** influence, signage
8	Service workers	**job, employee,** intention, **company,** relationship, **industry**, system, marketing, information, **level**
9	Consumer value	**consumer,** relationship, intention, **loyalty, trust, value,** influence, information, product, internet
10	Brand value	relationship, intention, l**oyalty,** system, **value,** performance, management, **brand,** orientation, **marketing**
11	Facility equipment	loyalty, influence, value, intention, **user,** restaurant, **facility**, **system,** relationship, **performance**

Note) Bold type indicates the keywords used to determine the topic name