ISSN: 2288-7709 © 2020 KODISA & ICMA.
JEMM website: https://acoms.kisti.re.kr/jemm
doi: http://dx.doi.org/10.20482/jemm.2022.10.6.1

Topic Modeling Analysis of Beauty Industry using BERTopic and LDA*

Hoe-Chang YANG [1], Won-Dong LEE [2]

Received: October 11, 2022. Revised: October 16, 2022. Accepted: December 05, 2022.

Abstract

Purpose: The purpose of this study is identifying the research trends of degree papers related to the beauty industry and providing information which can contribute to the development of the domestic beauty industry and the direction of various research about beauty industry. **Research design, data and methodology:** This study used 154 academic papers and 189 academic papers with English abstracts out of 299 academic papers. All of these papers were found by searching for the keyword "beauty industry" in ScienceON on August 15, 2022. For the analysis, BERTopic and LDA (Latent Dirichlet Allocation) analysis were conducted using Python 3.7. Also, OLS regression analysis was conducted to understand the annual increase and decrease trend of each topic derived with trend analysis. **Results:** As a result of word frequency analysis, the frequency of satisfaction, management, behavior, and service was found to be high. In addition, it was found that 'service', 'satisfaction' and 'customer' were frequently associated with program and relationship in the word co-occurrence frequency analysis. As a result of topic modeling, six topics were derived: 'Beauty shop', 'Health education', 'Cosmetics', 'Customer satisfaction', 'Beauty education', and 'Beauty business'. The trend analysis result of each topic confirmed that 'Beauty education' and 'Health education' are getting more attention as time goes by. **Conclusions:** The future studies must resolve the extreme polarization between the structure of the small beauty industry and beauty stores. Furthermore, the researches have to direct various ways to create the performance of internal personnel. The ways to maximize product capabilities such as competitive cosmetics and brands are also needed attentions.

Keywords : Beauty Industry, Research Trends, Topic Modeling, BERTopic, LDA

JEL Classification Code : C80, I10, I30, L10, L66.

1. Introduction

If human interest in beauty and health can be classified into Maslow (1987) 's pyramid, it must be closer to esteem or self-actualization rather than human basic needs such as love & belongingness, safety needs, and physiological needs. In this respect beauty and health needs can be divided into one of self-actualization, which increases as they are satisfied, compared to physiological, safety, love and belongingness, which no longer work as a motive. Therefore, the beauty industry is a representative industry that seeks to meet human needs for beauty and health.

Compared to other industries, the beauty industry contributes greatly to the national economy due to its high value-added inducement effect due to production activities. Also, it can induce the employment inducement effect by its labor-intensive characteristics, which boosts domestic demand by resolving employment difficulties (Health

[1] First Author. Ph.D. Assistant Professor, Dept. of Distribution Management, Jangan University, South Korea, Email: pricezzang@jangan.ac.kr

[2] Corresponding Author. Professor, Department of Logistics Trade, Jangan University, Email: wdlee@jangan.ac.kr

Industry Trend, 2012). Thus the cosmetics and beauty service industries forming the beauty industry are closely linked to the value chain and can induce high synergy through convergence and convergence between industries.

According to health industry statistics, Korean cosmetics market is estimated to be $12.3 billion, accounting for about 3% of the global cosmetics market, sales of the domestic beauty service industry are 6.74 trillion won as of 2018, and the number of workers is about 22.9 million (Health Industry Brief, 2021). It is reported that interest of consumers in the beauty industry has increased significantly as purchasing power of consumers continues to improve due to economic development and increased national income. Also, the competition between companies in the beauty industry is also fierce. Therefore, beauty service companies can be competitive only by quickly responding to changes in their beauty-related knowledge and consciousness and providing valuable knowledge, skills, and services that can quickly and accurately satisfy needs of consumers.

The purpose of this study is identifying the research trends of degree papers related to the beauty industry and providing information which can contribute to the development of the beauty industry based on the main topics of interest in the domestic beauty industry. To accomplish these ends, this study used 154 academic papers and 189 academic papers with English abstracts out of 299 academic papers. All of these papers were found by searching for the keyword "beauty industry" in ScienceON on August 15, 2022. For the analysis, BERTopic and LDA (Latent Dirichlet Allocation) analysis were conducted using Python 3.7. Also, OLS regression analysis was conducted to understand the annual increase and decrease trend of each topic derived with trend analysis.

The results of this study are expected to provide insight into which areas to pay more attention as well as the direction of future studies related to the beauty industry.

2. Literature Review

2.1. Beauty Industry

Korean beauty industry has been showing rapid growth since 2010. The interest in beauty has expanded due to the increase in women's social advancement and men's desire for beauty at a time when the proportion of the service industry increases due to Korea's rapid economic growth (Kim & Han, 2021). In terms of industrial analysis, the beauty industry is increasing its growth potential as export tourism content, new consumption trends such as wellness-oriented and emotional consumption and being a major industry contributing to expanding domestic demand and creating jobs (Kim & Yang, 2014).

The beauty industry can be defined as an industry related to the manufacture, production, and development of cosmetics, beauty products, and devices used to provide services and services to manage the human body in healthy and beautiful ways (Health Industry Trend, 2012). The beauty service industry is defined as the beauty industry of consultation because the beauty industry covers not only cosmetics but also services such as hair, skin beauty, nail beauty, and makeup beauty, and manufacturing related beauty devices and supplies (Bae & Lee, 2013). Therefore, the beauty industry can be defined as an overall industry from producing tangible or intangible products and selling them to consumers to realize human aesthetic needs and expressions (Cho & Young, 2017).

As shown in <Figure 1>, the beauty industry can be largely divided into beauty service industries such as hair, skin, nails, and makeup, and beauty manufacturing industries such as cosmetics, beauty products, and devices. Also beauty industry related to medical, tourism, fashion, and food can be divided into beauty-related industries (Health Industry Trend, 2012).

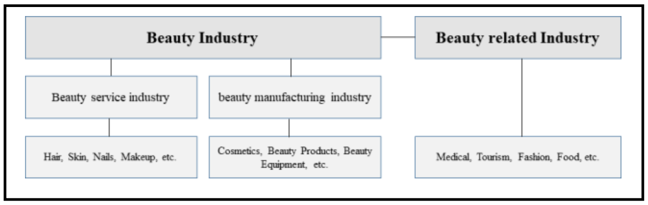

Beauty Industry
Beauty related Industry
Beauty service industry
beauty manufacturing industry
Hair, Skin, Nails, Makeup, etc.
Cosmetics, Beauty Products, Beauty Equipment, etc.
Medical, Tourism, Fashion, Food, etc.

Figure 1: Scope of the Beauty Industry

As of 2008, the beauty industry had a value-added inducement coefficient of 0.92, which is higher than the total service industry of 0.87. Also, the employment inducement coefficient was 14.3, which is much higher than 6.6 of manufacturing, and 12.6 of service industry. Thus, the beauty industry has a great effect on resolving employment difficulties and boosting domestic demand (Bank of Korea, 2008). With the growing industrial importance, the government announced a "plan to strengthen the competitiveness of the beauty industry" at the 18th National Competitiveness Reinforcement Committee (2009.10.28) with the aim of reorganizing the beauty industry and fostering the beauty industry as tourism and export products (Health Industry Trend, 2012). In addition, global interest in K-Beauty is increasing due to the recent exposure of Korean celebrities overseas along with the global Korean Wave. Accordingly, the president is paying a lot of attention and efforts in policy by launching "Brand-K" as Korean representative co-brand (Park & Kim, 2017). As such, Korean beauty industry is not only a future growth industry that can create high added value and reflect the trend of higher quality, segmentation, and specialization (Park & Kim, 2017).

According to the Korea Health Industry Promotion Agency, Korean cosmetics market is estimated to be $12.3 billion in 2019. Also, the production performance over the past five years has shown a high growth trend with an average annual growth rate of 16%, but has slowed to 16.2633 trillion won in 2019. Meanwhile, exports of cosmetics in 2019 rose 4.2% year-on-year to $6.52 billion, imports fell 2.1% year-on-year to $1.58 billion, and exports of cosmetics rose 15.7% year-on-year to $7.57 billion despite the worsening external environment caused by COVID-19.

Table 1: Cosmetics Industry Status

Division		2015	2016	2017	2018	2019
Market Size (Billions of dollars)	World Market	351.7	366.7	383.4	402.2	420.3
	Domestic	11.1	12.0	12.1	12.2	12.3
Production Performance (100 million won)		107,3	130,5	135,2	155,0	162,6
Import and Export (million dollars)	Export	2,910	4,178	4,944	6,260	6,524
	Import	1,397	1,434	1,527	1,615	1,581

Source: Health Industry Brief (2021).

In particular, the beauty industry is expected to continue to grow as the desire and expectation for beauty are expanding. However, the trend-sensitive nature makes the industry to find ways to accomplish its continuous development and ways to create new values through connection with various industries such as health, medical, culture, science, life, and emotion (Cho & Young, 2017).

Nevertheless, Korean beauty industry is limited in securing international competitiveness because of insufficient brand globalization, weak external recognition, small beauty industry structure, and insufficient overseas support system. It is evaluated that there is a lack of policy consideration for fostering industries, such as excessive and unreasonable laws and institutional operations and regulations that do not reflect reality, as well as lack of support to strengthen industrial competitiveness (Cho & Young, 2017).

In addition, it is reported that there are several problems internally in the beauty industry. Since the opening of the beauty market in 1994, franchises of beauty stores have caused extreme polarization among beauty store, and the spread of the Internet has made it easier for customers to collect beauty-related information, making it difficult to realize customer satisfaction (Kim & Kim, 2021). Despite the beauty industry has the advantage of excellent job creation because it is a labor-intensive service industry with relatively low entry barriers, it exposes various operational problems such as smallness and limitations in securing excellent human resources (Cho & Youn, 2017).

As one of the ways to find clues to solve these problems, it is very important to confirm the research trends in the beauty industry. Therefore, this study attempted to explore what fields are interested in and what directions are developed through research trends of papers related to the beauty industry.

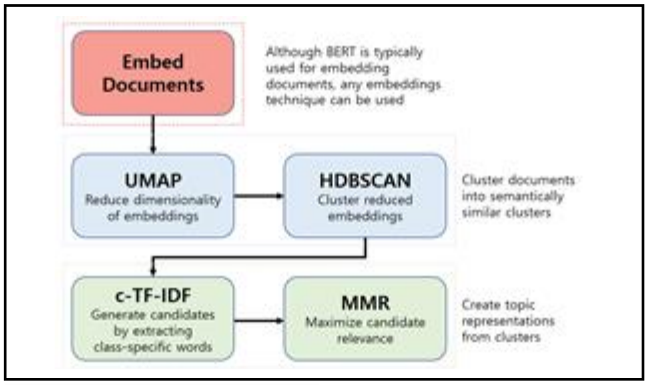
2.2. Topic Modeling

Topic modeling is an algorithm that extracts the subject matter of a document based on the keywords that make up each document from a vast collection of documents. As a result, it can derive the subject matter ratio of each document, and the distribution of the words contained in each topic (Blei, 2012; Yang & Yang, 2022a).

2.2.1. BERTopic

BERTopic is a topic modeling algorithm that utilizes BERT embeddings and class-based TF-IDF to create dense clusters, assuming that multiple documents contain topics of similar meaning (Yang & Yang, 2022a). In particular, BERTopic is known to be useful as a topic modeling algorithm by recording the highest coherence score compared to LDA, Non-negative Matrix Factorization (NMF), Correlated Topic Model (CTM), and Top2Vec (Abuzzed & Al-Khalifa, 2021; Grootendorst, 2022).

The specific implementation process of BERTopic introduced by Yang and Yang (2022a) is shown in <Figure 2>. The process is composed of Document Embeddings, Document Clustering, and Topic Representation in order. First, the document embedding process uses a Sentence-BERT (SBERT) framework that yields document embedding results from BERT as a pre-trained language model. Next, UMAP dimensionally reduce the results of high-dimensional document embeddings and cluster them with HDBSCAN for document clustering. In this case, HDBSCAN determines noise data as an outlier as a soft clustering technique. Finally, TF-IDF is modified to select the representative word of each topic for topic representation. In other words, BERTopic derives the distribution of representative words by modeling the importance of words for each cluster using TF-IDF on a cluster-by-cluster basis (Yang & Yang, 2022a).

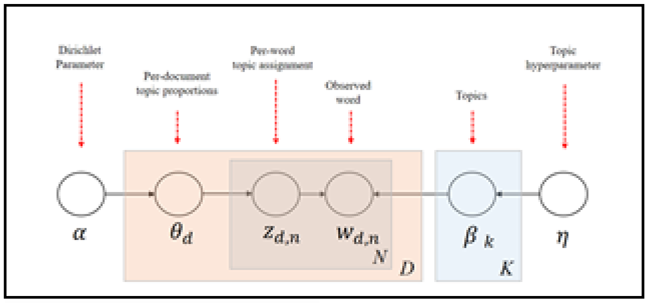

Embed Documents
Although BERT is typically used for embedding documents, any embeddings technique can be used
UMAP
Reduce dimensionality of embeddings
HDBSCAN
Cluster reduced embeddings
Cluster documents into semantically similar clusters
c-TF-IDF
Generate candidates by extracting class-specific words
MMR
Maximize candidate relevance
Create topic representations from clusters

Source: Yang and Yang (2022a).

Figure 2: Principles of BERTopics

2.2.2. LDA (Latent Dirichlet Allocation)

LDA is an algorithm that follows a Dirichlet distribution. The word weight of the extracted topic and the topic weight of the document have positive real numbers as elements and all the elements add up to 1 (Blei, 2012: Yang & Cho, 2022; Yang et al., 2022; Yang & Yang, 2022a). In particular, it is evaluated as a suitable technique for modeling real-life phenomena because one document is an algorithm that corresponds to multiple topics simultaneously (Yang & Cho, 2022; Yang et al., 2022; Yang & Yang, 2022b).



Source: Yang and Yang (2022a).

Figure 3: Principles of LDA

As shown in <Figure 3>, LDA discovers the subject hidden in the document through the observed variables, the words *Wd,n*. Hidden parameters β are used to extract words and hyperparameters, α and η. Finally, there are hidden variable, θ and *z*, that cannot be directly observed in the document. *z* of LDA is generated from θ, which represents the topic ratio for each document. Meanwhile, θ is a value that follows the Dirichlet distribution determined by the value of α. Finally, the β, the probability of word generation for each topic, is determined by η and the Dirichlet distribution of β is determined by η. As a result, the word *Wd,n* is determined by *z*, which represents the topic of each word, and β, which represents the probability of word generation for each topic.

LDA is used not only for research related to research trends (e.g., Yang & Cho, 2022; Yang et al., 2022; Yang & Yang, 2022b), but also for identifying trends in specific industries (e.g., Barua et al., 2014).

3. Research Procedure

In this study, LDA was additionally performed on outlier documents classified in BERTopic to derive topic modeling results. This is because the outlier documents identified in BERTopic are excluded from the topic modeling results, which is not suitable for the purpose of this study (Yang & Yang, 2022a). Therefore, in this study, BERTopic and LDA were sequentially used to increase the representativeness of each topic and to understand research trends, as in the study of Yang and Yang (2022a). The research procedure of this study is shown in <Figure 4>.

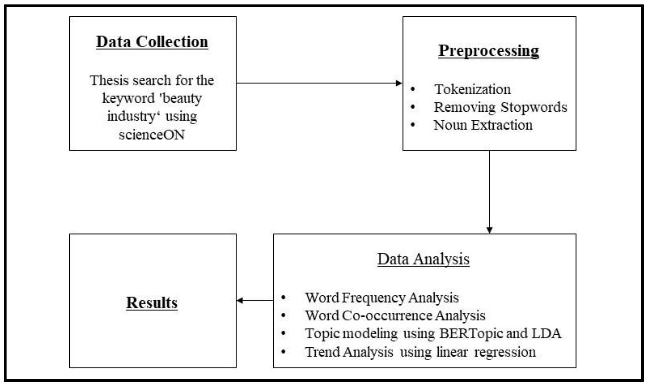

Data Collection
Thesis search for the keyword 'beauty industry' using scienceON
Preprocessing
• Tokenization
• Removing Stopwords
• Noun Extraction
Data Analysis
• Word Frequency Analysis
• Word Co-occurrence Analysis
• Topic modeling using BERTopic and LDA
• Trend Analysis using linear regression
Results

Figure 4: Resarch Process

3.1. Data Collection

In this study used 189 academic papers with English abstracts out of 154 academic papers and 299 academic papers for analysis. Those papers are searched on August 15, 2022, at ScienceON by using keyword "beauty industry". The reason for selecting the thesis is that research trends in the beauty industry can be confirmed from a more diverse perspective because the total number of papers is relatively less restrictive than academic journals. The number of thesis papers by year is presented in <Table 2>.

Table 2: Number of Thesis Publications by Year

Year	Num.	Year	Num.	Year	Num.	Year	Num.
2002	1	2010	6	2015	11	2020	16
2004	1	2011	12	2016	16	2021	23
2007	2	2012	7	2017	11	2022	16
2008	5	2013	11	2018	20	Sum	189
2009	3	2014	9	2019	19		

3.2. Data Preprocessing

For data preprocessing, the thesis to be analyzed was tokenized based on words, and then special symbols, numbers, and stopwords were removed. In the case of stopwords, expressions that appear frequently in academic paper abstracts such as aim, purpose, research and implications, and 'Beauty' and 'industry', which are search keywords, were also considered as stopwords.

4. Results and Discussion

4.1. Results of Word Frequency Analysis

As a result of word frequency analysis, it was confirmed that the frequency of use was high in the order of satisfaction (509), management (411), behavior (380), and service (377). The results of frequency analysis for the top 20 words are presented in <Appendix 1>. The results of word crowding for word frequency are presented in <Figure 5>.

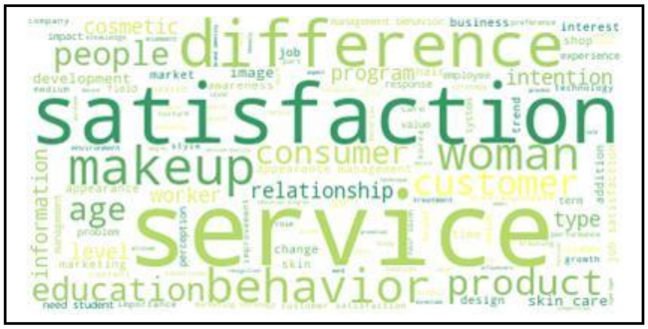


Figure 5: Results of Word Clouding

4.2. Results of Word Co-occurrence Frequency Analysis

As a result of analyzing the frequency of simultaneous appearance based on counts to check the relationship between words, it was confirmed that 'service', 'satisfaction', and 'customer' appeared frequently, indicating that the beauty industry was more interested in the service sector than in the manufacturing sector.

Table 3: Results of Word Co-Occurrence Frequency Anaysis

Rank	Word	Freq.	Rank	Word	Freq.
1	satisfaction service	53	6	care skin	42
2	customer service	45	7	age difference	40
3	program satisfaction	44	8	education program	40
4	difference satisfaction	42	9	program service	40
5	customer satisfaction	42	10	relationship satisfaction	40

4.3. Results of Topic Modeling

As a result of BERTopic, <Topic 1> was named 'Beauty shop', expecting that words such as 'hair', 'service', 'salon', and 'shop' would be related to beauty shops. <Topic 2> expects words such as 'health', 'makeup', 'type', and 'education' to be related to health, and named them 'Beauty health education'. <Topic 3> consists of 'cosmetics', 'brand', 'product', 'care', 'color', and 'tone', which are named 'Cosmetics'. On the other hand, <Topic 4>, which was derived from the LDA results for outliers, consists of 'customer', 'satisfaction' and 'service' expected to be related to customers using the beauty industry, and it was named 'Customer satisfaction'. <Topic 5> is composed of words such as 'makeup', 'education', 'care' and 'program'. It is expected to be related to beauty education, and it is named 'Beauty education'. <Topic 6> noted words such as 'job', 'service', 'behavior' and 'woman' named them 'Beauty business' in which women enter (See Appendix 2).

4.3. Trend Analysis for Each Topic

For trend analysis on topics, the ratio of each topic allocated to a specific paper was first calculated through topic modeling. After calculating the average ratio of each topic by year using the smoking year of each paper, the results of visualizing the increase or decrease of the topic over time based on this are presented in <Figure 6>. The results of visualization appear to be that <Topic 1> (Beauty shop), <Topic 5> (Beauty education), and <Topic 6> (Beauty business) have recently been studied as relatively interested topics.

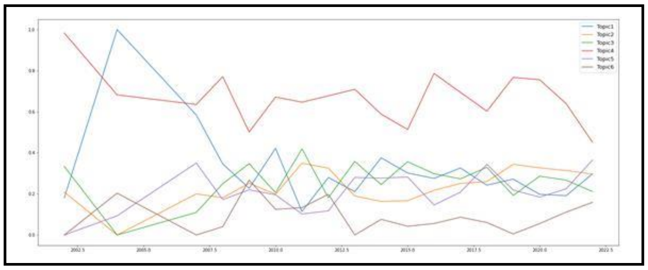


Figure 6: Variation of Tipocs by year with papers

However, in order to check more specific topic trends, the independent variable was set to the year of publication of the paper. Also, the dependent variable was set to average the weight of the topic for the year, and the OLS regression analysis was conducted. Furthermore, the annual trend of each topic was confirmed using the derived regression coefficient value. Griffiths and Steyvers (2004) stated that if the regression coefficient is a statistically significant positive coefficient, then a Hot topic can be interpreted as a Neutral topic (Yang & Yang, 2022a).

Table 4: Trend Analysis Results of Each Topic through OLS Regression Analysis

Topic		Coefficient	p-value	Trend
1	Beauty shop	-0.0150	0.068	-
2	Health education	0.0087	0.010	Hot
3	Cosmetics	0.0037	0.387	-
4	Customer satisfaction	-0.0077	0.134	-
5	Beauty education	.0088	0.025	Hot
6	Beauty business	-0.0010	0.766	-

As a result of the regression analysis, 'Beauty education’ ($\beta = 0.0088$, $p<.05$)’, ‘Health education’ ($\beta = 0.0087$, $p<.05$) were found that those are receiving more attention over time in beauty industry research.

5. Conclusions

The purpose of this study is to identify the research trends of degree papers related to the beauty industry, and to provide clues to contribute to the development of the domestic beauty industry and the direction of various research in the future. To this end, the following results and implications were derived as a result of conducting BERTopic, LDA, and trend analysis on a total of 189-degree papers in English.

Firstly, the analysis of word frequency showed that the frequency of the words such as satisfaction, management, behavior, service was high. In the analysis of the frequency of simultaneous word appearance, it was found that 'service', 'satisfaction', and 'customer' were frequently linked to programs and relationships. These results indicate that degree papers related to the beauty industry in Korea are more interested in the service sector than in the manufacturing sector. Since the beauty industry is an emotional consumption industry, this result seems to be a desirable phenomenon in responding to the sensitivity of trends. However, the beauty industry is likely to fall behind if technological innovation is not achieved. Therefore, research needs to be conducted in connection with various industries such as health, medical care, and culture to cope with consumers' desire and expectations for beauty.

Secondly, as a result of topic modeling, six topics were derived: 'Beauty shop', 'Health education', 'Cosmetics', 'Customer situation action', 'Beauty education', and 'Beauty business'. These results also mean that researchers' interest in the Korean beauty industry is showing more interest in related industries such as beauty service and health than in the manufacturing industry.

Thirdly, the trend analysis results of each topic confirmed that 'Beauty education' and 'Health education' are receiving more attention over time. It means the beauty industry is linked not only to the service sector but also to health and medical care, and that the beauty industry is a labor-intensive industry with relatively low entry barriers due to the nature of the industry. In other words, the beauty industry that researchers are interested in is expanding into the beauty-related industry as well as the beauty service sector, as classified in <Figure 1>. From another perspective, they are paying more attention to beauty and health education in relation to women's start-ups. However, in order to resolve the extreme polarization between the structure of the small beauty industry and beauty stores, it is necessary not only to pay attention to the support system to strengthen the competitiveness of the beauty industry, but also to study various ways to create internal manpower. In addition, it is necessary to pay attention to ways to maximize product capabilities such as competitive cosmetics and brands.

Although this study identified the areas of interest of researchers by examining the research trends in the beauty industry, further studies need to supplement this research. Firstly, this study targets only the degree thesis with the expectation that the absolute amount of the thesis will be more than that of the journal, and also that many of the thesis will be published in the journal. This hypothesis may limit to properly reflect the trend of the beauty industry. Therefore, in future studies, it is necessary to analyze research trends based on more diverse data such as comparison with papers in overseas academic journals along with newspaper articles and reports. Secondly, although BERTopic has the advantage of excluding keywords that are not related to LDA and determining topics through coherence scores, there is a limit to the subjectivity of topic modeling. Therefore, in future studies, it is necessary to apply a more objective method for determining the number of topics. Finally, if a comparative study with related industries in a supplementary and substitution relationship is conducted in an industrial trend study, more various implications can be derived.

References

Abuzayed, A., & Al-Khalifa, H. (2021). BERT for Arabic topic modeling: An experimental study on BERTopic technique. *Procedia Computer Science, 189*, 191-194.

Bae, K. H., & Lee, Y. J. (2013). Analysis of economic effects of beauty industry by input-output table. *The Journal of the Korea Contents Association, 13*(4), 350-360.

Bank of Korea (2008). *Industry association table*. Seoul: Bank o Korea.

Barua, A., Thomas, S. W., & Hassan, A. E. (2014). What are developers talking about? an analysis of topics and trends in stack overflow. *Empirical Software Engineering, 19*(3), 619-654.

Blei, D. M. (2012). Probabilistic topic models. *Communications of the ACM, 55*(4), 77-84.

Cho, Y. O., & Youn, C. S. (2017). A study on factor analysis of platform business model in beauty industry using AHP. *The Journal of Humanities and Social science, 8*(4), 185-206.

Griffiths, T. L., & Steyvers, M. (2004). *Finding scientific topics*. Proceedings of the National academy of sciences, 101(suppl_1), 5228-5235.

Grootendorst, M. (2022). BERTopic: Neural topic modeling with a class-based TF-IDF procedure. *arXiv preprint arXiv: 2203.05794.*

Health Industry Brief (2021). *Basic survey for promotion of cosmetics and beauty service industry and mutual growth plan (vol.325)*. Seoul: Korea Health Industry Promotion Agency.

Health Industry Trend (2012). *2012 Health industry major performance and issues: Beauty and cosmetic industry*. Seoul: Korea Health Industry Promotion Agency.

Kim, J. Y., & Han, C. J. (2021). Influence of beauty service workers' psychological capital on their creative behavior according to empowerment and shared leadership. *Journal of The Korean Society of cosmetology, 27*(5), 1255-1266.

Kim, C. B., & Kim, H. S. (2021). The effect of the quality of education service on the performance of education service through relationship commitment in franchise beauty academy: Moderating effect of trust level. *Asia-Pacific Journal of Business Venturing and Entrepreneurship (APJBVE), 16*(3), 193-211.

Kim, M. S., & Yang, C, K. (2014). A study on the franchise trend and perception of co-branding strategy for beauty industry. *Journal of Health and Beauty, 8*(2), 35-43.

Maslow, A. H. (1987). *Motivation and Personality* (3rd ed.). Boston, MA: Addison-Wesley.

Park, J. W., & Kim, S. W. (2017). Effects of SNS wom information characteristics on brand attitude, brand image and purchase intention. *Global Business Administration Review, 14*(5), 229-249.

Yang, H. C. (2021). Topic modeling analysis of franchise research trends using LDA algorithm. *The Korean Journal of Franchise Management, 12*(4), 13-23.

Yang, H. C., & Cho, H. Y. (2022). Topic modeling analysis of HMR research trends using LDA. *Korea Logistics Review, 32*(1), 81-92.

Yang, H. C., Ju, Y. H., & Cho, H. Y. (2022). Topic modeling of CVS research trends using LDA. *The Journal of Business Education, 36*(2), 121-143.
DOI: 10.34274/krabe.2022.36.2.006

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

Yang, W. R., & Yang, H. C. (2022b). Overseas research trends telated to 'research ethics' using LDA topic modeling. *Journal of Research and Publication Ethics, 3*(1), 7-11.

Appendixes

Appendix 1: Results of Word Frequency Analysis

No	Word	Freq.	No	Word	Freq.	No	Word	Freq.	No	Word	Freq.
1	satisfaction	509	6	makeup	327	11	cosmetic	290	16	image	275
2	management	41	7	appearance	320	12	type	282	17	education	274
3	behavior	380	8	difference	313	13	woman	281	18	intention	251
4	service	377	9	hair	312	14	skin	280	19	relationship	246
5	customer	335	10	job	300	15	care	277	20	product	227

Note: Freq.: Frequency

Appendix 2: Results of Topic Modeling

Division	Topic	Topic Name	keyword
BERTopic	1	Beauty shop	customer, **hair,** satisfaction, **service, salon,** relationship, **shop,** intention, job, marketing
	2	Health education	management, job, appearance, satisfaction, behavior, difference, **health, makeup, type, education**
	3	Cosmetics	**cosmetic, brand,** market, skin, **product, care, c**onsumer, age, **color, tone**
LDA	4	Customer satisfaction	cosmetic, **customer,** intention, **satisfaction, service,** product, type, hair, skin, market
	5	Beauty education	management, **makeup,** image, behavior, appearance, **education, care,** relationship, woman, **program**
	6	Beauty business	satisfaction, **job, service,** difference, **behavior,** appearance, type, hair, **woman,** age

Note: Bold type word is a word that contributed to the derivation of the topic name.