RESEARCH ARTICLE

Open Access

What Semiotic Values Do Cosmetics Consumers Consume?: Analysis of Cosmetic Brand Reputation and Brand Selection Attributes Using Text Mining

Man Seok Song

Department of Financial Accounting Office, University Of Gyeongnam Namhae, Gyeongnam, Korea

Corresponding author: Man Seok Song,
Department of Financial Accounting Office,
University Of Gyeongnam Namhae, 30,
Hwajeon-ro 78beon-gil, Namhae-eup,
Namhae-gun, Gyeongnam 52422, Korea
Tel.: +82 55 254 2570
Fax: +82 55 254 2579
Email: consulting89@hanafos.com

Received May 8, 2021
Revised May 31, 2021
Accepted June 09, 2021
Published June 30, 2021

Abstract

Purpose: This study aims to empirically analyze how consumers perceive their consumption of the semiotic value from cosmetics that they purchased by using text-mining analysis techniques-word cloud analysis, semantic network analysis, and sensitivity analysis-and examining the subjective emotional big data that consumers have left as discourse reviews concerning the brand reputation and selection attributes of cosmetics in social media. **Methods:** R version 3.6.2-RStudio Version 1.4.1103 was used to collect and analyze the review data on the semiotic value experience that cosmetics consumers left after their purchases. The research was conducted in three stages: the pre-processing of collected data and setting of stopwords, the execution of modeling, and the result analysis. **Results:** The word cloud analysis evinced that the word “skin” appeared the most, followed by “purchase” and “nature”. The fact that such keywords featured prominently suggests that cosmetics consumers described their experience of their purchased cosmetics based on a semiotic value. The semantic network analysis revealed that “nature”, “product”, “real”, “cover”, “use”, and “skin cosmetics” had high levels of degree centrality, betweenness centrality, and closeness centrality. Finally, through the sensitivity analysis, 4,742 words showing positive reviews, 2,039 words showing negative reviews, and 14,146 showing neutral reviews were found. **Conclusion:** In the word cloud analysis, the top 30 keywords represented the purchase goal and selection attribute of cosmetics consumers well and are considerably important for consumption factors related to cosmetics. “Real”, “product”, and “skin cosmetics”, all of which were high in betweenness centrality, degree centrality, and closeness centrality in the semantic network analysis, play an important role in spreading and connecting “meaning” toward particular cosmetics brands on a daily basis. Finally, the sensitivity analysis found that positive emotional reviews appeared approximately 2.33 times more often than negative ones.

Keywords: Semiotic value, Text mining, Word cloud analysis, Semantic network analysis, Sensitivity analysis

Introduction

프랑스의 포스트모더니즘 사상가 Baudrillard (1972)는 소비자들은 상품을 더 이상 사용가치, 교환가치 중심인 기능적 사용으로 인한 가치획득이 목적이 아닌 상품이 지니고 있는 상징적 의미인 기호적 관점에서 소비하기 시작하고 소비를 통해 자신의 존재감을 드러내며 외부로 표현하는 시대로 접어드는 소비가치와 더불어 상징적 의미로

브랜드 또는 기호도 소비하게 된다고 하였다. 이러한 현상은 현대 소비가 갖는 사회적 커뮤니케이션의 한 과정이자 계층의 분류 및 사회적 지위의 차별화 과정이라 할 수 있으며, 최근의 기업은 브랜드에 기호의 의미를 강화하는 방식으로 유사한 다른 기업 브랜드와의 차별화를 만들어 내기 시작했다.

소비자는 기호가치로 소비한 화장품에 대한 경험적 가치인 주관적 감성을 비정형화된 자기표현을 담은 형식으로 인터넷과 소셜 미디어

에서 구매한 화장품 브랜드의 평판과 선택 속성에 대해 만족/불만족의 경험 리뷰를 공유하고 확산도 하여 시장에 영향력을 행사하고 자신만의 정체성을 추구하는 새로운 소비문화를 창출하고 있어 소비자가 직접 작성한 리뷰는 제품 또는 서비스에 대해 소비자의 솔직한 생각을 알 수 있는 중요한 정보 원천으로서 오늘날 전자상거래에서 중요한 부분으로 자리 잡고 있다(Kostyra *et al.*, 2016).

인터넷과 소셜 미디어에 담론 형식인 소비자 리뷰로 생성된 비정형 데이터는 연구자의 개입이나 편견이 반영되지 않은 자발적으로 생성되는 실시간 데이터로 있는 그대로의 인식, 태도, 행동 등을 발견할 수 있기 때문에 전통적인 양적 및 질적 데이터를 보완하는 중요한 자료로써 인식되고 있으며(Lee, 2015; George *et al.*, 2014; Kostyra *et al.*, 2016), 제품 또는 서비스에 대해 사용자의 솔직한 생각을 알 수 있는 중요한 정보 원천으로 오늘날 전자상거래에서 중요한 부분으로 자리 잡고 있다.

브랜드에 대한 긍정적인 평판은 소비자에게 브랜드에 대한 기대를 가능하게 하며 구매 위험을 감소시킬 수도 있고 소비자의 구매의사결정 과정에 중요한 영향을 미치며 기업의 성과 향상에 전략적으로 중요한 역할을 한다(Walsh *et al.*, 2006; Koporcic & Halinen, 2018). 따라서 개별 화장품 브랜드가 보유한 브랜드 파워도 중요하지만 소비자가 갖는 호의적인 지각된 브랜드 선택 속성의 구축 또한 기업 브랜드의 경쟁력 강화를 위한 장기계획수립 및 집행에 매우 효과적이라고 할 수 있다.

최근 소비자는 화장품 구매시 인터넷 또는 소셜 미디어에서 제품에 대한 정보를 직접 검색하고 타인의 구매후 평가 리뷰를 수렴하기도 하며 자신이 구매한 제품에 대한 리뷰를 담론 형식으로 공유하는 상호작용적인 커뮤니케이션 활동을 하며, 화장품 구매시 화장품이 지니고 있는 기능적 가치뿐만 아니라 사용가치(유용성)와 더불어 화장품에 담긴 상징적 의미인 브랜드 또는 기호가치(상징적 의미)도 소비한다. 즉, 사용가치는 소비의 전제조건에 불과하며 기호가치가 현대사회에서 소비의 진정한 의미라고 할 수 있다(Lee, 2011).

소비자가 기호 소비한 제품을 평가할 때 브랜드만으로 평가하는 경우가 많기 때문에 높은 브랜드 평판을 가지고 있는 화장품 브랜드는 경쟁에서 더욱 유리하다. Kim & Kim (2018) 그리고 Walsh *et al.* (2006) 등은 기업의 브랜드 평판이 소비자의 의사결정 과정에 중요한 영향을 미친다고 하였으며, Gray & Balmer (1998)와 Weiss *et al.* (1999)은 평판은 단기간이 아닌 오랜 시간 형성된 지각과 일치하며 내부 및 외부적으로 다양한 이해관계자와 관련이 깊다고 하였고, 소비자에게 제품에 대한 기대를 가능하게 하며 구매 위험을 감소시킬 수도 있다(Padberg *et al.*, 1974).

화장품 브랜드와 관련한 기존 연구들은 주로 화장품 선택 속성과 만족도 및 충성도와 선행요인들에 관련된 인과관계 연구(Namgung *et al.*, 2017; Oh, 2020; Lee *et al.*, 2015; Hyun & Park, 2012; Kim & Hwang, 2016)로서 수행되어져 왔으며, 소비자의 화장품 선택 의도를 설명할 수 있는 새로운 변인들을 밝히려는 시

도(Kim *et al.*, 2018; Song, 2018)도 하여 왔다.

이들 연구는 화장품 이용 소비자가 브랜드 평판과 브랜드 선택 속성에 대해 지각하는 정도를 파악하여 어떠한 속성이 구매의도에 영향을 미치는가를 규명하는 소위 화장품 브랜드 선택 속성에 대한 소비자의 절대적 가치평가와 종속변수로서 만족도 또는 충성도 그리고 구매의도와 인과관계를 제시한 연구들이었다.

그러나 이러한 연구는 소비자만의 공간에서 능동적이며 경험적 문화가치가 생성되고 전파되어지는 기호가치 감성소비를 하는 소비자 행동에 대한 태도의 탐색이나 설명에는 이르지 못하고 있으며, 브랜드 강화 전략이 절실히 요구되는 상황에서 의사결정 시스템으로서의 전략적 시사점을 제공하기에는 부족함이 많다.

인터넷과 소셜 미디어에 고객이 담론 형식으로 남긴 비정형 데이터는 연구자의 개입이나 편견이 반영되지 않은 자의적인 데이터로 공존하고 있어 고객 리뷰는 제품 또는 서비스에 대해 사용자의 솔직한 생각을 알 수 있는 중요한 정보 원천으로 자리 잡고 있기 때문에(Kostyra *et al.*, 2016) 많은 연구자와 기업은 고객 리뷰를 지능적으로 분석하여 전통적인 연구 방법을 보완하고자 시도하고 있다(Trenz & Berger, 2013; Ghose & Ipeirotis, 2007).

이에 본 연구는 화장품 이용 소비자가 가장 많이 이용하는 소셜 미디어를 선정하여 담론 형식으로 장기간에 걸쳐 존재하는 기호가치 소비의 경험가치 리뷰 빅데이터(big data)를 크롤링(crawling)하여 소비자가 화장품 구매시 우선적으로 고려하는 화장품 브랜드 평판과 선택 속성과 관련된 유의미한 감성 단어를 파악하기 위해 워드클라우드 분석(word cloud analysis)과 의미연결망 분석(semantic network analysis) 그리고 감성 분석(sentiment analysis)을 적용하여 접근하고자 한다.

Methods

1. 데이터의 수집과 전처리

본 연구인 텍스트 마이닝(text mining)을 이용한 화장품 브랜드 평판과 브랜드 선택속성 분석에 대한 연구를 위해 연구자의 개입이나 편견이 반영되지 않은 화장품 소비자 자의에 의해 자연어 담론 형식으로 소셜 미디어에 사용 후 그대로의 인식, 태도, 행동 등이 장기간에 걸쳐 작성된 방대한 감성 리뷰의 비정형 빅데이터 수집이 필요하였다.

이를 위해 소비자 주도형 화장품 구매 우선순위와 솔직한 사용후 리뷰가 담론 형태로 존재하고 있는 글로윙크와 네이버 쇼핑 화장품 카테고리에 일 특정 화장품 브랜드에 대한 경험가치 리뷰 빅데이터를 크롤링하여 소비자가 화장품 구매후 남긴 화장품 브랜드에 대한 평판 그리고 선택 속성과 관련된 유의미한 감성 단어의 파악과 연관어 분석을 위해 워드클라우드 분석과 의미연결망 분석 그리고 감성 분석의 기법을 적용하기로 하였다.

이들 분석기법의 적용을 위해 우선되어야 하는 빅데이터의 수집을 위한 분석 도구로는 R version 3.6.2-RStudio Version 1.4.1103 프로그램을 이용하여 화장품 소비자가 구매후 남긴 기호가치 경험 리뷰 데이터를 수집하였으며, 수집된 데이터에 대해 데이터의 전처리와 불용어의 설정, 모델링의 실행, 결과 해석까지 총 3단계에 걸쳐 연구를 진행하였다.

R은 뉴질랜드 오클랜드 대학의 로버트 젠틀맨(Robert Gentleman)과 로스 이하카(Ross Ihaka)에 의해 시작된 통계와 그래픽 그리고 빅데이터 분석에 널리 사용되는 오픈 소스 프로그래밍 언어이자 소프트웨어 환경으로 오픈소스(open source)로 쓰여졌으며 무료로 누구나 R의 홈페이지(<https://r-project.org>)에서 다운로드 받을 수 있으며, R에서 사용할 수 있는 수많은 통계 패키지 개발되어 있어 사용자가 제작한 패키지를 추가하여 기능을 확장할 수 있다.

RStudio 프로그램은 캐나다의 R-Tools Technology Inc.에서 개발한 것으로 통계 컴퓨팅, 그래픽스를 위한 프로그래밍 언어인 R을 사용하는 데 있어 편리하고 유용한 기능을 제공하고 있고 오픈소스 버전과 상용 버전 두 가지로 이루어져 있으며, RStudio사의 홈페이지에서 다운로드 받을 수 있다(<https://www.rstudio.com>). 본 연구에서는 RStudio의 오픈소스 에디션을 이용하였다.

전처리란 크롤링된 데이터를 분석하는 작업 전에 데이터를 분석하기 좋은 형태로 만드는 과정으로 데이터 분석 과정에서 데이터 전처리가 중요한 이유는 아무리 좋은 분석 기법이나 도구를 사용하더라도 품질이 낮은 데이터로는 좋은 분석 결과를 얻기 힘들기 때문이다(Kim *et al.*, 2021)에 수집 데이터는 분석하기 용이한 형태로 정제하는 과정을 거쳐야 한다.

구체적으로 데이터 전처리 과정에선 수집 텍스트 내에서 동일하거

나 유사하지만 다르게 표현된 단어들을 하나의 단어 형태로 통일하는 정규화(normalization) 작업과 분석에 영향을 미칠 수 있는 불필요한 단어 및 어구를 삭제하는 작업을 수행하였다. 이를 위해 수집한 빅데이터(big data)를 텍스트마이닝(tm) 라이브러리(library)를 적용하여 말뭉치(corpus)로 변환한 후 데이터 전처리 과정을 수행하였으며, 한글 자연어 처리 KoNLP 라이브러리 함수를 적용하였고, 형태소 사전으로는 한국정보화진흥원 사전(NIADic)을 이용하였으며, 품사 구분을 위해서는 KoNLP에 있는 SimplePos09()를 적용하였다.

한국어로 이뤄진 텍스트 마이닝 분석을 위해서는 명사와 조사 혹은 동사의 어간과 어미 등으로 구성된 어절을 의미 기능 부분과 문법 기능 부분으로 분리하는 작업을 거쳐야 하나 이는 외형적으로 판단하기 어렵기(Lee & Gil, 2019) 때문에 토픽 모델링을 활용한 다수의 텍스트 마이닝 선행 연구에서 텍스트 내에서 명사만을 추출해 분석하고 있으며, 이런 관점과 더불어 특정 주제를 구성하거나 이와 연관 의미를 내포하는 키워드가 대부분 명사 형태를 가지는 경우가 많기 때문에(Bae *et al.*, 2014) 본 연구에서의 분석을 위한 형태소를 선행연구에서와 같이 명사로 한정하였으며 나머지는 불용어로 처리하였다.

2. 분석 방법

소비자가 남긴 빅데이터를 분석하는 연구가 활발하게 수행되면서 다양한 학문 연구 분야에서 빅데이터 분석 방법인 텍스트 마이닝 기법을 활용한 연구가 활발하게 수행되고 있다(Zhu *et al.*, 2018; Kim & Kim, 2016). 텍스트 마이닝은 정보검색과 정보 추출 및 자연어 처리 기술을 사용하여 알려지지 않은 유용한 패턴과 지식을 텍스트로부터 발견하는 것을 말하며(Han & Yoon, 2016), 텍스트 내에서 메

Table 1. Top 30 keywords and frequency of word cloud analysis

NO	Keywords	Frequency	Cumulative frequency	NO	Keywords	Frequency	Cumulative frequency
1	Skin	4,922	4,922	16	Cover	1,901	43,561
2	Purchase	3,721	8,640	17	Lasting power	1,875	45,436
3	Nature	3,705	12,348	18	Last year	1,847	47,283
4	Essence	3,453	15,801	19	UV-rays	1,813	49,096
5	Makeup	2,976	18,777	20	Retouch	1,800	50,896
6	Once	2,886	21,663	21	For correction	1,798	52,694
7	Tone up	2,715	24,378	22	Blemishes	1,774	54,468
8	Product	2,621	26,999	23	Brighter	1,725	56,193
9	Face	2,413	29,412	24	Formulation	1,575	57,768
10	Really	2,264	31,676	25	Repurchase	1,412	59,180
11	Cosmetics	2,085	33,761	26	Only once	1,352	60,532
12	Skin cosmetics	2,074	35,835	27	Cream	1,268	61,800
13	Morning	1,963	37,798	28	About	1,249	63,049
14	Use	1,949	39,747	28	Think	1,236	64,285
15	Covering power	1,913	41,660	30	Completely	1,125	65,410

시지를 형성하는 단어들은 어떻게 사용했는지에 대한 구조의 패턴과 의미를 분석함으로써 관계의 형태 속에 내포되어있는 메시지의 내용까지 도출하는 분석 방법이다(Cha, 2015; Han, 2003).

본 연구의 목적인 화장품 소비자가 소셜 미디어에 작성한 화장품 기호가치 구매 후 사용한 화장품의 평판과 선택 속성에 대한 리뷰를 텍스트 마이닝 분석 기법인 워드클라우드 분석, 의미연결망 분석, 그리고 감성 분석을 이용하여 일 특정 화장품 브랜드에 대한 기호가치 소비에 대한 논의를 전개하는 연구 문제의 수행을 위해 R version 3.6.2-RStudio Version 1.4.1103 프로그램을 이용하여 Figure 1, Figure 2, 그리고 Figure 3과 같이 분석 결과를 시각화할 수 있었으며, Table 1, Table 2, 그리고 Table 3과 같은 정량화된 분석 결과를 얻을 수 있었다.

Results and Discussion

본 연구에서는 소비자가 주도하는 화장품 브랜드 구매 우선순위와 리뷰가 존재하고 있는 글로우픽과 네이버 쇼핑 화장품 카테고리에서 담론 형식으로 장기간에 걸쳐 존재하는 일 특정 화장품 브랜드에 대한 기호가치 소비의 경험가치인 리뷰 빅데이터를 R version 3.6.2-RStudio Version 1.4.1103 프로그램을 이용하여 도출 결과를 직관적으로 파악하기 위해 시각화 작업을 진행하였다.

Figure 1인 워드클라우드 분석의 결과로 나타나는 주제가 분석, Figure 2의 의미연결망 분석, 그리고 Figure 3과 같이 기계학습을 통해서 이루어지는 감성 분석을 실시하여 화장품 기호가치 소비에 대한 시각화 분석 결과의 정량적인 설명을 위해 Table 1, Table 2, 그리고 Table 3으로 제시하였다.

1. 워드클라우드 분석 결과

워드클라우드와 관련한 빈도 분석에서는 수집된 단어 전체를 전처리 한 후 R 프로그램의 wordcloud 2 라이브러리 함수를 사용하여 워드클라우드 분석을 실행하여 소비자가 구매 후 남긴 담론에 대한 텍스트 데이터를 Figure 1과 같이 시각적으로 확인할 수 있었으며, 모델링의 대상이 되는 수집 담론에 대해 출현 빈도가 높은 키워드 상위 30개 주요어에 대한 빈도수와 누적 빈도수의 순위는 Table 1과 같이 확인할 수 있다. Figure 1의 워드클라우드는 단어의 글자 크기가 클수록 해당 단어의 빈도수가 높다는 것을 의미하며, 워드클라우드의 중심부로 갈수록 빈도수가 높은 단어이며 외곽에 배치되거나 글자 크기가 작은 것은 상대적으로 빈도수가 낮은 단어이다.

본 연구를 위해 크롤링한 전체 말뭉치에서 각 단어가 총 몇 회 등장했는지 R 프로그램으로 도출한 결과 도출 빈도수가 높은 상위 30개 단어는 Table 1과 같으며, 이를 도출 빈도순으로 정렬하여 누적 빈도수와 함께 표기하였다. 이중 가장 도출 빈도가 높은 상위 5개의 단어는 30개 단어 수의 약 28.7%를 차지하고 있는 것으로 나타났다.

Figure 1. Wordcloud analysis results.

As a result of word cloud analysis, top keywords represent the purchase goal and selection attribute of cosmetics consumers, which are important cosmetic consumption factors. The word “skin” appeared the most, followed by “purchase” and “nature.” Such prominently featured keywords suggest that experience value is well-reflected in consumers’ selection attribute of cosmetics they purchased based on semiotic value.

이들 중 '피부'라는 단어가 가장 높게 나타났으며 그 다음으로는 '구입', '자연', '에센스', '화장'의 순으로 높은 빈도를 보이고 있으며, 하위 5개 그룹은 '하나', '크림', '정도', '생각', '완전'의 순으로 나타났다. 이들 주제가 두드러지게 나타난 것은 기호가치로 소비한 화장품 이용 소비자의 화장품 선택 요인의 경험 가치를 다양하게 나타내고 있으며 이를 확인할 수 있었다.

2. 의미연결망 분석 결과

의미연결망 분석은 네트워크 정보학의 하나로 의미를 갖는 텍스트를 분석 대상으로 하여 언어 메시지에서 의미가 있는 단어를 추출하고 그들 간에 구성되는 의미적 연결 관계를 토대로 네트워크를 구성하고 네트워크의 분석지표를 활용하여 언어 메시지의 다양한 의미적 특성을 분석하는 방법을 말하는 것으로 사회 연결망 분석(social network analysis)을 확장한 개념으로 단어의 출현 빈도와 관계를 확인하여 메시지에 내재되어 있는 의미 혹은 개념적인 특성을 추출하고 이들 간의 관계적 속성을 파악하는 데 유용한 분석 방법이다(Lee, 2014).

단어 사이의 연관 관계를 확인하기 위한 지표로는 매개중심성(betweenness centrality)과 연결중심성(degree centrality) 그리고 근접중심성(closeness centrality) 등을 사용한다. 의미연결망을 통하여 도출되는 노드(node)가 그래프의 중앙에 위치할수록 전체 연결망에서 중심적인 역할을 하는데 연결중심성은 하나의 단어에 직접 연결 단어들의 합으로 연결망에서 중요한 역할을 하는 단어로 파악할 수 있다. 근접중심성은 한 단어에서 다른 단어에 도달하기까지 필요한 최소 단계 수로 각 단어 간 거리를 근거로 중심성을 측정한다. 매개중심성은 연결망 내에서 한 단어가 담당하는 중개자 역할의 정도

Figure 2. Semantic network analysis results.

As a result of semantic network analysis, “nature,” “real,” “cover,” “use,” and “skin” appeared to have high degree, betweenness, and closeness centralities and played an important role in spreading and connecting “meaning” in the semantic network.

를 뜻하며 매개중심성이 높을수록 연결망에서 의미의 확산에 중요한 역할을 한다.

데이터 전처리 과정을 통하여 추출한 화장품 이용 소비자의 답변 형식으로 남겨진 고객 리뷰를 통해 브랜드 선택 속성에 대한 아이텐티티 25개의 노드에 대한 연결 관계를 Figure 2와 같이 의미연결망 분석을 시도하여 시각화하였다. 본 연구의 키워드 중에서 출현 빈도가 상위 그룹에 속하는 주제어를 중심으로 주제어 간의 의미연결망 구조와 중심성을 파악하였다. 데이터 중 너무 많은 키워드를 선정하게 되면 프로그램 운용이 어려울 뿐만 아니라 전체적인 구조의 분석 의미가 광범위하게 분산되고 연결망 구조의 특징 및 중심성과 관련 분석의 의미를 찾기 어렵기 때문에 각 주제어의 출현 빈도 수위와 연구 목적에 적합하게 전처리 과정을 거쳐서 선정하였다.

주제어를 중심으로 의미연결망을 분석한 결과 Figure 2와 같은 노드와 선(line)으로 이루어진 의미연결망이 도출되었다. 텍스트 내에서 도출 빈도가 높을수록 노드에 대한 연결망이 복잡해지고 상대적으로 출현 빈도가 낮으면 노드와 연결되는 선의 연결망 수는 작아진다. 노드와 노드를 연결하는 선은 노드 간의 연관 정도를 나타내고 있으며, 핵심 주제어 간 연관성이 높으면 노드와 노드를 이어주는 선이 많아지고, 연관성이 없으면 선이 연결되지 않는다.

Figure 2와 Table 2의 도출 결과에서 화장품을 기호가치 소비한 답변 형식의 리뷰를 크롤링한 댓글 중에서 매개중심성이 가장 높게

나타난 것은 '재구매'로 나타났으며, 그다음으로는 '진짜', '제품', '커버', '스킨 화장품'의 순으로 나타났다. 이들은 전체 연결망 내에서 담당하는 중개자 역할에서 높은 위치를 나타내고 있음을 뜻하며 연결망에서 의미의 확산에 중요한 역할을 한다. 가장 낮은 값으로 나타난 것은 '작년'으로 나타났으며, 그 다음으로는 '수정용', '화사', '에센스', '보정'의 순으로 나타났다.

연결중심성이 가장 높게 나타난 키워드는 '자연', '진짜', '사용', '스킨 화장품', '제품' 등의 순으로 나타났으며, 이때 '사용'과 '스킨 화장품' 그리고 '제품' 등은 동등하게 각 15개의 연결중심성을 가지고 있는 것으로 나타나 이들이 전체 주제어에서 매우 높은 역할을 하고 있음을 보여주고 있으며, '자외선', '작년', '잡티' 등은 3개의 연결중심성을 가지고 있어 상대적으로 가장 낮은 연결 정도의 강도를 가지고 있는 것으로 나타났다.

근접중심성이 높게 나타난 것은 '진짜', '제품', '얼굴', '스킨 화장품', '화장' 등의 순서로 나타났는데 이들은 다른 단어에 도달하기까지 필요한 최소한의 근접중심성이 크기 때문에 의미연결망의 중앙에 위치한다. 근접중심성이 가장 낮게 나타난 것은 '작년'이며 그다음으로는 '잡티', '지속력', '자외선', '화사'의 순으로 근접중심성 정도가 낮게 나타났다.

그리고 '진짜', '제품', '스킨 화장품' 등의 경우 본 연구의 의미연결망 분석에서 연결중심성과 매개중심성 그리고 근접중심성 모두가 높

게 나타나 본 연구를 위한 의미의 확산과 연결에 중요한 역할을 하고 있는 것으로 나타났으며 수집된 담론의 워드클라우드 분석 결과와 유사한 유의미한 중심성 강도의 표현으로 나타났다. '작년'과 '화사', '지속력', '자외선', '잡티' 등은 본 연구에서 상대적으로 낮은 정도의 매개중심성과 연결중심성 그리고 근접중심성 강도를 가지고 있는 것으로 나타났다.

3. 감성 분석 결과

화장품 브랜드 생산 기업에 있어 고객의 감성을 분석하는 것은 기업 성과분석에 기여하는 바가 크며 엄청난 가치를 제공하지만 이를 수행하는 것은 매우 어려운 과제이기 때문에 문제 해결을 위한 방법으로 텍스트 마이닝에 의한 고객 감성 및 의견에 대한 연구는 비교적 일찍부터 시작되었다(Morinaga *et al.*, 2002; Tong & Koller, 2001).

감성 분석은 다양한 기계학습을 통해서 이루어지고 있으며, 텍스트에 나타난 감정을 분류하거나 수치화하는 객관적인 정보로 바꾸는 기술을 말하며(Kim & Song, 2016), 긍정과 부정 등 감성에 대한

분석을 sentiment analysis라 하며 국내에서는 감성 분석이라는 용어로 번역되어 사용되고 있다(Oh & Chae, 2015). 앞서 연구 분석한 워드클라우드 기법과 의미연결망 분석이 텍스트가 다루는 대상을 파악하는 텍스트 마이닝 기법이었다면, 감성 분석은 텍스트에 담겨져 있는 태도를 추정하는 텍스트 마이닝 기법이다.

본 연구의 감성 분석을 위해 딥 러닝(deep learning) 기법을 이용하여 대용량의 학습 데이터를 생성하는 방안으로 접근하고자 한다. 딥러닝은 인공신경망의 학습기법으로(LeCun *et al.*, 2015; Schmidhuber, 2015) 최근 인공지능 분야에서 많은 주목을 받고 있으며, 인간 뇌의 정보처리 과정을 수학적인 모델링을 통해 묘사한 모형으로 인공신경망 학습에서 발생하던 기존의 제약사항들을 부분적으로 해소함으로써 페이스북은 사용자의 얼굴인식에 적용하여 정확도를 크게 높였으며(Taigman *et al.*, 2014), 음성인식 분야에서도 활발히 활용되고 있고(Hinton *et al.*, 2012; Graves *et al.*, 2013), 텍스트 마이닝의 한 영역인 감성분석에도 다양한 딥러닝 기법이 적용되어 의미있는 성과를 거두고 있다(Lai *et al.*, 2015).

감성분석은 텍스트 데이터에 내포되어 있는 텍스트 작성자의 의

Table 2. Results of semantic network analysis

No	Keywords	Betweenness centrality	Degree centrality	Closeness centrality
1	Skin	12.900	13	0.0263
2	Repurchase	64.200	6	0.0222
3	Nature	7.790	17	0.0238
4	Essence	3.890	9	0.0233
5	Makeup	31.500	14	0.0270
6	Face	31.000	14	0.0294
7	Cosmetics	6.430	13	0.0263
8	Morning	17.600	10	0.0250
9	Really	47.400	16	0.0303
10	Formulation	15.000	11	0.0263
11	Lasting power	7.240	4	0.0169
12	Cover	41.900	11	0.0263
13	UV-rays	14.400	3	0.0192
14	For correction	0.739	5	0.0208
15	Tone up	18.700	8	0.0244
16	Brighter	2.660	5	0.0196
17	Only once	19.800	11	0.0256
18	Retouch	4.350	6	0.0217
19	Covering power	9.970	9	0.0244
20	Purchase	9.510	9	0.0244
21	Use	22.100	15	0.0270
22	Last year	0.167	3	0.0156
23	Skin cosmetics	38.700	15	0.0286
24	Product	42.500	15	0.0298
25	Blemishes	15.200	3	0.0167

Figure 3. Sensitivity analysis results.

As a result of sensitivity analysis, “positive,” “neutral,” and “negative” values are indicated by blue, green, and red colors, respectively, in the pie chart. There were 4,742 words showing positive reviews, 2,039 words showing negative reviews, and 14,146 showing neutral reviews.

견이나 감성, 평가, 태도 등을 분석하는 일련의 과정으로(Liu, 2012) 본 연구에서는 기존에 구축되어 있는 감성사전의 어휘를 사용하여 충분한 양의 학습 데이터를 구축하고 딥러닝 기법으로 학습하여 그 모델을 이용한 감성 분석을 통해 본 연구를 위한 결과물을 얻도록 하였다.

이용 가능한 우리나라 감성사전으로 많이 활용되고 있는 것은 서울대학교에서 구축한 한국어 감성 분석 코퍼스(Korean Sentiment Analysis Corpus, KOSAC)와 군산대학교에서 구축한 한국어 감성사전(KNU Korean Sentiment Lexicon, KNU-KSL, 이하 KNU로 표기)이 있으며, 이들 감성사전을 통해 학습 데이터를 구축하면 크롤링 데이터에 내재된 감성있는 많은 양의 레이블 학습 데이터를 추출할 수 있다는 이점이 있으며, 높은 수준의 감성 분석을 할 수 있을 뿐만 아니라 많은 시간을 단축할 수 있다는 장점이 있다. 본 연구에서는 KNU를 이용한 감성 분석을 실시하였으며, 그 결과는 Figure 3 그리고 Table 3과 같다.

일 특정 화장품 브랜드 기업에 대한 기호가치 소비한 화장품 소비자의 경험적 담론 리뷰에 대한 감성 분석 결과는 긍정적 반응을 보인 것은 4,742단어이며, 부정적인 반응 결과 단어는 2,039, 그리고 중

립으로 나타난 단어는 14,146으로 나타났다. 감성 분석 결과 긍정적인 감성 반응이 부정적인 감성 반응보다 약 2.33배 높게 나타난 것은 연구 대상인 일 특정 화장품 브랜드 기업의 브랜드 평판이 시장에서 좋은 반응을 보이는 것으로 판단되며 해당 기업에게는 상당히 긍정적인 시장 반응으로 볼 수 있으나 지속적인 시장 경쟁력 강화를 위해서는 세심한 관리가 요구되어진다고 할 수 있다.

Conclusion

IT 기술의 발달과 함께 소비자들의 구매 채널이 오프라인에서 온라인으로 급격히 변화되고 있는 가운데 예기치 못한 COVID-19 팬데믹 환경은 소셜 미디어로의 소비 트렌드를 더 가속화하고 있다. 본 연구는 화장품 소비자가 기호가치 소비로 소셜 미디어에서 구매하는 화장품의 브랜드 평판과 선택 속성을 분석하기 위해 자연어 말뭉치 텍스트 리뷰로 남겨진 빅데이터를 크롤링하여 실증분석 과정을 거치면 소비자는 그에 대해 어떻게 지각하고 있는가? 에 대한 의문점에서 시작하였다.

먼저 이들 문제점 해결을 위해서 소비자 주도형 화장품 구매 우선 순위와 솔직한 사용후 경험가치 리뷰가 담론 형태로 존재하고 있는 글로워픽과 네이버 쇼핑 화장품 카테고리에 장기간에 걸쳐 존재하는 일 특정 화장품 브랜드에 대한 기호가치 소비한 비정형화 자기표현의 담론을 크롤링하였다.

화장품 브랜드 평판과 브랜드 선택속성의 분석을 위해서는 다양한 방법이 있으나 본 연구에서는 화장품이라는 특수성에 맞추어 기호가치 소비에 대한 특성에 초점을 두고 기존의 설문이나 인터뷰 방식이 아닌 텍스트 마이닝을 이용하여 화장품 이용 소비자가 담론 형식으로 장기간에 걸쳐 남긴 사용 후 그대로의 인식, 태도, 행동 등으로 나타난 평판과 선택 속성과 관련한 유의미한 단어의 파악과 연관이 분석을 위해 R version 3.6.2-RStudio Version 1.4.1103 프로그램을 이용하여 크롤링하고 워드클라우드 분석, 의미연결망 분석, 그리고 감성 분석을 하였다는 점에 의의가 있다.

첫째, 워드클라우드 분석에서 도출된 상위 30위의 주제어들은 화장품 이용 소비자가 기호가치로 소비한 화장품 구매 목적과 선택 속

Table 3. Results of sensitivity analysis

Positive	Neutral	Negative
4,742	14,146	2,039
1. In the pie chart, ‘positive’ is set in blue, ‘neutral’ in green, and ‘negative’ in red.		
2. R atypical text sensitivity analysis package combining text mining technique with collected data is used.		
3. Sensitivity analysis : Emotional scores are calculated by comparing crawled discourse reviews about emotional consumption based on semiotic value with emotional dictionary of KNU: 1 point or higher is set as positive, 0 point or less as negative, and 0 point as neutral and then textmining technique is applied.		

KNU, Kunsan national university.

성을 잘 나타내고 있는 것으로 보인다. 본 연구 분석 결과에서 전체 문장과 문단이 나타나지 않은 관계로 이와 연결하여 설명하기 힘들지만, 크롤링한 전체 담론에서의 화장품 소비를 위한 요소는 '피부', '자연', '에센스', '톤업', '스킨', '커버력', '지속력', '보정', '잡티', '제형', '크림' 등의 키워드가 중요시되고 있음을 알 수 있다.

둘째, 의미연결망 분석으로 도출된 결과에서 모두 높게 나타난 '진짜', '제품', '스킨 화장품', '재구매', '자연' 등의 주요어는 매개중심성, 연결중심성, 근접중심성 등의 의미연결망에서 의미의 확산과 연결에 중요한 역할을 하고 있으며, 수집된 담론의 표현인 워드클라우드 분석 결과와 유사한 유의미한 중심성 강도의 표현으로 나타났으며, 기호가치 소비에 의해 자주 언급되는 상품과 주요어를 통해서 화장품 유통 플랫폼에서의 소비자들이 선택 구매하는 제품의 선택 속성을 파악할 수 있었다.

따라서 기업에서 이러한 키워드를 중심으로 화장품 기호가치 구매 소비자를 대상으로 하는 브랜드 평판 강화를 위한 노력을 한다면, 이들은 브랜드만으로 평가하는 경우가 많기 때문에 최소한의 노력으로 시장에서 구매위험을 감소시킬 수 있을 뿐만 아니라 기호가치 소비 의사결정을 증폭시킬 수 있어 성공적인 시장 경쟁력 강화와 기업 성과 향상에 전략적으로 도움이 될 것이다.

셋째, 텍스트 마이닝한 기호가치 소비자 담론의 감성 분석 결과 긍정적인 감성 반응이 4,742개 단어로 나타나 2,039개 단어로 나타난 부정적인 감성 표현 반응 보다 약 2.33배 높게 나타난 것은 연구 대상인 일 특정 화장품 브랜드 기업의 브랜드 평판이 시장에서 좋은 반응을 보이고 있는 것으로 판단되며 해당 기업에게는 상당히 긍정적인 시장 반응으로 볼 수 있으나 지속적인 시장 경쟁력 강화를 위해서는 14,146개의 단어로 나타난 중립의 기호가치 소비자 담론을 긍정적인 감성 반응으로 개선하는 세심한 배려가 요구된다. 이는 소비자에게 긍정적으로 지각되는 브랜드 선택 속성의 구축은 브랜드 경쟁력 강화와 장기 계획 수립 및 집행에 매우 효과적이기 때문이다.

연구의 한계점으로는 본 연구는 일 특정 화장품 브랜드 기업을 대상으로 실시한 연구로 전체 화장품 시장에 적용하는 데는 한계가 있어 보인다. 다른 화장품 브랜드군과의 비교 분석 또한 연구의 가치를 강화하는 도움이 필요할 수 있다. 이러한 부분을 보완하기 위해 저가 화장품 브랜드군과 고급화장품 브랜드군 그리고 수입화장품 브랜드군으로 시장을 세분화하여 비교하는 후속 연구도 필요할 것이다.

또한 본 연구는 R version 3.6.2-RStudio Version 1.4.1103 프로그램을 이용하여 독자적으로 전처리 작업과 의미 해석을 진행하였다. 좀 더 많은 여유가 주어진다면 객관적 논리의 의미 해석을 위해 학제적인 전문가 집단과의 공동 작업도 요구되며, 전처리과정에서 형태소를 선행 연구와 같이 명사만으로 한정하였으나 앞으로의 연구에서는 형용사와 동사도 포함하는 연구와 함께 단어와 단어와의 맥락 연결을 찾아내는 바이그램 분석도 같이 진행하는 것도 의미 있는 연구가 될 것이다.

그럼에도 불구하고 본 연구는 화장품 산업 분야에서 워드클라우드

분석, 의미연결망 분석, 그리고 감성 분석을 같이 수행한 초기의 연구라는 점에서 의미를 부여할 수 있으며 이러한 분석을 통해서 일 특정 화장품 브랜드 기업뿐만 아니라 같은 경쟁시장에 있는 기업에게도 시장경쟁력 강화를 위한 가이드라인과 중요한 단서를 제공할 수 있을 것이다.

Acknowledgements

This work was supported by the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea(NRF-2019S1A5B5A07093402).

Author's contribution

MSS alone contributed to this work. The author designed all experiments, performed a pre-test, and collected analyzed and interpreted the data regarding physical and written responses. The author also wrote the manuscript and oversaw the project.

Author details

Man Seok Song (Part-time Lecturer), Department of Financial Accounting Office, University Of Gyeongnam Namhae, 30, Hwajeon-ro 78beon-gil, Namhae-eup, Namhae-gun, Gyeongnam 52422, Korea.

References

- Bae JH, Han NG, Song M. Twitter issue tracking system by topic modeling techniques. *Journal of Intelligence and Information Systems*, 20: 109-122, 2014.
- Baudrillard J. Pour une critique de l'economie politique du signe. Gallimard, Paris, pp23-25, 1972.
- Cha MK. Semantic network analysis of "arts management" in newspaper articles: from 1990 to 2014. *The Journal of Cultural Policy*, 29: 168-200, 2015.
- George G, Haas MR, Pentland A. Big data and management. *Academy of Management Journal*, 57: 321-326, 2014
- Ghose A, Ipeirotis PG. Designing novel review ranking systems: predicting the usefulness and impact of reviews. In Proceedings of the ninth International Conference on Electronic Commerce (ICEC'07). Association for Computing Machinery, New York, pp303-310, 2007.
- Graves A, Mohamed AR, Hinton G. Speech recognition with deep recurrent neural networks. 2013 IEEE

- International Conference on Acoustics, Speech and Signal Processing(ICASSP), Vancouver, pp6645-6649, 2013.
- Gray ER, Balmer JMT. Managing corporate image and corporate reputation. *Long Range Planning*, 31: 695-702, 1998.
- Han JS, Yoon JH. Activation strategies of the 20th BIFF using social big data text mining analysis. *Journal of Tourism Science*, 40: 133-145, 2016.
- Han KJ. The meaning and research agenda in network analysis as social science methodology: based on semantic network analysis. *Research in Social Studies Education*, 10: 219-235, 2003.
- Hinton G, Deng L, Yu D, Dahl GE. Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups. *IEEE Signal Processing Magazine*, 29: 82-97, 2012.
- Hyun IG, Park SJ. 20s~30s Chinese women's cosmetic decision: importance performance analysis on shanghai consumer`s cosmetics attribute selection. *Journal of Distribution and Management Research*, 5: 79-87, 2012.
- Kim DH, Jung YJ, Joh WI. A study on the characteristics of consumers' choice of cosmetics in online shops using AHP. *Korea Business Education Review*, 33: 387-406, 2018.
- Kim JY, Na HS, Park KH. Topic modeling of profit adjustment research trend in Korean accounting. *Journal of Digital Convergence*, 19: 125-139, 2021.
- Kim KJ, Hwang SJ. The influence of price sensitivity, bundle discount type and price level of male cosmetics on quality perception. *Journal of the Korean Society of Costume*, 66: 1-14. 2016.
- Kim KR, Kim SS. Trend analysis of research on inclusive education in Korea using text mining. *Korean Journal of Physical, Multiple, & Health Disabilities*, 59: 133-157, 2016.
- Kim MJ, Kim NS. A study on the brand value and corporate social responsibility activities that influence the company reputation. *Korean Review of Corporation Management*, 9: 279-304, 2018.
- Kim YY, Song M. A study on analyzing sentiments on movie reviews by multi-Level sentiment classifier. *Journal of Intelligence and Information Systems*, 22: 71-89, 2016.
- Koporcic N, Halinen A. Interactive network branding: Creating corporate identity and reputation through interpersonal interaction. *IMP Journal*, 12: 392-408, 2018.
- Kostyra DS, Reiner J, Natter M, Klapper D. Decomposing the effects of online customer reviews on brand, price, and product attributes. *International Journal of Research in Marketing*, 33: 11-26, 2016.
- Lai S, Xu L, Liu K, Zhao J. Recurrent convolutional neural networks for text classification. In Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence (AAI'15). AAI Press, pp2267-2273, 2015.
- Lee JH, Gil WY. News agenda classification and media diversity analysis using topic-modeling: based on news on the presidential new year press conference. *Korean Journal of Broadcasting and Telecommunication Studies*, 33: 161-196, 2019.
- Lee SA. Consumption mode of body image as sign value in mass-culture: focused on idol groups in Korea. *The Korean Journal of Cultural Sociology*, 11: 193-235, 2011.
- Lee SS. Libraries and big data: libraries and big data analytics. *Korean Library Association Journal*, 55: 14-25, 2014.
- LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature*, 521: 436-444, 2015.
- Lee SI, Park YS, Du LJ. A comparative analysis on the choice attributes of cosmetics between Korean and Chinese university students using the IPA technique. *Global Business Administration Review*, 12: 43-75, 2015.
- Lee SK. A review of big data analysis based on marketing perspective. *Korea Journal of Business Administration*, 28: 21-35, 2015.
- Liu B. Sentiment analysis and opinion mining. *Synthesis Lectures on Human Language Technologies*, 5: 1-167, 2012.
- Morinaga S, Yamanishi K, Tateishi K, Fukushima T. Mining product reputations on the web. In Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining (KDD'02), New York, pp341-349. 2002.
- Namgung JE, Kim BR, Joh WI. A study on the attributes of purchasing cosmetics in road shop using AHP. *Korean Business Education Review*, 32: 65-84, 2017.
- Oh YJ, Chae SH. Movie rating inference by construction of movie sentiment sentence using movie comments and ratings. *Journal of Internet Computing and Services*, 16: 41-048, 2015.
- Oh YK. Effects of cosmetics purchase decision factors of

- female university students in Korea and Taiwan on the Korea cosmetics satisfaction. *Asian Journal of Beauty and Cosmetology*, 18: 619-628, 2020.
- Padberg DI, Walker FE, Kepner KW. Measuring consumer brand preference. *Journal of Farm Economics*, 56: 38-47, 1974.
- Schmidhuber J. Deep learning in neural networks: an overview. *Neural Networks*, 61: 85-117, 2015.
- Song MS. A study on the repositioning strategy for competitiveness of enterprise in low-cost cosmetics market. *Asian Journal of Beauty and Cosmetology*, 16: 405-416, 2018.
- Taigman Y, Yang M, Ranzato M, Wolf L. Deepface: Closing the gap to human-level performance in face verification. 2014 IEEE Conference on Computer Vision and Pattern Recognition, IEEE, pp1701-1708, 2014.
- Tong S, Koller D. Support vector machine active learning with applications to text classification. *Journal of Machine learning research*, 2001: 45-66, 2001.
- Walsh G, Dinnie K, Wiedmann KP. How do corporate reputation and customer satisfaction impact customer defection? A study of private energy customers in Germany. *Journal of Services Marketing*, 20: 412-420, 2006.
- Weiss AM, Anderson E, MacInnis DJ. Reputation management as a motivation for sales structure decisions. *Journal of Marketing*, 63: 74-89, 1999.
- Zhu D, Lappas T, Zhang J. Unsupervised tip-mining from customer reviews. *Decision Support Systems*, 107: 116-124, 2018.

국문초록

화장품 소비자는 어떤 기호가치로 소비를 하는가?: 텍스트 마이닝을 이용한 화장품 브랜드 평판과 브랜드 선택속성 분석

송만석

경남도립남해대학 금융회계사무과, 경남, 한국

목적: 본 연구는 화장품 소비자가 소셜미디어에 작성한 화장품 브랜드 평판과 선택 속성에 대한 담론적 리뷰로 남긴 주관적 감성 빅데이터를 크롤링하여 소비자 자신이 구매한 화장품 기호가치 소비를 어떻게 인식하고 있는가에 대해 텍스트 마이닝 분석 기법인 워드클라우드 분석, 의미연결망 분석, 그리고 감성 분석을 실시하였다. **방법:** 빅데이터의 수집과 분석 도구로는 R version 3.6.2-RStudio Version 1.4.1103 프로그램을 이용하여 화장품 소비자가 구매후 남긴 기호가치 경험 리뷰 데이터를 수집하였으며, 수집된 데이터에 대해 데이터의 전처리와 불용어의 설정, 모델링의 실행, 결과 해석까지 총 3단계에 걸쳐 연구를 진행하였다. **결과:** 워드클라우드 분석 결과 '피부'라는 단어가 가장 높게 나타났으며 '구입', '자연'이 다음으로 비중이 높게 나타났다. 이들 주제가 두드러지게 나타난 것은 화장품 이용 소비자가 기호가치에 의해 구매한 화장품 선택 요인의 경험 가치를 잘 반영하고 있다 할 수 있다. 의미연결망 분석 결과에서 '자연', '진짜', '제품', '커버', '사용', 그리고 '스킨 화장품' 등의 경우 연결중심성과 매개중심성, 그리고 근접중심성이 높게 나타났으며, 감성 분석 결과 긍정적 반응을 보인 것은 4,742 단어이며, 부정적인 반응 결과 단어는 2,039, 그리고 중립으로 나타난 단어는 14,146으로 나타났다. **결론:** 워드클라우드 분석에서 상위 30위의 주제어들은 화장품 이용 소비자의 구매 목적과 선택 속성을 잘 나타내고 있는 것으로 보이며 화장품 소비를 위한 요소에서 매우 중요하다는 것을 의미하고 있다. 의미연결망 분석에서 매개중심성, 연결중심성, 근접중심성 등에서 모두 높게 나타난 '진짜', '제품', 그리고 '스킨 화장품' 등은 일 특정 화장품 브랜드에 대한 의미의 확산과 연결에 중요한 역할을 한다. 감성 분석 결과 긍정적인 감성 반응이 부정적인 감성 반응보다 약 2.33배 높게 나타난 것은 연구 대상인 일 특정 화장품 브랜드 기업의 브랜드 평판이 좋은 것으로 판단된다.

핵심어: 기호가치, 텍스트 마이닝, 워드클라우드 분석, 연결의미망 분석, 감성 분석

This work was supported by the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea(NRF-2019S1A5B5A07093402).

참고문헌

- 김근정, 황선진. 가격민감도와 번들할인 유형, 남성화장품의 가격수준이 품질지각에 미치는 영향. *복식*, 66: 1-14, 2016.
- 김기용, 김삼섭. 텍스트마이닝을 이용한 국내 통합교육 연구 동향 분석. *지체. 중북. 건강장애연구*, 59: 133-157, 2016.
- 김동현, 정윤진, 조원일. AHP를 활용한 온라인 화장품 물에서 제품구매 시 선택속성에 관한 연구. *경영교육연구*, 33: 387-406, 2018.
- 김문준, 김노사. 기업평판에 영향을 미치는 브랜드 자산과 기업의 사회적 책임활동에 관한 연구. *기업경영리뷰*, 9: 279-304, 2018.
- 김유영, 송민. 영화 리뷰 감성분석을 위한 텍스트 마이닝 기반 감성 분류기 구축. *지능정보연구*, 22: 71-89, 2016.
- 김지연, 나홍석, 박경환. 텍스트 마이닝을 이용한 이익조정 연구동향 토픽모델링. *디지털복합연구*, 19: 125-139, 2021.
- 남궁지은, 김보람, 조원일. AHP를 활용한 로드샵 화장품 구매 선택속성에 관한 연구: 한·중 소비자 비교 중심으로. *경영교육연구*, 32: 65-84, 2017.
- 배정환, 한남기, 송민. 토픽 모델링을 이용한 트위터 이슈 트래킹 시스템. *지능정보연구*, 20: 109-122, 2014.

- 송만석. 저가화장품 시장에서의 기업 경쟁력 강화를 위한 리포지셔닝 전략에 관한 연구. *아시아뷰티화장품학술지*, 16: 405-416, 2018.
- 오연주, 채수환. 영화평가 평점을 이용한 감성문장구축을 통한 영화 평점 추론. *인터넷정보학회논문지*, 16: 41-048, 2015.
- 오윤경. 한국과 대만 여대생의 화장품 구매결정요인이 한국화장품의 만족도에 미치는 영향. *아시아뷰티화장품학술지*, 18: 619-628, 2020.
- 이서구. 빅데이터 분석에 관한 마케팅적 접근. *대한경영학회지*, 28: 21-35, 2015.
- 이수안. 대중문화에서 기호가치로서 몸 이미지의 소비양식: 아이돌 그룹을 중심으로. *문화와 사회*, 11: 193-235, 2011.
- 이승인, 박윤서, 두려교. IPA를 활용한 한국과 중국 대학생의 화장품 선택속성 비교 연구. *글로벌경영학회지*, 12: 43-76, 2015.
- 이수상. 도서관과 빅데이터: 도서관과 빅데이터 분석. *도서관문화*, 55: 14-25, 2014.
- 이종혁, 길우영. 토픽모델링을 이용한 뉴스 의제 분류와 미디어 다양성 분석: 대통령 신년 기자회견 관련 뉴스 분석을 통해. *한국방송학보*, 33: 161-196, 2019.
- 차민경. 국내 언론에 나타난 '예술경영'관련 이슈의 의미연결망 분석: 1990년부터 2014년까지 국내 일간지 기사 분석을 중심으로. *문화정책논총*, 29: 168-200, 2015.
- 한관중. 사회과학 방법론으로서의 연결망 분석기법 적용의 의의와 연구과제: 의미와 연결망 분석(semantic network analysis)을 중심으로. *사회과학연구*, 10: 219-235, 2003.
- 한진성, 윤지환. 소셜 빅데이터 텍스트 마이닝을 활용한 부산국제영화제(BIFF) 활성화 방안 연구: 영화제 개최 전·후 비교 분석을 중심으로. *관광학연구*, 40: 133-145, 2016.
- 현인규, 박소진. 중국 20~30대 여성소비자들의 화장품 선택요인 연구. *유통경영학회지*, 15: 79-87, 2012.

中文摘要

化妆品消费者消费哪些嗜好价值? : 利用文本挖掘分析化妆品品牌声誉和品牌选择属性

宋晚碩

庆南南海大学金融会计事务科, 庆南, 韩国

目的: 本研究旨在通过文本挖掘分析技术-词云分析、语义网络分析和敏感性分析并检验消费者对所购买化妆品的嗜好价值的消费, 实证分析消费者如何感知他们购买的化妆品的嗜好价值。消费者在社交媒体上留下了关于化妆品品牌声誉和选择属性的话语评论。**方法:** 使用R版本3.6.2-RStudio 1.4.1103版本对化妆品消费者购买后留下的嗜好价值体验的评论数据进行收集和分析。研究分三个阶段进行: 收集数据的预处理和停用词的设置、建模的执行和结果分析。**结果:** 词云分析表明, “皮肤”一词出现次数最多, 其次是“购买”和“自然”。此类关键字突出显示的事实表明, 化妆品消费者根据嗜好价值描述了他们购买的化妆品的体验。语义网络分析表明, “自然”、“产品”、“真实”、“封面”、“使用”和“皮肤化妆品”具有高度的度中心性、中介中心性和接近中心性。最后, 通过敏感性分析, 找到了 4,742 个表示正面评价的词, 2,039 个表示负面评价的词, 以及 14,146 个表示中性评价的词。**结论:** 词云分析中, 前30个关键词很好地代表了化妆品消费者的购买目标和选择属性, 对化妆品相关消费因素具有重要意义。“真实”、“产品”和“皮肤化妆品”在语义网络分析中的中介中心性、度中心性和接近中心性都很高, 在“意义”向特定化妆品品牌的传播和连接方面发挥着重要作用以一天为周期。最后, 敏感性分析发现, 正面情绪评论的出现频率大约是负面评论的 2.33 倍。

关键词: 嗜好价值, 文本挖掘, 词云分析, 语义网络分析, 敏感性分析