



Topic Modeling Analysis of Social Media Marketing using BERTopic and LDA

Woo-Ryeong YANG¹, Hoe-Chang YANG²

Received: January 15, 2021. Revised: November 29, 2021. Accepted: December 05, 2021.

Abstract

Purpose: The purpose of this study is to explore and compare research trends in Korea and overseas academic papers on social media marketing, and to present new academic perspectives for the future direction in Korea. **Research design, data and methodology:** We used English abstract of research paper (Korea's: 1,349, overseas': 5,036) for word frequency analysis, topic modeling, and trend analysis for each topic. **Results:** The results of word frequency and co-occurrence frequency analysis showed that Korea researches focused on the experiential values of users, and overseas researches focused on platforms and content. Next, 13 topics and 12 topics for Korea and overseas researches were derived from topic modeling. And, trend analysis showed that Korean studies were different from overseas in applying marketing methods to specific industries and they were interested in the short-term performance of social media marketing. **Conclusions:** We found that the long-term strategies of social media marketing and academic interest in the overall industry will necessary in the future researches. Also, data mining techniques will necessary to generate more general results by quantifying various phenomena in reality. Finally, we expected that continuous and various academic approaches for volatile social media is effective to derive practical implications.

Keywords : Social Media Marketing, Research Trends, Topic Modeling, BERTopic, LDA

JEL Classification Code : C10, C49, C60, L82

1. Introduction

전통적 미디어에 의존하던 기업 브랜드의 고객들이 보다 능동적이고 선택적인 태도로 자신이 선호하는 정보를 수집하게 되면서 소셜 미디어가 하나의 마케팅 채널로 자리잡게 되었다. 특히 이러한 성향이 강하게 나타나는 MZ 세대의 등장(Lee & Kwon, 2022)과 코로나19 이후 더욱 증가한 언택트

마케팅 활동(Bae & Chang, 2021)은 소셜 미디어가 브랜딩과 관련한 다양한 기능을 탑재하도록 이끌었다. 이처럼 점차 발전하는 소셜 미디어를 활용한 마케팅(이하 소셜 미디어 마케팅)은 고객에게 정보 수집의 편의성을 제공하고 기업에게도 마찬가지로 저비용으로 제품과 서비스의 홍보가 가능하도록 한다(Kim et al., 2012).

소셜 미디어 마케팅에 관한 국내외 연구는 기업

1 First Author. Ph.D. Candidate, Dept. of Business Informatics, Hanyang University, Korea. Email: wooryeong325@naver.com

2 Corresponding Author. Assistant Professor, Department of Distribution Management, Jangan University, South Korea. Email: pricezzang@jangan.ac.kr

© Copyright: The Author(s)
 This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>) which permits unrestricted noncommercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

혹은 고객 관점에 대하여 다수 이루어졌다. 예를 들어, 기업의 관점에서는 인기도가 높은 콘텐츠의 발행 전략 탐색(e.g., Deng et al., 2021), 콘텐츠의 인기도 예측(e.g., Kennedy et al., 2021) 및 인플루언서 마케팅(e.g., Martínez-López et al., 2020)과 같이 소셜 미디어에서의 다양한 마케팅 방식의 효과성을 측정하여 향후 기업이 나아가야 할 방향을 제시하였다. 또한 고객의 관점에서는 고객의 특성이나(e.g., Saef et al., 2018) 그들이 체감하는 다양한 경험적 가치(e.g., Nadeem et al., 2021)가 구매 의도와 같은 마케팅 성과에 미치는 영향을 측정하는 접근이 주를 이루고 있다. 그러나 이러한 학술적인 노력에도 불구하고, 소셜 미디어의 역동성을 탐색할 수 있는 다양한 관점의 추가 연구가 이뤄져야 한다는 한계점이 여전히 남아있는 실정이다(Li et al., 2021; Voorveld et al., 2018).

따라서 소셜 미디어의 변동성과 성과 예측의 어려움을 극복하기 위해서는 열린 시각으로 해결 방안을 강구하는 학술적 접근이 동반되어야 한다. 예를 들어, 연구 동향 파악을 통해 소셜 미디어 마케팅의 전문가들의 시각을 확인하고 비교함으로써 학술적으로 외부 환경의 변화에 대응하고 적응하며 발전할 수 있는 방안을 마련할 수 있다(Yang, 2021).

이에 본 연구는 소셜 미디어 마케팅의 연구 동향을 파악하여, 현 시점의 주된 연구 관심 주제와 이를 기반으로 향후 연구의 방향성을 제시하는 것이 소셜 미디어 마케팅 연구에 대한 유용한 통찰력을 제시할 수 있을 것이라 판단하였다. 따라서 본 연구는 소셜 미디어 마케팅의 연구 동향을 탐색하는 것을 목적으로 하며, 그러한 연구 동향을 총체적인 관점으로 바라보기 위해 연구의 범위를 국내로 한정하지 않고 국내외 동향에 대해 탐색 및 비교하는 것으로 하였다.

본 연구는 국내외 소셜 미디어 마케팅 연구 흐름의 방향을 확인하기 위한 수단으로 관련 분야의 국내외 학술논문에 주목하였다. 이러한 접근을 통해

소셜 미디어 마케팅 전문가들의 주된 연구 관심 주제를 도출할 수 있기 때문에, 앞서 언급한 소셜 미디어 마케팅 연구에 현존하는 한계점을 극복할 수 있는 단서로 활용하기에 적합할 것이라고 판단하였다.

따라서 본 연구는 국내외 소셜 미디어 마케팅의 연구 동향을 파악하고 비교하기 위해 ScienceON에서 수집 가능한 소셜 미디어 마케팅 관련 국내외 논문의 영문초록을 수집하였다. 분석 기법으로는 카운트 기반의 단어 빈도분석 및 BERTopic과 LDA(Latent Dirichlet Allocation)를 활용한 토픽 모델링을 수행하였다. 마지막으로 국내외 논문에 대해 도출한 각 토픽의 연도별 증감 추이를 파악하기 위해 단순회귀분석을 수행하였다.

본 연구의 결과는 향후 소셜 미디어 마케팅 관련 연구가 나아가야 할 방향뿐만 아니라 소셜 미디어 마케팅 관련 국내 연구가 범위를 확장하고 다양성을 추구할 수 있도록 학술적 관점의 새로운 기준을 제시할 수 있을 것으로 기대하였다.

2. Literature Review

소셜 미디어 마케팅 관련 연구는 마케팅 방법론의 효과성 측정이나 영향 요인 탐색의 방향이 주를 이루고 있다. 한편, 변동이 큰 소셜 미디어 시장에서 연구 관점 확장의 필요성이 지속적으로 대두됨에 따라, 본 연구는 소셜 미디어 마케팅에 대한 연구 방향의 현 위치를 확인하고 국내외 연구 동향을 파악하는 것이 그러한 관점 확장을 위한 단서를 제공할 것으로 기대하였다. 이를 위한 분석 기법으로는 문서의 특정 주제를 자동으로 추출하는 비지도 학습 기법인 토픽 모델링 알고리즘을 활용하였다.

2.1. 소셜 미디어 마케팅

소셜 미디어는 개인의 경험을 온라인으로 공유

하는 인터넷 기반 응용 프로그램을 의미하며(Hays et al., 2013) 응용 프로그램 기반의 콘텐츠 생성 및 소비 플랫폼으로 정의된다(Kaplan & Haenlein, 2010). 특히 정보의 개방, 사용자들 간의 정보 공유와 참여, 정보의 가치를 지속적으로 증대 시키는 것을 목표로 하는 웹 2.0의 시대가 도래하면서 네트워크가 확장되고 포털 중심의 독점적 소통 생태계에서 개인 미디어인 소셜 미디어로의 변화가 가속화되었다(Thackeray et al., 2008).

소셜 미디어 마케팅은 연구분야와 상관없이 ‘상호작용성’, ‘정보성’, ‘개인화’, ‘유행성’, ‘오락성’이 주된 구성요소로 언급된다(Kim & Ko, 2012; Yadav & Rahman, 2017). 이는 소셜 미디어가 이용자의 일상적 측면을 넘어서 사회, 경제, 문화 등을 불문하고 다양한 분야에서의 커뮤니케이션을 위한 활용성이 높다는 것을 의미한다.

코로나19가 초래한 언택트 환경은 성숙기에 접어들면서 성장세가 둔화된 듯했던 소셜 미디어 시장이 또 다른 성장 동력을 맞이하게 하였다(Bae & Chang, 2021). 또한 개인의 관심사 위주로 공유하는 3세대 소셜 미디어에서는 그 안에서의 정보 수집을 통해 실제 제품 및 서비스 구매 활동으로 이어지는 경우가 점차 증가하였다. 예를 들어, 3세대 소셜 미디어에 속하는 인스타그램에서는 90% 이상의 이용자가 1개 이상의 브랜드 공식 계정을 팔로우 하였으며, 72% 이상의 이용자는 브랜드 콘텐츠를 접한 뒤 실제 구매 행동을 취하였다(Huang & Copeland, 2020)

소셜 미디어 마케팅과 관련한 연구는 고객의 제품 및 서비스 구매의도나 구전의도의 선행 요인을 탐색(e.g., Gautam & Sharma, 2017)하거나 패션, 스포츠, 식품, 교육 등 다양한 업종에서 소셜 미디어 마케팅의 효과성을 측정(e.g., Bolat & O’Sullivan, 2017)하는 주제가 주를 이루고 있다. 한편, 특정 산업에 관한 소셜 미디어 마케팅의 국내 연구 동향을 분석한 연구(e.g., Park et al., 2019)는 일부 보고되고 있으나 이

를 총체적인 관점에서 바라본 국내외 연구동향 분석은 거의 이루어지지 않고 있다.

2.2. 토픽모델링

토픽모델링은 방대한 양의 문서집합에서 각 문서를 이루고 있는 키워드들을 바탕으로 문서의 주제를 추출하는 알고리즘이며 결과적으로 전체 문서 집합의 주제들, 각 문서 별 주제의 비율, 각 주제에 포함된 단어들의 분포를 도출할 수 있다(Blei, 2012).

2.2.1. BERTopic

BERTopic은 다수의 문서가 서로 유사한 의미의 토픽을 내포하고 있음을 가정하며 BERT embeddings와 클래스 기반(class-based) TF-IDF를 활용하여 조밀한 클러스터를 만드는 토픽 모델링 알고리즘이다. 특히 BERTopic은 LDA, NMF(Non-Negative Matrix Factorization), CTM(Correlated Topic Model), Top2Vec과 비교하였을 때 가장 높은 coherence score를 기록하여 토픽 모델링 알고리즘으로서 유용성이 입증된 바 있다(Abuzayed & Al-Khalifa, 2021; Grootendorst, 2022).

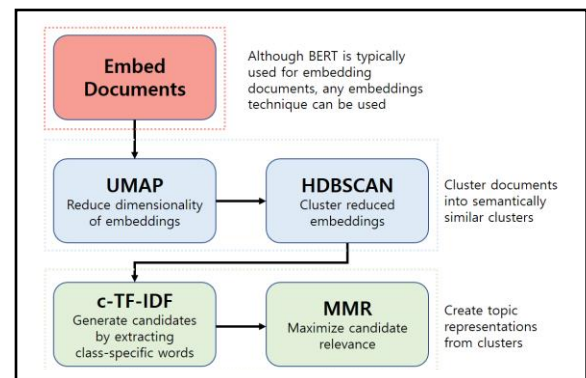


Figure 1: Principles of BERTopic

<그림 1>과 같이 BERTopic의 구체적인 구현 과정은 문서 임베딩(Document Embeddings), 문서 클러스터링(Document Clustering), 토픽 표현(Topic Representation) 순으로 이루어진다. 먼저 문서 임베딩 과정에서는 사전 훈련된 언어 모델로 BERT로부터 문서 임베딩

결과를 산출하는 SBERT(Sentence-BERT)의 프레임워크를 사용한다. 다음으로 문서 클러스터링을 위해 UMAP으로 먼저 고차원의 문서 임베딩 결과를 차원 축소하며 HDBSCAN으로 클러스터링 한다. 이때, HDBSCAN는 소프트 클러스터링 기법으로서 노이즈 데이터를 이상치로 판별한다. 마지막으로 토픽 표현을 위해 TF-IDF를 변형하여 각 토픽의 대표 단어를 선출한다. 즉, BERTopic은 문서가 아닌 클러스터 단위로 TF-IDF를 사용하여 각 클러스터별 단어의 중요성을 모델링하여 대표 단어의 분포를 도출한다.

2.2.2. LDA (Latent Dirichlet Allocation)

LDA는 추출된 토픽의 단어 비중과 문서의 토픽 비중이라는 두 변수가 모두 양의 실수를 요소로 가지며, 모든 요소를 더한 값이 1이 되는 Dirichlet 분포를 따르는 알고리즘이다(Blei, 2012). 특히 하나의 문서가 여러 토픽과 동시에 대응되는 알고리즘이기 때문에 현실의 현상에 대한 모델링에 적합한 기법으로 평가받고 있다(Yang & Yang, 2022).

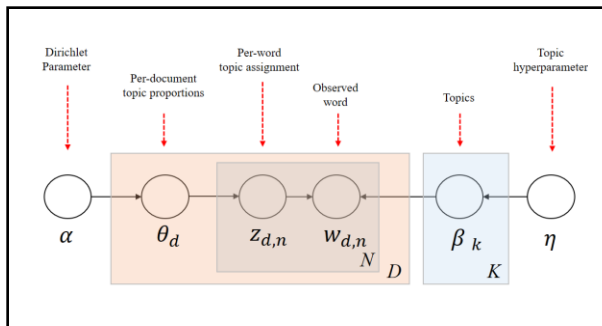


Figure 2: Principles of LDA

<그림 2>와 같이 LDA는 관찰되어지는 변수인 단어 $w_{d,n}$ 을 통해 문서에 숨겨진 주제를 발견한다. Hyperparameter인 α, η 과 단어를 추출하는 데는 Hidden parameter인 β 를 사용한다. 마지막으로 문서 내에서 직접적으로 관찰할 수 없는 Hidden variable인 θ, z 가 있다. LDA의 z 는 문서 별 토픽 비율인 θ 로부터 생성되고 θ 는 Dirichlet 분포를 따르는 값으로서

α 값에 의해 형태가 정해지게 된다. 마지막으로 토픽 별 단어 생성확률인 β 는 η 에 의해 결정되어지므로 η 에 따라 β 의 Dirichlet 분포가 결정된다. 결과적으로 각 단어의 토픽을 나타내는 z 와 토픽 별 단어 비율인 β 에 따라 단어 w 가 결정된다.

LDA는 연구 동향 관련 연구(e.g., Yang & Yang, 2022) 뿐만 아니라 특정 산업의 트렌드를 파악(e.g., Barua et al., 2014)하는 목적으로도 다수 이용되고 있다.

3. Research Procedure

본 연구의 연구 절차는 <그림 3>과 같다. BERTopic과 LDA을 활용한 분석은 Python 3.7에서 제공하는 패키지를 import 하여 수행하였다.

본 연구에서는 BERTopic에서 분류된 이상치 문서에 대해 LDA를 추가 수행하여 토픽모델링 결과를 도출하였다. 왜냐하면 BERTopic에서 판별된 이상치 문서는 토픽모델링 결과에서 제외되기 때문에 각 토픽의 대표성을 높일 수 있지만, 이상치 문서들이 갖는 토픽을 고려하지 않는 것은 연구 동향 파악이라는 본 연구의 목적에 부합하지 않기 때문이다. 따라서 본 연구에서는 각 토픽의 대표성을 높이면서 연구 동향의 파악이 가능하도록 BERTopic과 LDA를 순차적으로 사용하였다.

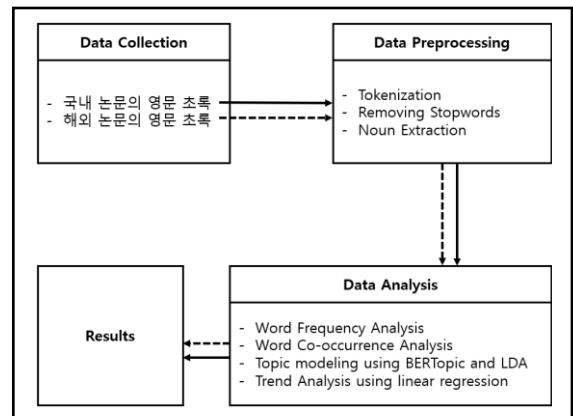


Figure 3: Research process

3.1. 데이터 수집

본 연구의 수행을 위해 2022년 4월 2일 기준 ScienceON(구 NDSL: National Digital Library)에서 검색 가능한 국내외 학술논문의 영문초록을 수집하였다. 검색 키워드는 “Social media marketing(소셜 미디어 마케팅)”, “SNS marketing(SNS 마케팅)”, “Social media branding(소셜 미디어 브랜딩)”, “SNS branding(SNS 브랜딩)”, “Social media commerce(소셜 미디어 커머스)”, “SNS commerce(SNS 커머스)”으로 구분하였다. 왜냐하면 활용 범위의 확대에 따라 소셜 미디어와 SNS의 의미가 동일시되고 있으며(Boyd & Ellison, 2007) 그러한 플랫폼에서의 마케팅 활동은 기업의 영위를 위한 커머스 활동과 일맥상통하기 때문이다(Curty & Zhang, 2013). 그 결과, 총 2,156편의 국내논문과 6,877편의 해외논문의 영문초록을 수집하였다. 그 중, 영문초록이 탑재되어 있지 않은 논문을 제거하여 총 1,349편의 국내논문과 5,036편의 해외논문을 분석 대상으로 하였다. 연도군 별 국내외 논문의 개수는 <표 1>과 같다.

Table 1: Number of domestic and overseas academic papers by year group

Year group	Domestic	%	Overseas	%
2020~2022	337	25%	1,772	35%
2015~2019	531	39%	2,219	44%
2010~2014	428	32%	789	16%
1991~2009	53	4%	241	4.7%
1973~1990			15	0.3%
Total	1,349	100%	5,036	100%

3.2. 데이터 전처리

본 연구에서는 데이터 전처리를 위해 먼저 분석 대상의 논문을 단어 기준으로 토큰화한 후 특수기호, 숫자, 불용어를 제거하였다. 불용어의 경우, research, implications처럼 학술논문 초록에 빈번히 등

장하는 표현들과 검색 키워드에 해당하는 단어들을 불용어로 간주하였다.

4. Research Results

4.1. 단어 빈도분석 결과

ScienceON에서 키워드로 검색한 국내외 논문에 대해 단어 빈도분석을 수행한 결과는 <표 2>와 같다. 분석 결과, 국내논문의 상위 단어는 의도(intention), 사용자(user), 커뮤니케이션(communication), 가치(value), 만족(satisfaction)순으로 나타났으며, 해외논문의 상위 단어는 사용자(user), 제품(product), 커뮤니케이션(communication), 콘텐츠(content), 플랫폼(platform) 등의 순으로 나타났다.

Table 2: Results of word frequency analysis

Domestic					
Rank	Word	Freq	Rank	Word	Freq
1	intention	998	11	image	516
2	product	937	12	fashion	488
3	user	846	13	purchase	463
4	communication	748	14	advertising	460
5	characteristic	706	15	experience	430
6	content	694	16	activity	428
7	value	590	17	business	412
8	market	556	18	method	398
9	satisfaction	522	19	tourism	393
10	type	519	20	survey	386
Overseas					
Rank	Word	Freq	Rank	Word	Freq
1	user	2,360	11	technology	1,368
2	product	2,279	12	intention	1,328
3	communication	2,241	13	facebook	1,311
4	business	1,874	14	engagement	1,233
5	content	1,776	15	value	1,227
6	campaign	1,582	16	market	1,151
7	health	1,551	17	behavior	1,128
8	platform	1,529	18	message	1,105
9	activity	1,454	19	twitter	1,080
10	community	1,450	20	system	1,040

4.2. 단어 동시출현빈도분석 결과

구체적인 단어 간의 관계를 확인하기 위해 카운트 기반의 동시출현빈도를 분석한 결과는 국내에 대한 <표 3>, <그림 4>과 해외에 대한 <표 4>, <그림 5>에 제시하였다. 상위 20개의 단어 쌍을 보았을 때, 국내논문에서는 'user', 'characteristic', 'survey' 와 같은 고객 관점이 두드러지며, 이와 달리 해외논문에서는 'content', 'platform' 등 기업 관점의 마케팅 요소가 다수 포함되었다.

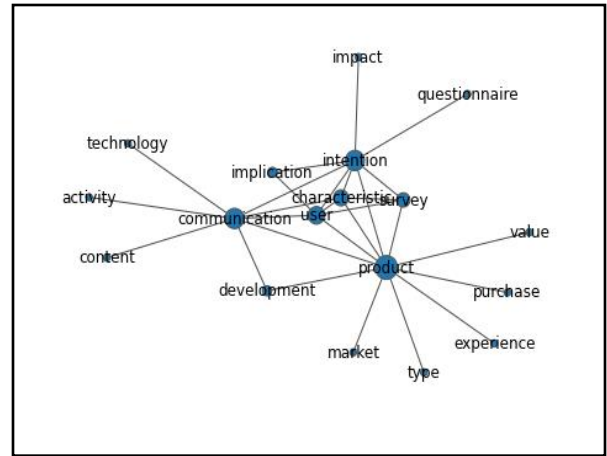


Figure 4: Visualizing results of co-occurrence network analysis for domestic academic papers

Table 3: Results of co-occurrence network analysis for domestic academic papers

Rank	Word	Freq	Rank	Word	Freq
1	intention	130	11	characteristic	97
	survey			communication	
2	intention	127	12	product	97
	user			user	
3	communication	124	13	communication	94
	user			content	
4	intention	120	14	survey	93
	purchase			user	
5	characteristic	114	15	communication	92
	intention			development	
6	characteristic	113	16	intention	92
	user			product	
7	market	110	17	questionnaire	90
	product			survey	
8	product	106	18	development	90
	purchase			product	
9	communication	106	19	characteristic	90
	product			survey	
10	characteristic	105	20	product	90
	product			survey	

Table 4: Results of co-occurrence network analysis for overseas academic papers

Rank	Word	Freq	Rank	Word	Freq
1	content	291	11	communication	235
	user			user	
2	platform	283	12	communication	234
	user			platform	
3	product	278	13	activity	233
	user			communication	
4	business	265	14	content	230
	product			platform	
5	platform	262	15	business	223
	product			communication	
6	market	248	16	content	222
	product			product	
7	communication	247	17	business	219
	technology			platform	
8	communication	240	18	facebook	218
	product			twitter	
9	communication	239	19	application	207
	content			user	
10	business	239	20	business	204
	technology			market	

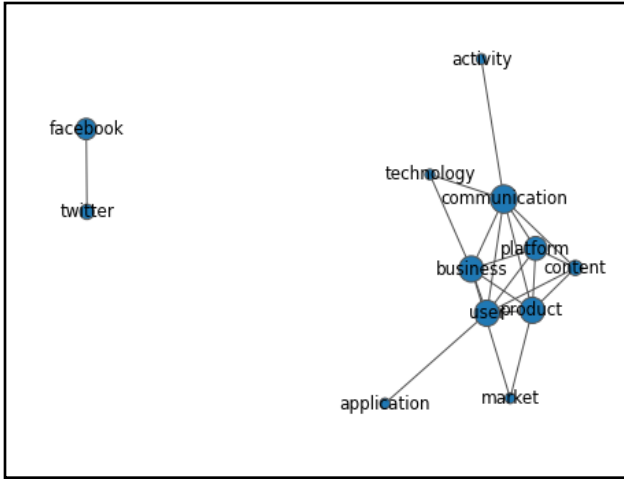


Figure 5: Visualizing results of co-occurrence network analysis for overseas academic papers

4.3. 토픽 모델링 결과

먼저 국내의 논문에 BERTopic을 적용했을 때, 국내논문에 대한 19개의 토픽과 해외논문에 대한 56개의 토픽이 도출되었다. 이때, 국내외 토픽의 비교가 가능하도록 토픽 수가 비교적 적은 국내논문을 기준으로 토픽 수를 축소할 필요가 있다고 판단하였다. 토픽 수의 축소에 앞서 분류의 명확성을 위해 국내논문에서 도출한 토픽의 중복 가능성을 검토한 결과, 19개의 토픽에서 10개의 토픽으로 축소하는 것이 적합하다고 판단하여 국내외 논문에 대한 토픽 수를 10개로 지정하고 BERTopic을 재수행 하였다. 또한 LDA에서도 분석 결과 해석의 용이성을 위해 토픽 수를 2~5개로 한정하였다. LDA 결과, 국내논문에 대한 Coherence value는 토픽의 개수가 3일 때 약 0.20으로 가장 높았으며 해외논문에 대해서는 토픽의 개수가 2일 때 약 0.37로 가장 높았다. 결과적으로 <표 5>와 같이 국내논문에 대한 총 13개의 토픽과 해외논문에 대한 총 12개의 토픽을 도출하였다. 각 토픽에 대한 주제명 결정을 위해 마케팅을 전공한 J대학 교수 2명과 함께 검토하였으며, 그 결과는 Appendix 1과 Appendix 2에 정리하였다.

Table 5: Number of representative papers on each topic for domestic and overseas

Topic	Domestic		Overseas		Topic	Count	%
	Count	%	Topic	Count			
1	146	11%	1	335	7%		
2	140	10%	2	278	6%		
3	84	6%	3	245	5%		
4	82	6%	4	182	4%		
5	76	6%	5	143	3%		
6	64	5%	6	132	3%		
7	59	4%	7	130	3%		
8	53	4%	8	105	2%		
9	48	4%	9	102	2%		
10	37	3%	10	97	2%		
11	142	11%	11	1,701	34%		
12	269	20%	12	1,586	31%		
13	149	11%					
Total	1,349	100%	Total	5,036	100%		

4.3.1. 국내논문에 대한 토픽모델링 결과

국내논문에 대한 13개의 토픽 중 먼저 토픽 1은 ‘user’, ‘movie’, ‘film’, ‘communication’로 이루어져 ‘SNS 영화마케팅(SNS film marketing)’으로 명명하였다. 토픽 2는 ‘intention’, ‘shopping’, ‘purchase’, ‘product’ 등의 단어를 통해 ‘SNS 쇼핑서비스(SNS shopping service)’으로 명명하였다. ‘food’, ‘restaurant’, ‘franchise’ 등의 토픽 3은 ‘SNS 식품마케팅(SNS food marketing)’으로 명명하였다. 토픽 4는 ‘tourism’, ‘destination’, ‘experience’로 이루어져 ‘SNS 관광마케팅(SNS tourism marketing)’으로 정의하고 토픽 5는 ‘identity’, ‘society’, ‘design’, ‘csr’을 통해 ‘SNS CSR활동(SNS CSR activity)’으로 정의하였다. ‘fashion’, ‘luxury’, ‘trend’, ‘clothing’ 등의 단어로 구성된 토픽 6은 ‘SNS 패션마케팅(SNS fashion marketing)’으로 정의하였다. 토픽 7은 ‘advertising’, ‘type’, ‘facebook’, ‘content’ 등을 통해 SNS의 특성에 맞게 콘텐츠를 제품 및 서비스의 홍보 수단으로 활용하는 ‘SNS 네이티브광고(SNS native advertising)’가 적합하며 ‘ecommerce’, ‘application’, ‘platform’, ‘smartphone’ 등을 통해 토픽 8

은 ‘SNS 모바일커머스(SNS mobile commerce)’가 적합할 것이라 판단하였다. ‘beauty’, ‘cosmetic’, ‘collaboration’에 관한 토픽 9는 ‘SNS 뷰티마케팅(SNS beauty marketing)’으로 명명하였다. 토픽 10은 ‘sport’, ‘team’, ‘league’를 통해 ‘SNS 스포츠마케팅(SNS sports marketing)’으로 정의하였으며 Topic 11은 ‘influencer’, ‘message’, ‘interaction’ 위주의 단어로 이루어져 ‘SNS 인플루언서마케팅(SNS influencer marketing)’으로 명명하였다. 마지막으로 ‘quality’, ‘technology’, ‘system’ 등의 대표 단어로 보아 토픽 12는 SNS 이용 방법 습득을 위한 별도의 노력이 요구되지 않는다는 측면의 ‘SNS 이용편의성(SNS ease of use)’이 적절하며 Topic 13은 ‘performance’, ‘design’, ‘video’로 보아 ‘SNS 영상마케팅(SNS video marketing)’으로의 명명이 적합할 것으로 판단하였다.

4.3.2. 해외논문에 대한 토픽모델링 결과

다음으로 해외논문에서 도출한 12개의 토픽 중 토픽 1은 ‘ecommerce’, ‘purchase’, ‘shopping’, ‘online’ 등을 통해 ‘SNS 쇼핑 서비스(SNS shopping service)’로 명명하였다. 토픽 2는 ‘sentiment’, ‘opinion’, ‘review’, ‘recommendation’ 로 이루어져, 웹사이트와 소셜미디어에서 특정 주제에 대한 다양한 정보를 수집 및 분석하여 평판을 도출하는 데이터 마이닝 기법인 ‘SNS 오피니언마이닝(SNS opinion mining)’으로 정의하였다. ‘tourism’, ‘destination’, ‘visitor’ 등으로 이루어진 토픽 3은 ‘SNS 관광 마케팅(SNS tourism marketing)’으로 정의할 수 있었다. 토픽 4는 ‘engagement’, ‘facebook’, ‘page’, ‘motivation’ 등으로 구성되었으므로, 마케팅 활동에 대한 고객의 능동적이고 긍정적인 반응을 의미하는 ‘SNS 고객인게이지먼트(SNS consumer engagement)’가 적절할 것으로 판단하였다. 토픽 5는 ‘tobacco’, ‘cigarette’, ‘campaign’ 등의 단어를 통해 ‘SNS 담배마케팅(SNS tobacco marketing)’으로 정의하였다. 토픽 6은 ‘food’, ‘child’, ‘adolescent’, ‘beverage’ 로 이루어져 ‘SNS 식품마케팅

(SNS food marketing)’으로 정의하며, ‘twitter’, ‘diffusion’, ‘algorithm’, ‘propagation’ 등의 단어로 이루어진 토픽 7는 ‘SNS 구전효과(SNS word of mouth)’로 명명하였다. ‘student’, ‘university’, ‘education’, ‘learning’ 등의 토픽 8은 ‘SNS 교육적활용(SNS educational application)’이 적절할 것으로 판단하였다. 토픽 9는 ‘advertising’, ‘communication’, ‘attitude’, ‘value’ 위주이므로 ‘SNS 네이티브광고(SNS native advertising)’가 적절하며 ‘sport’, ‘fan’, ‘athlete’ 등의 토픽 10은 ‘SNS 스포츠 마케팅(SNS sports marketing)’이 적절할 것으로 판단하였다. 마지막으로 토픽 11은 ‘system’, ‘site’, ‘time’, 등을 통해 인터페이스의 반응 속도와 관련한 ‘SNS 시스템 품질(SNS system quality)’이 적절하며 ‘community’, ‘engagement’, ‘behavior’, ‘way’ 등의 단어로 이루어진 토픽 12는 ‘SNS 커뮤니티활동(SNS community activity)’이 적합할 것으로 판단하였다.

4.4. 각 토픽의 트렌드 분석

본 연구에서는 토픽 모델링을 통해 특정 논문에 할당된 각 토픽의 비율을 산출하였다. 이렇게 산출된 비율과 각 논문의 발행연도를 활용하여 국내외 논문에서 도출한 각 토픽의 연도별 평균 비율을 계산하고, 이를 기반으로 시간의 흐름에 따른 토픽의 증감을 시각화 한 결과는 Appendix 3과 Appendix 4와 같다. 그 결과, 국내논문에서는 토픽 1(SNS 영화마케팅), 토픽 2(SNS 쇼핑서비스), 토픽 13(SNS 영상마케팅) 순으로 연도별 변동이 가장 큰 것으로 보이며, 해외논문에서는 토픽 2(SNS 오피니언마이닝), 토픽 1(SNS 쇼핑서비스) 순으로 변동이 큰 것으로 나타났다.

그러나 Appendix 3과 Appendix 4에 제시한 시각화 결과는 연도별 토픽의 동향을 대략적으로 파악할 수 있지만 해당 토픽이 학술논문에서 전반적으로 활발히 다루지는 것인지에 대한 판단이 어렵다. 따라서 본 연구에서는 국내외 논문에서 추출한 각 토픽

픽의 연도별 추이에 대한 상승 및 하락 추세를 판단하기 위해 단순회귀분석을 수행하였다. 이를 위해 독립변수는 논문의 발행연도로 설정하고 종속변수를 해당 연도 토픽의 비중을 평균 내어 설정한 후, 회귀분석을 실시하고 도출된 회귀계수 값을 이용하여 각 토픽들의 연도별 추세를 확인하였다. 분석 결과로 나타난 회귀계수가 통계적으로 유의미한 양의 계수이면 Hot 토픽이고 음의 계수이면 Cold 토픽이며, 통계적으로 유의미하지 않은 경우는 Neutral 토픽으로 해석할 수 있다(Griffiths & Steyvers, 2004).

먼저 국내논문에 대한 토픽의 트렌드 분석 결과는 <표 6>과 같다. Hot 토픽은 토픽 2(B=.002, p<.05), 토픽 3(B=.001, p<.01), 토픽 4(B=.002, p<.01), 토픽 8(B=.001, p<.01), 토픽 9(B=.002, p<.01), 토픽 13(B=.009, p<.01)이며 이 외의 토픽은 Neutral 토픽인 것으로 나타났다. 즉, 국내에서 시간이 흐름에 따라 더욱 관심이 주목되고 있는 연구 주제는 SNS 쇼핑서비스, SNS 식품마케팅, SNS 관광마케팅, SNS 뷰티마케팅, SNS 스포츠마케팅, SNS 영상마케팅인 것으로 확인하였다.

Table 6: Results of trend analysis for domestic academic papers' each topic

	Topic	Coef	p	Trend
T1	SNS film marketing	.002	.218	-
T2	SNS shopping service	.002	.049	Hot
T3	SNS food marketing	.001	.006	Hot
T4	SNS tourism marketing	.002	.002	Hot
T5	SNS CSR activity	-.002	.315	-
T6	SNS fashion marketing	-9.45e-05	.941	-
T7	SNS native advertising	-.0001	.930	-
T8	SNS mobile commerce	.001	.002	Hot
T9	SNS beauty marketing	.002	.000	Hot
T10	SNS sports marketing	.0002	.838	-
T11	SNS influencer marketing	-.01	.247	-
T12	SNS ease of use	-.001	.863	-
T13	SNS video marketing	.009	.000	Hot

다음으로 해외논문에 대한 토픽의 트렌드 분석

결과는 <표 7>과 같이 토픽 1(B=.001, p<.01), 토픽 2(B=.001, p<.01), 토픽 3(B=.001, p<.05), 토픽 4(B=.0004, p<.01)이 Hot 토픽이며 이 외에는 Neutral 토픽으로 나타났다. 즉, 해외에서는 SNS 쇼핑서비스, SNS 관광마케팅, SNS 고객인게이지먼트와 SNS 오피니언마이닝에 대해 꾸준한 학술적 관심을 보이고 있었다.

Table 7: Results of trend analysis for overseas academic papers' each topic

	Topic	Coef	p	Trend
T1	SNS shopping service	.001	.000	Hot
T2	SNS opinion mining	.001	.000	Hot
T3	SNS tourism marketing	.001	.012	Hot
T4	SNS consumer engagement	.0004	.000	Hot
T5	SNS tobacco marketing	-.0001	.829	-
T6	SNS food marketing	.0002	.226	-
T7	SNS word of mouth	.002	.054	-
T8	SNS educational application	8.65e-05	.827	-
T9	SNS native advertising	.0003	.139	-
T10	SNS sports marketing	.0002	.260	-
T11	SNS system quality	-.001	.587	-
T12	SNS community activity	.001	.585	-

5. Results and Discussion

본 연구에서는 소셜 미디어 마케팅에 관한 국내외 학술논문의 영문초록을 활용하여 학술적 관심 주제에 대해 탐색 및 비교하였다. 분석 결과를 통해 도출한 인사이트는 다음과 같다.

첫째, 단어 빈도분석 및 동시출현빈도분석 결과, 국내논문은 고객의 구매 의도나 그들의 특성에 초점을 맞추며, 해외논문은 고객과의 커뮤니케이션 수단인 플랫폼과 콘텐츠에 주된 관심을 둔다는 차이가 있었다. 이러한 결과는 소셜 미디어 마케팅에 관한 국내외 연구가 고객의 관점과 기업의 관점을 포괄하여 이루어지고 있으며 각 관점이 모두 현실성 있는 전략을 제시하기에 유용하기 때문이라고 추측할 수 있다. 따라서 소셜 미디어 마케팅 연구의 관

점은 기업 혹은 고객 중 한쪽으로 치우치는 것이 아닌 각 관점 안에서 지속적으로 세분화하는 것이 적합한 연구 방향임을 알 수 있다.

둘째, 국내외 논문에 대해 공통으로 도출된 토픽은 ‘SNS 쇼핑서비스’, ‘SNS 식품마케팅’, ‘SNS 관광마케팅’, ‘SNS 네이티브광고’, ‘SNS 스포츠마케팅’이었다. 이런 결과는 쇼핑 기능을 탑재한 소셜 미디어의 효용성이 학술적으로도 주목받고 있기 때문이라고 예상할 수 있다. 또한 소셜 미디어 마케팅은 다양한 산업이 동등한 조건에서 활용할 수 있으며 언택트 마케팅이 점차 증가하고 있는 현황 속에서 식품, 관광, 스포츠 분야의 소셜 미디어 마케팅 활용이 국내외로 공통적인 학술적 관심을 받고 있음이 드러나는 결과이다. 트렌드 분석 결과에서도 국내외 논문에서 공통적으로 나타난 Hot 토픽은 ‘SNS 쇼핑서비스’와 ‘SNS 관광 마케팅’이었다. 따라서, 소셜 미디어에서 제공하는 쇼핑 기능의 효용성을 연구하거나, 관광 산업 등 본 연구에서 도출된 토픽과 관련된 산업들에 대한 연구가 앞으로도 활발히 지속될 것이라고 예상할 수 있다. 또한 국내외 논문의 공통 관심 주제인 만큼, 내외국인을 모두 대상으로 한 연구가 실효성 높은 전략 도출에 유리할 것으로 보인다.

셋째, 토픽 모델링 결과 중 국내논문에서만 나타난 토픽은 ‘SNS 영상마케팅’, ‘SNS CSR활동’, ‘SNS 패션마케팅’, ‘SNS 모바일커머스’, ‘SNS 뷰티마케팅’, ‘SNS 인플루언서마케팅’, ‘SNS 이용편의성’, ‘SNS 영상마케팅’이고 해외논문에서만 나타난 토픽은 ‘SNS 오피니언마케팅’, ‘SNS 고객인게이지먼트’, ‘SNS 담배마케팅’, ‘SNS 구전효과’, ‘SNS 교육적활용’, ‘SNS 신속성’, ‘SNS 커뮤니티활동’이다. 트렌드 분석 결과에서도 토픽 모델링 결과처럼 국내논문에서는 식품, 관광 등 산업에 특화된 마케팅에 관심이 높고 해외논문에서는 SNS 고객인게이지먼트, 오피니언마케팅 등의 주제가 상승 추세를 보였다. 이를 통해

국내논문에서는 ‘SNS 인플루언서마케팅’과 같은 단기적 성과나 이를 특정 산업에 한정하여 활용하는 것에 초점을 맞추고 있는 반면, 해외논문에서는 ‘고객인게이지먼트’나 ‘구전효과’와 같이 장기적 관점의 자산 축적과 관련이 있는 요소에 초점을 두고 있음을 알 수 있다. 한편, 소셜 미디어 마케팅의 단기 성과를 위한 전략보다 장기 전략이 수익 창출이나 충성 고객 확보가 더 용이하며(Gautam & Sharma, 2017), 소셜 미디어에서의 전략은 특정 산업이나 타겟층에 국한하여 설계되더라도, 기업의 의도와 상관없이 그 타겟층이 일반화된 형태로 빠르게 확대된다(Berthon et al., 2012)고 알려져 있다. 이러한 주장을 본 연구의 결과와 비교해보면, 국내논문에서는 제 3자가 아닌 기업이 자체적인 내부 계획으로 수행하는 소셜 미디어 마케팅을 통한 장기 전략이나 산업 전체를 총체적으로 바라보는 방향으로 연구의 범위를 넓힐 필요가 있음을 알 수 있다. 또한 산업 전체를 망라하는 관점을 적용하면, 오피니언 마이닝과 같이 과학적 접근으로 현실 세계의 현상을 정량화할 수 있는 관점이 더해진다면 연구 결과의 일반화 가능성을 높일 수 있을 것이다.

한편, 본 연구에서 수행한 국내외 토픽 트렌드 분석 결과에서는 감소하는 추세인 토픽이 드러나지 않고 증가 혹은 중립의 토픽만이 도출되었다. 이러한 결과로부터 향후 연구의 방향성을 추가적으로 제안할 수 있다. 향후 연구에서는 본 연구에서 중립으로 나타난 각 토픽에 대한 연구 관점의 변화를 구체적으로 규명하거나, 더욱 구체적인 토픽 트렌드 분석을 위해 각 토픽을 세분화할 수 있는 연구 동향의 분석이 필요할 것이다. 즉, 본 연구에서 도출한 증가 추세의 토픽은 소셜 미디어 마케팅의 향후 연구 방향을 제시하는데 도움이 되며, 명확한 추세가 나타나지 않은 토픽은 연구 동향 탐색의 방향을 구체화하는데 활용할 수 있다.

본 연구의 학술적, 실무적 시사점은 소셜 미디어

마케팅에 대한 현시점의 국내외 연구 관심 주제와 이에 대한 공통점을 규명하고 향후 국내외 연구의 흐름을 예측할 뿐만 아니라, 국내외 연구 관심의 차이점을 통해 국내 연구가 나아가야 할 방향을 제안하였다는 것이다. 이는 실질적인 연구설계의 기초 자료로서 유용할 것이라고 기대하였다. 또한 소셜 미디어 마케팅 연구 동향 분석의 관점을 다양화할 필요가 있음을 추가적으로 제안했다는 점에서 연구 관점의 다양화에 기여하였다는 것이다.

이러한 시사점에도 불구하고, 본 연구는 다음과 같은 한계점을 갖는다. 첫째, 본 연구에서는 ScienceON에서 특정 키워드로 검색하여 수집 가능한 학술논문 데이터만 활용했기 때문에 향후 연구에서는 여러 개의 수집 채널을 이용하거나 검색 키워드를 확장하는 등의 노력이 필요할 것이다. 둘째, 본 연구에서는 토픽을 세분화하기 위해 BERTopic와 LDA를 혼용하여 토픽모델링을 수행하였지만, 토픽 수를 결정하는 과정에서의 객관성 확보에는 한계를 가지므로 향후 연구에서 이를 보완할 필요가 있다. 마지막으로 본 연구에서는 국내외 논문에서 도출한 토픽의 대표 단어들을 참고하여 토픽명을 부여하고 연구 결과를 도출하였으나 각 토픽명은 연구자의 자의성을 내포하고 있기 때문에 향후 연구에서는 토픽 간의 유사성을 정량화 하여 연구 결과의 타당성을 높일 필요가 있다.

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, S. Y., & Chang, P. J. (2021). The effect of coronavirus disease-19 (COVID-19) risk perception on behavioural intention towards 'untact' tourism in South Korea during the first wave of the pandemic (March 2020). *Current Issues in Tourism*, 24(7), 1017-1035.
- 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.
- Berthon, P. R., Pitt, L. F., Plangger, K., & Shapiro, D. (2012). Marketing meets Web 2.0, social media, and creative consumers: Implications for international marketing strategy. *Business horizons*, 55(3), 261-271.
- Blei, D. M. (2012). Probabilistic topic models. *Communications of the ACM*, 55(4), 77-84.
- Bolat, E., & O'Sullivan, H. (2017). Radicalising the marketing of higher education: learning from student-generated social media data. *Journal of Marketing Management*, 33(9-10), 742-763.
- Boyd, D. M., & Ellison, N. B. (2007). Social network sites: Definition, history, and scholarship. *Journal of computer-mediated Communication*, 13(1), 210-230.
- Curry, R. G., & Zhang, P. (2013). Website features that gave rise to social commerce: a historical analysis. *Electronic commerce research and applications*, 12(4), 260-279.
- Deng, Q., Hine, M. J., Ji, S., & Wang, Y. (2021). Understanding consumer engagement with brand posts on social media: The effects of post linguistic styles. *Electronic Commerce Research and Applications*, 48, 101068.
- Gautam, V., & Sharma, V. (2017). The mediating role of customer relationship on the social media marketing and purchase intention relationship with special reference to luxury fashion brands. *Journal of Promotion Management*, 23(6), 872-888.
- 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*.
- Hays, S., Page, S. J., & Buhalis, D. (2013). Social media as a destination marketing tool: its use by national tourism organisations. *Current issues in Tourism*, 16(3), 211-239.
- Huang, O., & Copeland, L. (2020). GEN Z, INSTAGRAM INFLUENCERS, AND HASHTAGS' INFLUENCE ON PURCHASE INTENTION OF APPAREL. *Academy of marketing studies journal*, 24(3), 1-14.
- Kaplan, A. M., & Haenlein, M. (2010). Users of the world, unite! The challenges and opportunities of Social Media. *Business horizons*, 53(1), 59-68.
- Kennedy, H., Kunkel, T., & Funk, D. C. (2021). Using Predictive Analytics to Measure Effectiveness of Social Media Engagement: A Digital Measurement Perspective. *Sport Marketing Quarterly*, 30(4), 265-277.
- Kim, A. J., & Ko, E. (2012). Do social media marketing activities enhance customer equity? An empirical study of luxury fashion brand. *Journal of Business research*, 65(10), 1480-1486.
- Kim, N., Park, J., Kim, K. H., Aiello, G., & Donvito, R. (2012). A study on measuring and defining customer equity of complex shopping malls. *Journal of Global Fashion Marketing*, 3(2), 71-80.
- Lee, J., & Kwon, K. H. (2022). Mobile shopping beauty live commerce changes in COVID-19 pandemic focused on fun contents of MZ generation in Republic of Korea. *Journal of Cosmetic Dermatology*. 21(6), 2298-2306.
- Li, F., Larimo, J., & Leonidou, L. C. (2021). Social media marketing strategy: definition, conceptualization, taxonomy, validation, and future agenda. *Journal of the Academy of Marketing Science*, 49(1), 51-70.
- Martínez-López, F. J., Anaya-Sánchez, R., Fernández Giordano,

M., & Lopez-Lopez, D. (2020). Behind influencer marketing: key marketing decisions and their effects on followers' responses. *Journal of Marketing Management*, 36(7-8), 579-607.

Nadeem, W., Tan, T. M., Tajvidi, M., & Hajli, N. (2021). How do experiences enhance brand relationship performance and value co-creation in social commerce? The role of consumer engagement and self brand-connection. *Technological Forecasting and Social Change*, 171, 120952.

Park, S. B., Cho, K. M., & Byun, H. (2019). The effect of SNS marketing activity on emotional reaction, dance company image, consumer behavior. *Journal of Korean Association of Physical Education and Sport for Girls and Women*, 33(1), 15-39.

Saef, R., Woo, S. E., Carpenter, J., & Tay, L. (2018). Fostering socio-informational behaviors online: The interactive effect of openness to experience and extraversion. *Personality and Individual Differences*, 122, 93-98.

Thackeray, R., Neiger, B. L., Hanson, C. L., & McKenzie, J. F. (2008). Enhancing promotional strategies within social marketing programs: use of Web 2.0 social media. *Health promotion practice*, 9(4), 338-343.

Voorveld, H. A., Van Noort, G., Muntinga, D. G., & Bronner, F. (2018). Engagement with social media and social media advertising: The differentiating role of platform type. *Journal of advertising*, 47(1), 38-54.

Yadav, M., & Rahman, Z. (2017). Measuring consumer perception of social media marketing activities in e-commerce industry: Scale development & validation. *Telematics and Informatics*, 34(7), 1294-1307.

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, W. R., & Yang, H. C. (2022). Overseas Research Trends Related to 'Research Ethics' Using LDA Topic Modeling. *Journal of Research and Publication Ethics*, 3(1), 7-11.

Appendixes

Appendix 1. Results of topic modeling for domestic academic papers

Topic 1 SNS film marketing	Topic 2 SNS shopping service	Topic 3 SNS food marketing	Topic 4 SNS tourism marketing	Topic 5 SNS CSR activity
User(0.05)	Intention(0.07)	Food(0.12)	Tourism(0.23)	Identity(0.05)
Movie(0.05)	Shopping(0.06)	Restaurant(0.11)	Tourist(0.12)	Society(0.04)
Film(0.04)	Purchase(0.05)	Satisfaction(0.05)	Destination(0.09)	Design(0.04)
Communication(0.03)	Product(0.05)	Coffee(0.04)	Travel(0.06)	CSR(0.03)
Message(0.03)	Mall(0.03)	Franchise(0.04)	Experience(0.03)	Culture(0.03)
Topic 6 SNS fashion marketing	Topic 7 SNS native advertising	Topic 8 SNS mobile commerce	Topic 9 SNS beauty marketing	Topic 10 SNS sports marketing
Fashion(0.26)	Advertising(0.23)	Ecommerce(0.08)	Beauty(0.15)	Sport(0.28)
Luxury(0.05)	Type(0.05)	Application(0.05)	Cosmetic(0.14)	Team(0.11)
Trend(0.03)	Facebook(0.03)	Platform(0.05)	Color(0.06)	Player(0.05)
Value(0.03)	Content(0.03)	Trademark(0.04)	Appearance(0.05)	League(0.04)
Clothing(0.03)	Attitude(0.03)	Smartphone(0.04)	Collaboration(0.04)	Event(0.04)
Topic 11 SNS influencer marketing	Topic 12 SNS ease of use	Topic 13 SNS video marketing		
Influencer(0.02)	Market(0.03)	Performance(0.03)		
Message(0.02)	Quality(0.03)	Design(0.02)		
Survey(0.02)	Experience(0.01)	Application(0.02)		
Interaction(0.01)	Technology(0.01)	Video(0.01)		
Design(0.01)	System(0.01)	Space(0.01)		

Appendix 2. Results of topic modeling for overseas academic papers

Topic 1 SNS shopping service	Topic 2 SNS opinion mining	Topic 3 SNS tourism marketing	Topic 4 SNS consumer engagement	Topic 5 SNS tobacco marketing
Ecommerce(0.08)	Sentiment(0.13)	Tourism(0.17)	Engagement(0.08)	Tobacco(0.15)
Trust(0.07)	Opinion(0.08)	Destination(0.14)	Facebook(0.08)	Cigarette(0.07)
Purchase(0.07)	Review(0.06)	Travel(0.06)	Page(0.05)	Ecigarette(0.06)
Shopping(0.04)	Recommendation(0.04)	Visitor(0.04)	Motivation(0.05)	Health(0.04)
Online(0.03)	Text(0.04)	Management(0.02)	Behavior(0.03)	Campaign(0.04)
Topic 6 SNS food marketing	Topic 7 SNS word of mouth	Topic 8 SNS Educational application	Topic 9 SNS native advertising	Topic 10 SNS sports marketing
Food(0.23)	Twitter(0.09)	Student(0.18)	Advertising(0.21)	Sport(0.24)
Child(0.06)	Diffusion(0.08)	University(0.14)	Communication(0.03)	Fan(0.12)
Adolescent(0.05)	Algorithm(0.06)	Education(0.08)	Attitude(0.03)	Athlete(0.06)
Beverage(0.03)	Propagation(0.04)	Learning(0.05)	Advertiser(0.03)	Olympic(0.04)
Meat(0.03)	Event(0.03)	Classroom(0.04)	Value(0.03)	Player(0.04)
Topic 11 SNS system quality	Topic 12 SNS community activity			
System(0.01)	Community(0.01)			
Site(0.01)	Engagement(0.01)			
Time(0.01)	Behavior(0.01)			
Interaction(0.01)	Framework(0.01)			
Online(0.01)	Way(0.01)			

Appendix 3. Variation of topics by year with domestic academic papers

