ISSN 2466-2046 (Print)
ISSN 2466-2054 (Online) Asian J Beauty Cosmetol 2022; 20(4): 461-479 http://dx.doi.org/10.20402/ajbc.2022.0072

RESEARCH ARTICLE Open Access

Discrete Emotions Response after Purchase Cosmetics of Millennial Consumers: Evaluation of Satisfaction/Normal/Dissatisfaction Attributes Using Text Mining Techniques

Man Seok Song[1*], Yun-Jae Cho[2], Mi Ju Yim[3]
[1]Byuckkang Liberal Arts College, Kyungwoon University, Gumi-si, Gyeongsangbuk-do, Korea
[2]School of Start-up & Assets Convergence (Marketing), Changwon National University, Changwon-si, Gyeongsangnam-do, Korea
[3]College of Business Administration, Dong-A University, Busan, Korea

[*]**Corresponding author:** Man Seok Song , Byuckkang Liberal Arts College, Kyungwoon University, 70 Gangdong-ro, Sandong-eup, Gumi-si, Gyeongsangbuk-do 39160, Korea
Tel.: +82 54 479 1274
Fax: +82 54 479 1029
Email: consulting89@hanafos.com

Received September 25, 2022
Revised September 25, 2022
Accepted November 23, 2022
Published December 30, 2022

Abstract

Purpose: This study aimed to analyze the similarities and differences between each area using wordcloud analysis and semantic network analysis, which are text mining techniques. Further, confirmatory factor analysis will be conducted by crawling for word-of-mouth information on attribute reviews as satisfied, normal, or dissatisfied after purchases that are subjectively given by millennial generations. **Methods:** The R program version 4.1.2 was used as a big data collection and analysis tool, and text mining analysis was performed through preprocessing and stopword processing on the collected data. Further, using LISREL 8.80 we conducted confirmatory factor analysis on these results. **Results:** Wordcloud analysis revealed that the terms "skin," "products," and "skin" ranked first in the evaluation area of "satisfied," "normal," and "dissatisfied," respectively. Additionally, using confirmatory factor analysis, the correlation between the three latent variables of satisfaction, normal, and dissatisfaction was differentiated. **Conclusion:** The similarities and differences between the domains obtained through wordcloud and semantic network analyses and derived by classifying individual emotional responses of millennial consumers in social media into satisfied, normal, and dissatisfied domains are considered very meaningful. The keywords derived with high centrality in the semantic network for each domain is then refined and introduced as an observation variable for confirmatory factor analysis in accordance with the purpose of the study; this is helpful in research development for causal analysis in the future.

Keywords: Millennial consumers, Text mining, Wordcloud analysis, Semantic network analysis, Confirmatory factor analysis

Introduction

화장품 구매 후 소비자의 반응을 분석한 연구는 주로 화장품 이용 소비자의 브랜드 선택 속성에 대해 지각하는 정도를 파악하여 어떠한 속성이 구매의도에 영향을 미치는가를 규명하는 연구(Cho *et al*., 2021; Park & Lee, 2015; Kim *et al*., 2021)와 소비자의 절대적 가치평가와 종속변수로서 충성도나 구매의도 또는 만족도와의 인과관계를 제시한 연구(Namgung *et al*., 2017; Kim *et al*., 2018; Park & Ku, 2009), 그리고 화장품 브랜드 속성에 대한 탐색적 연구(Lee *et al*., 2021; Park & Lee, 2015; Ahn, 2021)들이었다.

화장품 구매 선택 속성은 화장품 구매 태도에 영향을 미치며(Seo & Kim, 2009), 다른 속성과 어떻게 차별화되는가를 알 수 있도록 하고(Day, 1980), 선택을 통해 사용하고 난 후의 지각되는 만족도를 의미(Ko & Seo, 2009)하기 때문에 매우 중요하다. 그렇다면 화장품 선택 속성에 대한 평가는 어떻게 측정할 수 있을까? 이와 같은 문제점에 대해 많은 선행연구에서는 다양한 연구 방법으로 이를 탐지하고 연구하기 위해 화장품 이용 소비자의 반응 분석에 대한 탐색적 연구와 인과분석 연구에 집중하고 있는 실정이다.

Copyright ⓒ Korea Institute of Dermatological Sciences.
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 non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

그러나 이와 같은 연구는 소비자만의 공간에서 능동적이며 경험적 문화가치가 생성되고 전파되는 인터넷과 소셜미디어 플랫폼의 정보 과부하 홍수 속에서 고관여 기호가치 소비를 하는 밀레니얼 화장품 소비자행동에 대한 태도의 탐색이나 설명에는 충분히 이르지 못하고 있을 뿐만 아니라 기업의 시장경쟁우위를 위한 브랜드 자산 강화전략이 절실히 요구되는 상황에서 복합적으로 접근하여야 하는 의사결정 시스템에 전략적 시사점으로 제공하기에는 부족함이 많다(Song, 2021).

이를 위해 강력한 빅데이터 분석 프로그램 중 하나인 R 프로그램을 이용하여 담론 형식으로 남겨진 개별정서 반응인 리뷰를 크롤링하여 텍스트 마이닝 분석 방법인 워드클라우드 분석과 의미연결망 분석을 실시하여 밀레니얼 화장품 이용 소비자 관점에서의 화장품 선택 속성에 대한 평가를 먼저 살펴보고자 한다.

그리고 이들 분석을 보완하는 방법으로 도출된 중심성이 높은 주제어가 포함된 전체 코퍼스(corpus)를 텍스트 측정변수로 정제하여 관측변수로 채택하고 이를 상관분석하고 이의 결과를 이용하여 구조방정식 모델 분석의 한 방법인 확인적 요인분석을 구조방정식 모델 전용 프로그램인 LISREL 프로그램을 이용하여 분석하게 되면 각 평가영역인 잠재변수에 대해 소비자 담론으로 이루어진 리뷰 코퍼스와의 성향과 인과관계를 확인할 수 있어 화장품 기업의 시장경쟁우위 전략 수립 방향성을 위한 기초정보도 제공할 수 있을 뿐만 아니라 구조방정식 모델 각 적합 지수가 만족 수준을 수용한다면 연구모델로서의 도입 가능성뿐만 아니라 관측변수로의 역할을 위한 측정변수로서의 도입 가능성에 대한 정당성을 확보할 수 있는 학문적 접근 방법으로서의 시사점도 제공할 수 있을 것이다.

밀레니얼 화장품 이용 소비자가 소셜미디어에 남긴 화장품 기호가치 소비에 대한 구매 후 담론 형식의 개별정서 반응을 만족/보통/불만족 영역으로 분류하여 Figure 1과 Figure 2 그리고 Figure 3의 워드클라우드 분석과 Figure 4와 Figure 5 그리고 Figure 6의 의미연결망 분석을 통하여 도출된 중심성이 높은 주제어를 Figure 7과 같은 만족/보통/불만족의 세 평가영역 잠재변수에 대한 관측변수인 측정항목으로의 도입을 위해 Table 7과 같은 텍스트 문장으로 정제하여 상관분석 과정을 거친 후 Table 8과 같은 결과를 도출할 수 있었다.

상관분석으로 얻은 결과인 Table 8을 이용하여 각 영역 잠재변수에 대한 인과관계를 정량적으로 판단할 수 있도록 Figure 7의 구조방정식 모델 분석을 적용한 확인적 요인분석의 결과는 Table 9와 같이 나타났다. 이때 도출된 연구모델 모델 적합치(model fit)가 권장 수준을 만족시키면 연구가설로 채택할 수 있어 향후 인과분석을 위한 연구모델로서의 정당성 확보와 측정 도구로서의 도입이 가능하고 텍스트 마이닝을 이용하여 수행한 연구와 인과분석 연구와의 차이점을 확인할 수 있기 때문에 이를 이용한 평가영역 간의 유사점과 차이점에 대해 연구하는 것은 발전된 디지털 테크놀로지 시대에 적합한 연구 방법이 될 것이다.

이는 설문조사와 직접 면담을 통한 심층 인터뷰보다 크롤링한 텍스트 데이터를 이용하여 분석해 보는 것이 의미 있는 연구가 될 것이라고 제안한 Kim (2018)과 동일한 맥락을 가지며 학문적 접근 방법으로서뿐만 아니라 사회 기여도 측면에서도 매우 탁월한 접근 방법이라고 할 수 있으며, 주관적이며 감성적인 다양한 가치소비에 대한 개별정서 반응으로 나타난 담론 형식의 리뷰를 이용한 연구 결과는 소비자를 대상으로 선행연구에서 정형화되어 있는 측정 문항을 이용하여 분석하는 연구 결과와는 다르게 의도되지 않은 심층 분석까지 가능할 것이다.

Methods

본 연구는 밀레니얼 소비자가 화장품 구매 후 소셜미디어에 담론으로 남긴 개별정서 반응인 평점 리뷰를 만족/보통/불만족의 세 영역으로 나누어 워드클라우드 분석과 의미연결망 분석 그리고 구조방정식 모델 분석 과정을 거치면 밀레니얼 화장품 이용 소비자는 각 평가영역에서 그들이 선택한 브랜드의 속성 평가가 어떤 지각된 유사점과 차이점이 있으며, 선행연구와는 또 다른 시사점을 도출할 수 있을까?에 대한 의문으로 시작하였다.

밀레니얼 세대는 Strauss & Howe (1991)가 처음 언급한 용어로 1980년대 초반부터 2000년대 초반에 출생한 세대를 가리키며, 태어날 때부터 컴퓨터와 핸드폰을 만지며 자라나 나만의 존재 가치를 위해 소셜미디어에 주도적으로 참여하는 트렌드의 중심에 있는 집단으로 남에게 과시하기 위한 욕구가 아닌 자기표현의 욕구로 음식점에 가서 밥을 먹는 순간 사진을 찍어 바로 소셜미디어에 올려 친구들에게 자신의 존재감을 전달하는 행위(Shin & Lee, 2018)도 하는 이들은 공통적인 사회적 환경 속에서 서로 상관관계(Kim, 2019)를 가지며 소유보다는 경험과 공유를 중요시하여 인터넷과 소셜미디어라는 채널을 통해 공유한다. 이러한 밀레니얼 세대의 경험과 지식은 그들의 생각과 태도, 가치, 신념, 행동, 라이프스타일 등에 영향을 미치기도 한다.

이상의 밀레니얼 세대 정의에서 밀레니얼 소비자는 전통적인 의사결정과정과는 다른 다양하고 복잡한 의사결정과정을 거쳐서 화장품을 구매하게 되는데 그들은 자신에게 필요한 구매 정보를 직접 수집하고 준거집단으로부터 끊임없이 그 정보를 검증하는 과정에서 인터넷과 소셜미디어에서 기호가치 소비를 위한 정보의 획득도 하고 있으며 또한 자신의 의견과 경험 그리고 정보를 생성하고 공유하면서 확산하고 있다. 화장품 브랜드에 대한 모든 담론이 인터넷상이나 소셜미디어에 리뷰로 다 담겨 있다고 할 수 없으나 밀레니얼 소비자가 어떻게 화장품을 소비하고 공감하는가에 대한 연구에서 소셜미디어에 개별정서 반응으로 남겨진 담론 형태의 리뷰는 적합한 분석 대상이 될 수 있다.

본 연구는 밀레니얼 소비자를 대상으로 인터넷과 소셜미디어에서의 소비자 행동과 욕망을 연구하는 것이 연구의 주목적이 아니라 이들

http://dx.doi.org/10.20402/ajbc.2022.0072

이 구매한 화장품에 대해 감성적 담론의 형식으로 작성한 개별정서 반응 리뷰를 텍스트 마이닝 분석 기법으로 만족/보통/불만족의 세 영역으로 나누어 워드클라우드 분석과 의미연결망 분석을 실시하여 각 영역 간의 유사점과 차이점을 분석하고 도출된 중심성이 높은 주제어 텍스트 문장을 연구의 목적에 맞게 관측변수로 정제한 후 구조방정식 모델 분석의 하나인 확인적 요인분석의 과정도 거치면서 밀레니얼 화장품 소비자의 기호가치 소비에 대한 개별감성 반응 결과의 논의를 전개하여 화장품 선택 속성과 관련된 주요 주제어를 분류하고 화장품과 관련하여 나타나는 주요 평가 속성 트렌드와 이슈에 대한 미래 신호를 탐지하여 새로운 학문적 접근 방법과 시장경쟁우위의 화장품 브랜드 자산 강화를 위한 예측 모델의 패러다임으로도 제시하고자 한다.

1. 데이터의 수집과 전처리

밀레니얼 소비자는 기호가치로 소비한 화장품에 대한 경험적 가치인 주관적 감성을 비정형화된 자기표현을 담론 형식으로 인터넷과 소셜미디어에 구매한 화장품 브랜드의 평가와 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).

구매 후 고객이 담론 형식으로 인터넷과 소셜미디어에 남긴 비정형 텍스트 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 많은 연구자와 기업은 고객 리뷰를 지능적으로 분석하여 전통적인 연구 방법을 보완하고자 시도하고 있다(Song, 2021).

화장품 기호가치 소비에 대한 개별감성 반응 결과의 논의를 전개하기 위해서는 소셜미디어에 시계열 자연어 담론 형식으로 장기간에 걸쳐 작성된 방대한 감성 리뷰 상태인 사용 후 그대로의 인식, 태도, 행동 등의 비정형 빅데이터 수집이 필요하다. 이를 위해 글로우픽과 네이버 쇼핑 기초화장품 카테고리에 소비자 주도형 화장품 구매 우선순위와 솔직한 사용 후기가 담론 형태로 장기간에 걸쳐 존재하고 있는 구매 후 화장품 브랜드에 대해 기호가치 소비의 경험 가치로 평가한 평점 1점과 2점은 불만족, 3점은 보통, 그리고 4점과 5점을 만족으로 하는 세 영역으로 나누어 크롤링하여 각 영역별 워드클라우드 분석과 의미연결망 분석을 실시하고자 한다.

그리고 구조방정식 모델을 이용한 확인적 요인분석을 위해서는 각 영역별 분석에서 중심성이 높게 도출된 주제어가 포함된 문장 텍스트를 Table 7과 같은 관측변수로 정제하고 이들 관측변수 텍스트로 상관분석 과정을 거치면 Table 8과 같은 상관계수가 도출된다. 이들을 LISREL 구조방정식 전용 분석 프로그램으로 확인적 요인분석 과정을 거쳐서 Figure 7과 Table 9와 같은 분석 결과를 얻을 수 있었다. 이때 도출된 분석 결과가 신뢰도 및 타당성 검증을 위한 모델 접합도 기준을 만족시키면 연구모델로 채택할 수 있고 화장품 구매 후 만족도 평가를 위한 인과분석 측정 항목으로의 도입도 가능할 것이다.

전처리란 크롤링 된 데이터를 분석하는 작업 전에 데이터를 분석하기 좋은 형태로 만드는 과정으로 데이터 분석 과정에서 데이터 전처리가 중요한 이유는 아무리 좋은 분석 기법이나 도구를 사용하더라도 원시 담론 상태의 텍스트로 이루어진 빅데이터가 가지고 있는 한계성으로 인해 좋은 분석 결과를 얻기 어려워 수집된 데이터는 연구목적에 맞게 분석하기 적합한 형태로 정제하는 과정이 필요하다.

구체적으로 데이터 전처리 과정에선 수집된 텍스트 내에서 동일하거나 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 수집된 빅데이터를 텍스트 마이닝 라이브러리(library)를 이용해 말뭉치(corpus)로 변환한 후 한글 자연어 처리 KoNLP 라이브러리 함수를 이용하여 데이터 전처리 과정을 수행하였으며, 형태소 사전으로는 한국정보화진흥원 사전(NIADic)을 이용하였다. 그리고 품사 구분을 위해서는 KoNLP에 있는 SimplePos09()를 사용하였다.

한국어로 이뤄진 텍스트 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 연구에서 텍스트 내에서 명사만을 추출해 분석하고 있으며, 이런 관점과 더불어 특정 주제를 구성하거나 이와 연관된 의미를 내포하는 키워드는 대부분 명사 형태를 가지는 경우가 많다(Bae *et al*., 2014).

본 연구에서의 분석을 위한 형태소를 처음에는 명사와 형용사 그리고 동사를 포함하여 워드클라우드 분석과 의미연결망 분석을 하였으나 명사만으로 한정하여 실시한 분석 결과와의 비교에서 큰 차이가 없었으며, 오히려 명사와 형용사 그리고 동사를 포함한 분석 결과가 필요 이상으로 복잡하게 도출되어 선행연구(Song, 2021; Lee & Gil, 2019; Bae *et al*., 2014)와 같이 명사만으로 한정하였으며 나머지는 불용어로 처리하였다. 그리고 한 글자로 이루어진 단어 역시 모두 불용어로 처리하였으나 "향"은 화장품 특성상 주요어이며, 높은 빈도로 도출되었기 때문에 전처리 과정에서 제외하도록 처리하였다.

2. 분석 방법

텍스트 마이닝은 리뷰, 신문 기사, 인터뷰 등의 텍스트로 구성된 데이터를 기반으로 축약된 정보로 추출하거나 새로운 정보를 찾아내는 기법으로(Aggarwal & Wang, 2011; Steyvers & Griffiths, 2007), 비정형 데이터로 이루어진 텍스트를 분석하면서 연구자 주관을 최대한 배제하고 통계적 방법을 통해 의미 구조를 추론할 수 있으며, 텍스트 문서 단어 간의 동시 출현 정보로부터 계산된 토픽과 단어의 확률 분포로부터 생성된다고 간주한다(Steyvers & Griffiths, 2007).

소비자가 남긴 빅데이터를 분석하는 연구가 전폭적으로 도입되면서 다양한 학문 연구 분야에서 빅데이터 분석 방법인 텍스트 마이닝 기법을 활용한 연구가 활발하게 수행되고 있다(Zhu *et al.*, 2018; Kim & Kim, 2016). 텍스트 마이닝은 정보 검색과 정보 추출 및 자연어 처리 기술을 사용하여 알려지지 않은 유용한 패턴과 지식을 텍스트로부터 발견하는 것을 말하며(Han & Yoon, 2016), 비정형으로 이루어진 텍스트 데이터를 자연어 처리와 형태소 분석 기술로 수집어를 정제하고, 단어를 추출해 빈도수를 제시하여 순위나 인식의 유사성, 일반성을 찾아내는 데 사용하며(Park, 2020), 텍스트 내에서 메시지를 형성하는 단어들이 어떻게 사용되었는지에 대한 구조의 패턴과 의미를 분석함으로써 관계의 형태 속에 내포되어있는 메시지의 내용까지 도출하는 분석 방법이다(Cha, 2015; Han, 2003).

본 연구의 목적인 밀레니얼 화장품 이용 소비자가 기호가치 소비로 구매하여 소셜미디어에 작성한 화장품의 평가와 선택 속성에 대해 어떻게 느끼는가에 대한 개별정서 반응 리뷰의 분석을 위해 텍스트 마이닝 분석 기법인 워드클라우드 분석, 의미연결망 분석, 그리고 구조방정식 모델 분석을 실시하여 화장품 브랜드 기호가치 소비 체계에서 논의를 전개하기 위한 연구 문제의 수행을 위해 글로우픽과 네이버 쇼핑 기초화장품 카테고리에 구매 후 남긴 평점을 기준으로 1점과 2점은 불만족, 3점은 보통, 그리고 4점과 5점은 만족으로 하여 영역별로 분류하여 R version 4.1.2 (RStudio Version 2021.09.2-382.exe) 프로그램을 이용하여 크롤링하였다.

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 쉽게 R의 홈페이지(https://r-project.org)에서 다운로드 받을 수 있으며, R에서 사용할 수 있는 수많은 통계 패키지자 개발되어 있어 사용자가 제작한 패키지를 추가하여 기능을 확장할 수 있다(Song, 2021).

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 Song (2021)과 같은 RStudio의 오픈소스 에디션을 이용하였다.

수집된 데이터에 대해 데이터의 전처리와 불용어의 설정, 토픽 모델링의 실행, 결과 해석까지 총 3단계에 걸치는 과정을 통하여 R 프로그램으로 워드클라우드 분석과 의미연결망 분석을 실시하였으며, 중심성이 높게 도출된 주제어를 중심으로 연구 목적에 맞게 주제어가 포함된 담론 리뷰 텍스트 문장을 Table 7과 같은 관측변수로 정제하여 이들 텍스트 문장을 상관분석하여 Table 8과 같은 결과를 도출하였고, 이를 연구가설로의 도입을 위해 구조방정식 전용 프로그램인 LISREL 8.80 프로그램을 이용하여 Figure 7과 같은 구조방정식 모델을 적용하는 확인적 요인분석을 실시하였으며 Table 9와 같은 측정모델에 대한 확인적 요인분석의 모델 적합도를 얻게 되었다.

화장품 소비자에 의해 작성된 담론 리뷰의 영역이 너무 넓을 뿐만 아니라 방대하여 분석 대상을 기초화장품으로 한정하였으며, 노출빈도가 높은 주요한 키워드를 시각화하기 위해 워드클라우드 분석을 실시하여 Figure 1, Figure 2, 그리고 Figure 3과 같이 도출할 수 있었으며, 이의 분석 결과인 각 영역 간의 주요한 주제어 빈도 순위 30위까지에 대해서는 Table 1, Table 2, 그리고 Table 3에 제시하였다.

만족/보통/불만족의 각 영역에서 도출된 주요 주제어가 상호 어떤 관계를 맺고 그들 사이에 어떤 연결상태를 나타내는지를 알기 위해 네트워크 연결 구조를 통해 담론 텍스트에 담긴 의도적인 의미의 흐름을 파악할 수 있는 의미연결망 분석을 실시하였다. 이들 분석 결과는 Figure 4, Figure 5, 그리고 Figure 6으로 나타났으며, 각 영역별 분석된 주제어에 대한 연관관계의 파악을 위한 지표인 매개중심성과 연결중심성 그리고 근접중심성에 대한 도출된 결과의 설명은 Table 4, Table 5, 그리고 Table 6으로 나타났다.

확인적 요인분석은 Jöreskog (1969)에 의해 처음 개발되었으며, 행동과학 및 사회과학 등 다양한 학문 분야 연구에서 활용되고 있는 중요한 양적 연구방법론(Anderson & Gerbing, 1988)으로 주로 연구를 위한 측정 구조의 측정값이 해당 구성의 특성에 대한 연구원의 이해와 일치하는지 여부를 위해 사전 확인하고 검증하는 데 유용하게 사용하고 있으며, 데이터가 가설 측정 모델에 적합하는가의 여부를 본 연구에 앞서서 사전 테스트한다(Preedy & Watson, 2009).

이때 연구자는 먼저 연구를 위한 기초 요인으로 연구모델과 함께 연구가설을 수립하고 연구모델에 대한 설정을 부과하는 선험적 가설을 확인적 요인분석으로 타당성과 신뢰도를 확인하며, 이러한 가설의 사전 설정을 부과하는 과정을 통하여 연구모델이 이론과 일치하도록 하며, 그런 다음 제안된 연구모델의 분석 과정을 거치면 연구모델 적합도 측정값을 얻을 수 있다. 이때 통계 분석 결과가 구조방정식 모델 적합도에 적합하지 않은 것으로 나타나면 그 연구모델은 기각하게 된다.

본 연구의 확인적 요인분석의 적용을 위해서는 확인적 요인분석을 처음 개발하여 도입한 스웨덴 웁살라(Uppsala) 대학의 Karl G Jöreskog와 Dag Sörbom가 개발한 구조방정식 모델링 분석을 위한 최초의 프로그램인 LISREL 프로그램의 8.80 Version을 사용하였다.

사실 구조방정식 모델에서 의미하는 인과관계는 엄밀히 말하면 변수 간 상관행렬이나 공분산행렬을 이용한 상관분해를 통해 인과관계로 설정된 변수 간 관계를 분석하는 것으로 연구모델과 변수 간 상관관계가 존재한다고 가정하여 상관의 분해를 위해서 먼저 인과계수와 상관계수를 명명한 후 상관분해를 이용하여 인과관계를 계산하게 된다(Yu, 2016). 이러한 인과관계 연구는 실험법을 통해 구할 수 있는 인과관계와 다른 개념의 인과관계임이 분명하다.

측정변수 간의 상관계수를 이용하여 인과분석을 하는 것에 대해 Guttman (1977)은 상관관계는 인과관계를 의미하지 않는다며 상관

http://dx.doi.org/10.20402/ajbc.2022.0072

관계를 인과관계로 해석하려는 연구 움직임을 강하게 비판하였고, Freedman (1987)과 de Leeuw(1985) 또한 이러한 움직임에 비판하고 있으나 이러한 비판의 의견에 대해 Bagozzi & Yi (1989)는 구조방정식 모델을 이용하여 실험법을 통해 얻은 데이터를 분석할 수 있음을 보여주었을 뿐만 아니라 Bollen (1989) 역시 비실험법을 통한 인과관계의 타당성을 설명하고 있으며, 오늘날 많은 연구자 또한 설문지를 이용한 비실험법적 연구를 인과관계 연구로써의 적합성을 인정하고 있으며, 지금도 국•내외적으로 다양한 구조방정식 모델 분석 도구를 이용한 인과분석 연구논문이 발표되고 있다.

인과분석을 위한 연구모델의 측정 도구 구성개념에 대한 타당성을 확인하는 방법으로 탐색적 요인분석과 함께 많이 사용되는 확인적 요인분석은 측정변수와 잠재변수와의 관계만을 탐색하는 탐색적 요인분석과는 다르게 측정변수를 통해 잠재변수를 외현화 시킬 때 잠재변수와 관측변수 사이의 관계를 검증하는 방법이다. 즉, 잠재변수와 관측변수 사이의 인과관계 모델을 통해 관측변수들 사이의 공분산을 적절히 설명할 수 있는지를 통계적으로 나타내는 것이 확인적 요인분석이다.

확인적 요인분석은 탐색적 요인분석이 탐색적 형식으로 관측 문항이 그들의 내재된 요인들과 어느 정도 연결되어 있는가를 결정하기 위해 진행하는 것과는 다르게 연구자가 설정한 연구모형을 분석하기 전 단계에서 분석을 위한 관측 문항들이 실제 자료와 부합하는지를 확인 또는 검증하는 통계기법으로 다 항목으로 측정한 척도의 타당성을 검증하기 위해 연구에 포함되는 잠재변수와 관측변수 간의 관계, 잠재변수와 잠재변수 간의 관계 및 방향 등을 가설로 설정한 연구모형을 기초로 분석하는 사전적인 분석 방법이다(Kim & Song, 2009; Park *et al*., 2014).

본 연구와 같이 구조방정식 모델 분석을 시행하게 되면 밀레니얼 화장품 이용 소비자의 특정 화장품 브랜드에 대한 사용 후 만족도와 충성도를 알 수 있을 뿐만 아니라 특정 화장품 브랜드에 대한 충성도와는 무관하게 만족/보통/불만족 평가영역에 대한 소비자의 성향을 객관적으로 판단할 수 있으며, 각 잠재변수에 영향을 미치는 다양한 원인을 정량적으로 분석하여 시각적으로 쉽게 파악할 수 있다는 것이다.

새로운 접근법의 적용을 위해 직접 설문지를 이용한 구조방정식의 확인적 요인분석과는 다르게 크롤링한 빅데이터에서 추출된 중심성이 높은 주제어 코퍼스를 중심으로 Table 7에 제시된 것과 같이 연구모델에 대한 관측변수로의 도입을 위해 텍스트로 구성된 12개의 측정항목을 상관분석하여 Table 8과 같은 상관계수를 도출할 수 있었으며, 도출된 결과를 LISREL 8.80 프로그램으로 Figure 7과 같은 확인적 요인분석을 실시하여 Table 9와 같은 경로분석의 결과와 연구모델의 각 적합지수의 확인을 통해 연구모델의 적합도를 알 수 있었다.

Results and Discussion

본 연구에서는 소비자가 주도하여 화장품 브랜드 순위와 리뷰가 존재하고 있는 글로우픽과 네이버 쇼핑 화장품 카테고리에서 개별 정서 반응이 담론 형식으로 장기간에 걸쳐 존재하는 화장품 브랜드에 대한 기호가치 소비의 경험가치인 리뷰 빅데이터를 R 프로그램과 구조방정식 전용 프로그램인 LISREL 프로그램을 이용하여 도출된 결과와의 유사성을 분석하고 직관적으로 파악하기 위해 시각화 작업을 진행하였다.

워드클라우드 분석의 결과로 나타나는 Figure 1과 Table 1은 밀레니얼 소비자가 화장품 구매 후 소셜미디어에 담론으로 남긴 개별정서 반응인 리뷰 평점을 4점과 5점으로 평가한 것으로 만족의 평가영역

Table 1. Satisfaction-Top 30 keywords and frequency in wordcloud analysis

NO	Keywords	Frequency	Cumulative frequency	NO	Keywords	Frequency	Cumulative frequency
1	Skin	3,395	3,395	16	Really	773	25,480
2	Product	2,898	6,293	17	Price	770	26,250
3	Use	2,649	8,942	18	Think	749	26,999
4	Cream	2,219	11,161	19	Face	746	27,745
5	Feel	1,851	13,012	20	Effect	719	28,464
6	Fragrance	1,845	14,857	21	Summer	693	29,157
7	Moisture	1,643	16,500	22	Recommendation	690	29,847
8	Toner	1,223	17,723	23	Truth	666	30,513
9	Absorption	1,181	18,904	24	Winter	623	31,136
10	Moisturizing	1,152	20,056	25	About	575	31,711
11	Skincosmetics	1,034	21,090	26	Ingredient	575	32,286
12	Lotion	1,007	22,097	27	Soothing	557	32,843
13	Trouble	881	22,978	28	Dry	516	33,359
14	Ease	875	23,853	28	Makeup	495	33,854
15	Hydrating	854	24,707	30	Purchase	495	34,349

Figure 1. Results of wordcloud analysis of the satisfaction domain.
Wordcloud analysis of the satisfaction domain revealed that the top keywords represent the purchasing goals and selection attributes of cosmetic consumers and are very important for cosmetic consumption factors. The word “Skin” appeared the most, followed by “Product” and “Use.” Such prominently featured keywords suggest that millennial cosmetic consumers associate their experience with the selection attribute of cosmetics that they purchased based on semiotic value.

을 나타내며, Figure 2와 Table 2는 평점 3점으로 평가한 것으로 보통의 평가영역을 나타내며, Figure 3과 Table 3은 평점 1점과 2점으로 평가한 것으로 불만족의 평가영역을 나타낸다.

그리고 의미연결망 분석 또한 Figure 4는 만족, Figure 5는 보통, 그리고 Figure 6은 불만족의 영역으로 분류한 분석 결과를 나타내고 있으며, 밀레니얼 소비자의 평점을 기준으로 분석한 결과에 대한 설명은 Table 4, Table 5, 그리고 Table 6으로 제시하였다.

확인적 요인분석을 위해 도입한 잠재변수와 텍스트로 이루어진 이들 12개의 관측변수는 Table 7과 같으며, 이들에 대한 상관분석 결과는 Table 8로 나타났고 확인적 요인분석을 위한 연구모델은 Figure 7과 같으며, 그의 분석 결과는 Table 9의 리스트로 제시하였다.

1. 워드클라우드 분석 결과

워드클라우드와 관련한 빈도 분석에서는 밀레니얼 소비자가 만족/보통/불만족 세 영역에서 담론 형식으로 남긴 리뷰를 영역별로 전처리한 후 워드클라우드 분석을 실행하여 Figure 1의 만족 워드클라우드 분석 결과와 Figure 2의 보통 워드클라우드 분석 결과 그리고 Figure 3의 불만족 워드클라우드 분석 결과와 같이 시각적으로 확인할 수 있었으며, 토픽 모델링의 대상이 되는 수집된 전체 담론에 대해 출현 빈도가 높은 키워드 상위 30위까지의 출현 주제어는 영역별로 Table 1과 Table 2 그리고 Table 3과 같이 나타났으며, 빈도순으로 정렬하여 이를 기반으로 표기하였다.

영역별 상위 5위까지의 주제어 누적 빈도는 Table에 제시된 바와 같이 만족 영역의 경우에는 약 37.9%를 차지하고 있었으며, 보통 영역의 경우에는 34.9%를 차지하고 있는 것으로 나타났고, 불만족 영역의 경우에서는 40.9%를 차지하고 있는 것으로 나타났다.

도출된 워드클라우드에서 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 연구에서는 Figure 1인 만족

Table 2. Normal-Top 30 keywords and frequency in wordcloud analysis

NO	Keywords	Frequency	Cumulative frequency	NO	Keywords	Frequency	Cumulative frequency
1	Product	1,620	1,620	16	Price	549	15,723
2	Skin	1,560	3,180	17	About	537	16,260
3	Use	1,488	4,668	18	Summer	489	16,749
4	Feel	1,413	6,081	19	Think	480	17,229
5	Cream	1,236	7,317	20	Face	471	17,700
6	Ease	1,212	8,529	21	Toner	438	18,138
7	Fragrance	1,116	9,645	22	Ingredient	426	18,564
8	Moisture	837	10,482	23	Ample	366	18,930
9	Effect	750	11,232	24	Purchase	324	19,254
10	Skincosmetics	747	11,979	25	Truth	312	19,566
11	Lotion	684	12,663	26	Blackhead	306	19,872
12	Absorption	678	13,341	27	Winter	291	20,163
13	Trouble	633	13,974	28	Soothing	288	20,451
14	Moisturizing	603	14,577	28	Smell	258	20,709
15	Hydrating	597	15,174	30	Makeup	243	20,952

http://dx.doi.org/10.20402/ajbc.2022.0072

Figure 2. Results of wordcloud analysis of the normal domain.
Wordcloud analysis of the normal domain revealed that the top keywords represent the purchasing goals and selection attributes of cosmetic consumers and are very important for cosmetic consumption factors. The word “Product” appeared the most, followed by “Skin” and “Use.” Such prominently featured keywords suggest that millennial cosmetic consumers reflect well their experience value with the selection attribute of cosmetics that they purchased based on semiotic value.

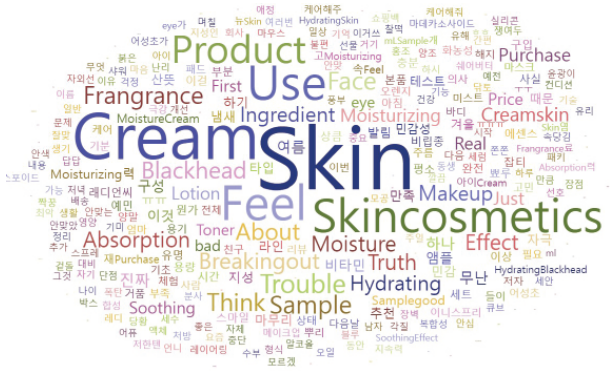

Figure 3. Results of wordcloud analysis of the dissatisfaction domain.
Wordcloud analysis of the moderate domain revealed that the top keywords represent the purchasing goals and selection attributes of cosmetic consumers and are very important for cosmetic consumption factors. The word “Skin” appeared the most, followed by “Cream” and “Use.” Such prominently featured keywords suggest that millennial cosmetic consumers reflect well their experience value with the selection attribute of cosmetics that they purchased based on semiotic value.

영역의 경우에는 '피부'가 월등하게 높은 빈도를 보이고 있으며, 그다음으로 '제품', '사용', '크림', '느낌', '향' 등의 주제어 순으로 도출되었다.

Figure 2인 보통 영역에 대한 상위 주제어를 살펴보면 '제품', '피부', '사용', '느낌', '크림' 등의 주제어 순으로 도출되었으며, Figure 3의 불만족 영역의 경우에는 '피부', '크림', '사용', '스킨', '제품' 등의 주제어 순으로 도출되었다. 이들 주제어가 두드러지게 나타난 것은 화장품 이용 소비자의 기호가치에 의해 구매한 화장품 선택 요인의 경험 가치를 잘 반영하고 있다 할 수 있다.

만족/보통/불만족의 세 영역에서 상위로 도출된 주제어의 우선순위가 유사하거나 동일한 순위로 있는 것은 밀레니얼 화장품 이용 소비자들이 그들이 사용하는 브랜드에 대한 선택 속성 우선순위가 높은 주제어를 의미하는 것으로 볼 수 있으나 담론 형식으로 남겨진 개인별 리뷰의 행간을 분석하면 그 내용은 사뭇 다르다.

전체 주제어 중 만족과 불만족 영역에서 1위이며, 보통 영역에서 2

Table 3. Dissatisfaction-Top 30 keywords and frequency in wordcloud analysis

NO	Keywords	Frequency	Cumulative frequency	NO	Keywords	Frequency	Cumulative frequency
1	Skin	957	957	16	Breakingout	171	5,952
2	Cream	603	1,560	17	Think	161	6,113
3	Use	558	2,118	18	Lotion	158	6,271
4	Skincosmetics	551	2,669	19	Just	140	6,411
5	Product	528	3,197	20	Soothing	140	6,551
6	Feel	315	3,512	21	Notgood	138	6,689
7	Fragrance	288	3,800	22	Truth	137	6,826
8	Blackhead	282	4,082	23	Purchase	135	6,961
9	Trouble	269	4,351	24	Toner	134	7,095
10	Moisturizing	264	4,615	25	About	129	7,224
11	Effect	254	4,869	26	Sample	126	7,350
12	Face	246	5,115	27	Price	125	7,475
13	Ingredient	236	5,351	28	Hydrating	123	7,598
14	Moisture	215	5,566	29	Irritate	116	7,714
15	Absorption	215	5,781	30	Makeup	104	7,818

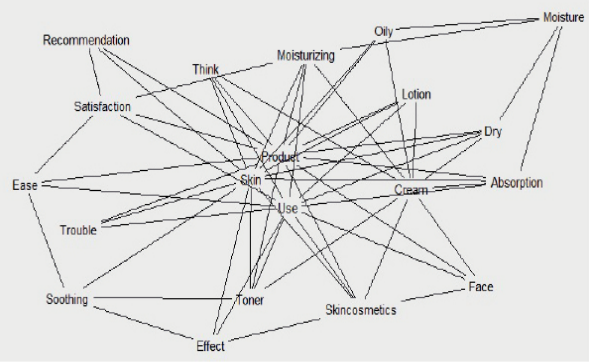

Figure 4. Semantic network analysis of the satisfaction domain. Semantic network analysis revealed that the words “Use,” “Face,” “Skincosmetics,” “Skin,” and “Effect” appeared high in the degree centrality, betweenness centrality, and closeness centrality and played an important role in spreading and connecting “meaning” in the semantic network analysis.

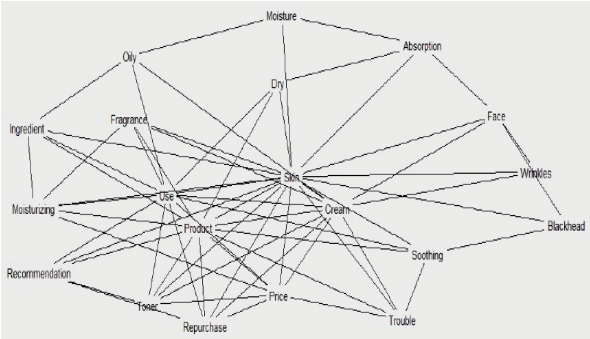

Figure 5. Semantic network analysis of the normal domain. Semantic network analysis revealed that “Skin,” “Absorption,” “Use,” “Face,” and “Moisture” appeared high in the degree centrality, betweenness centrality, and closeness centrality and played an important role in spreading and connecting “meaning” in the semantic network analysis.

위로 나타난 '피부'의 경우 각 영역에 남겨진 담론을 예시해 보면, 만족 영역에서는 "바르고 나면 피부 당김도 덜한 편이었고 자극 같은 것도 올라오지 않아서 좋습니다.","세안하고 바로 발라주면 피부가 오래도록 촉촉해요.", 그리고 "피부 진정 세럼으로 강력 추천합니다!" 등으로 나타났으며, 보통 영역에서는 "크게 피부가 좋아지는 개선점을 느끼지 못해서 아쉬움", "20대 초반의 진정이 필요한 피부에 좋을 거 같아요, 중건성에게는 조금 부족한 느낌이었어요." 그리고 "그냥 촉촉하고 그렇지만 피부를 드라마틱하게 변화시켜주진 못해요." 등으로 나타났으며, 불만족의 영역에서는 같은 주제어인 '피부'이지만, "유명해서 사봤는데 피부에 안 맞네요.", "둔감성(?) 피부인데도 이것만 쓰면 좁쌀 폭탄이 있습니다." 그리고 "엄마꺼 몰래 썼다가 피부과에서 돈 왕창 깨졌다." 등으로 나타났다. '피부'라는 주제어 외에도 다

Table 4. Satisfaction–Results of semantic network analysis

NO	Keywords	Betweenness centrality	Degree centrality	Closeness centrality
1	Satisfaction	9.730	5	0.017
2	Moisturizing	6.730	5	0.010
3	Skincosmetics	53.000	5	0.009
4	Absorption	15.700	5	0.009
5	Dry	31.400	4	0.008
6	Lotion	1.000	4	0.010
7	Think	23.000	4	0.008
8	Oily	21.000	4	0.008
9	Soothing	17.400	4	0.009
10	Toner	8.970	4	0.008
11	Use	114.000	3	0.009
12	Moisture	30.600	3	0.008
13	Face	65.700	3	0.009
14	Recommendation	1.000	3	0.014
15	Trouble	15.300	3	0.008
16	Effect	34.000	3	0.009
17	Ease	2.290	2	0.008
18	Skin	43.300	2	0.008
19	Product	27.000	1	0.008
20	Cream	18.400	1	0.007

http://dx.doi.org/10.20402/ajbc.2022.0072

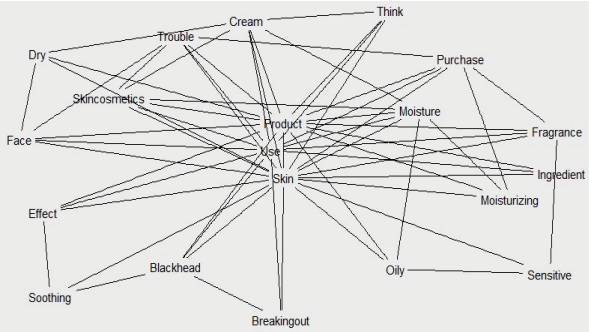

Figure 6. Semantic network analysis of the dissatisfaction domain.
Semantic network analysis revealed that the words "Product," "Cream," "Skin," "Skincosmetics," and "Trouble" appeared high in the degree centrality, betweenness centrality, and closeness centrality and played an important role in spreading and connecting "meaning" in the semantic network analysis.

른 주제어의 경우에도 마찬가지로 영역별로 다른 차별화된 담론의 리뷰를 남기고 있었다.

2. 의미연결망 분석 결과

의미연결망 분석은 네트워크 정보학의 한 영역으로 의미를 갖는 텍스트를 분석 대상으로 하여 언어로 된 메시지에서 의미를 가지고 단어를 추출하고 그들 간에 구성되는 의미적 연결 관계를 토대로 네트워크를 구성하며 네트워크의 분석지표를 활용하여 언어 메시지의 다양한 의미적 특성을 분석하는 방법을 말하는 것으로 사회 연결망 분석을 확장한 개념이며 단어의 출현 빈도와 관계를 확인하여 메시지에 내재하여 있는 의미 혹은 개념적인 특성을 추출하고 이들 간의 관계적 속성을 파악하는 데 유용한 분석 방법이다(Lee, 2014).

본 연구에서는 단어 사이의 연관 관계를 확인하기 위한 지표로 가장 많이 사용되는 매개중심성과 연결중심성 그리고 근접중심성 등을 사용하기로 한다. 의미연결망을 통하여 도출되는 노드(node)가 그래프의 중앙에 위치할수록 전체 연결망에서 중심적인 역할을 하는데 매개중심성은 연결망 내에서 한 단어가 담당하는 중개자 역할의 정도를 뜻하며 매개중심성이 높을수록 연결망에서 의미의 확산에 중요한 역할을 한다. 연결중심성은 하나의 단어(노드)에 직접 연결된 단어들의 합으로 연결망에서 중요한 역할을 하는 단어로 파악할 수 있으며, 근접중심성은 한 단어에서 다른 단어에 도달하기까지 필요한 최소 단계수로 각 단어 간 거리를 근거로 중심성을 측정한다.

많은 기업은 과거와는 다르게 소셜미디어에 텍스트 형태로 저장된 소비자가 전달하고자 하는 지각된 기업의 브랜드 아이덴티티와 이미지에 대한 평가의 측정을 위해 텍스트 마이닝 분석 기법의 적용을 시도하고 있다. 텍스트 마이닝 분석 기법의 하나인 의미연결망 분석에서 중심성은 권력과 영향력이라는 개념과 연결되어 가장 많이 쓰이는 지표 가운데 하나로서 중심성이 높은 개인은 특별한 사회 경제적

Table 5. Normal–Results of semantic network analysis

NO	Keywords	Betweenness centrality	Degree centrality	Closeness centrality
1	Price	18.200	7	0.022
2	Use	93.400	6	0.021
3	Repurchase	18.800	5	0.021
4	Soothing	23.000	5	0.023
5	Recommendation	1.000	5	0.021
6	Moisturizing	1.420	4	0.020
7	Ingredient	6.670	4	0.021
8	Face	60.300	4	0.017
9	Product	26.200	4	0.018
10	Toner	0.583	4	0.020
11	Fragrance	12.600	4	0.020
12	Dry	31.300	3	0.020
13	Blackhead	34.000	3	0.020
14	Wrinkles	1.000	3	0.015
15	Oily	28.000	3	0.018
16	Trouble	0.333	3	0.016
17	Absorption	152.000	3	0.022
18	Moisture	35.300	2	0.018
19	Cream	2.670	1	0.013
20	Skin	152.000	1	0.016

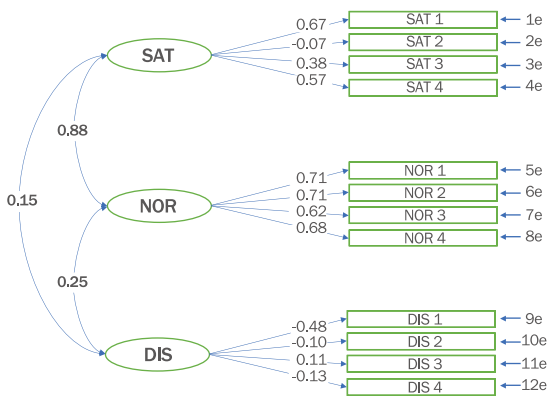

Figure 7. Research model for confirmatory factor analysis.
This is a confirmatory factor analysis using the results of correlation analysis by text mining. Confirmatory factor analysis satisfied all the fitness of the research model. The satisfaction and normal evaluation areas exhibit a high correlation between latent variables, so differentiation is low. However, they show a low correlation with the area of dissatisfaction evaluation, thereby showing high differentiation.

지위를 갖는 사람이고 조직의 경우에서도 중심성이 높을수록 생존율이 높거나 기업 성과가 좋다. 그렇기 때문에 중심성은 통계 분석에서 훌륭한 독립변수나 종속변수가 된다(An, 2009).

데이터 전처리 과정을 통하여 추출된 기호가치 화장품 이용 소비자의 구매 후 담론 형식으로 남겨진 고객 리뷰를 워드클라우드 분석과 함께 의미연결망 분석을 도입하여 각 노드에 대한 연결 관계를 시각화하였다. 이 과정 역시 워드클라우드 분석과 마찬가지로 명사로만 한정하여 단어 추출과 트랜잭션(transaction)을 생성하였으며, 의미있는 의미연결망 구조와 중심성을 파악하기 위해 전처리 과정을 거치면서 Figure 4와 Figure 5 그리고 Figure 6과 같이 시각화하였고 Table 4와 Table 5 그리고 Table 6과 같은 의미연결망 분석 결과를 도출하였다.

의미연결망 분석한 결과로 노드와 선으로 이루어진 의미연결망이 도출되었다. 텍스트 내에서 노드와 노드를 연결하는 선은 노드 간의 연관 정도를 나타내고 있으며, 핵심 주제어 간 연관성이 높으면 노드와 노드를 이어주는 선이 많아지고 연관성이 없으면 선은 연결되지 않는다. 따라서 도출된 빈도가 높을수록 노드에 대한 연결망이 복잡해지고 상대적으로 출현 빈도가 낮으면 노드와 연결되는 선의 연결망 수는 적어진다(Song, 2021). 결과로 나타난 중심성 중에서 한 노드에 연결된 엣지 개수가 많다고 무조건 중요한 노드는 아니지만, 연결망에서 중요한 역할을 하는 단어로 파악한다.

의미연결망 분석에서는 근본적으로 네트워크에서 노드 간의 연관관계에 따른 노드의 역할이 네트워크에 어떤 역할을 담당한다거나 어떤 영향을 미치는 것을 파악하는 것이 중요하다. 일반적으로 네트워크의 중심을 분석하기 위하여 중심성 분석(centrality analysis)을 실시하며(Sabidussi, 1966), 이 분석은 노드가 중심에 어느 정도 근접하

Table 6. Dissatisfaction–Results of semantic network analysis

NO	Keywords	Betweenness centrality	Degree centrality	Closeness centrality
1	Product	86.100	9	0.028
2	Cream	11.000	7	0.024
3	Skin	88.800	7	0.027
4	Purchase	51.600	6	0.029
5	Use	38.600	6	0.025
6	Moisture	20.500	6	0.023
7	Skincosmetics	4.380	6	0.026
8	Blackhead	14.400	6	0.023
9	Trouble	70.800	6	0.028
10	Face	2.640	5	0.024
11	Dry	1.680	4	0.022
12	Think	0.140	4	0.022
13	Fragrance	37.900	4	0.027
14	Effect	9.080	4	0.022
15	Sensitive	18.000	3	0.023
16	Moisturizing	0.330	3	0.022
17	Ingredient	0.710	3	0.021
18	Oily	2.620	3	0.024
19	Breakingout	2.370	3	0.020
20	Soothing	1.330	3	0.020

http://dx.doi.org/10.20402/ajbc.2022.0072

는가를 표현하는 방법이다. 본 연구를 위한 의미연결망 분석에서는 중심성 분석을 통해 의미연결망 분석의 중요성을 분석하기로 한다.

화장품을 기호가치 소비한 소비자에게 평점 4점과 5점을 받은 만족 영역의 의미연결망 분석 결과인 Figure 4와 Table 4를 살펴보면 매개중심성이 가장 높은 키워드는 '사용', '얼굴', '스킨', '피부', '효과' 등의 순으로 나타났고 이들은 전체 연결망 내에서 담당하는 중개자 역할에서 높은 위치를 나타내고 있음을 뜻하며 연결망에서 의미의 확산에 중요한 역할을 한다. 이는 네트워크 내에 한 노드가 매개자나 중계자의 역할을 담당하는 노드를 찾는 방법으로, 연결 정도가 많고 적음에 상관없이 정보유통이나 소통 역활을 담당한다. 가장 낮은 매개중심성 값을 가진 것은 '로션'과 '추천'으로 나타났으며 그다음으로는 '무난'으로 나타났다.

평점 3점을 받은 보통 영역의 의미연결망 분석을 나타내는 Figure 5와 Table 5의 도출된 결과에서 노드 간의 중개 역활을 담당하며, 연결고리 및 연관관계를 나타내며 의미의 확산에 중요한 역할을 하는 매개중심성이 가장 높게 나타난 것은 '피부'와 함께 '흡수'로 나타났고, 그다음이 '사용', '얼굴', '수분'의 순서로 나타났으며, 가장 낮게 나타난 것은 '트러블', '토너', 그리고 '추천'과 함께 '주름'의 순으로 나타났다. 이중 '피부'는 '크림'과 함께 가장 낮은 연결중심성을 가지고 있으며, 근접중심성 또한 낮게 도출되었으나 매개중심성이 가장 높은 것으로 나타나 전체 연결망내에서는 중개자로서 높은 역할과 위치를 보이고 있어 Figure 5의 중앙에 위치하고 있음을 알 수 있다.

만족과 보통의 의미연결망 분석 결과에서 '피부', '사용', '얼굴' 등의 세 주제어가 각 영역에서 상위 5위 이내로 나타난 결과는 이들 세 주제어가 밀레니얼 화장품 이용 소비자에게 높게 지각된 브랜드 평가 속성의 우선순위 중요도와 함께 이들 평가 속성에 대한 의미의 확산에 중요한 역할을 하고 있음을 보여주고 있다.

본 연구에서 화장품 이용 소비자가 소셜미디어에 사용 후 지각한 브랜드 평가에 대해 담론 형식으로 남긴 리뷰를 수집하여 텍스트 마이닝 기법의 하나인 의미연결망 분석을 이용하여 발견하고자 하는 것은 빅데이터 내에 숨겨진 패턴과 트렌드 그리고 상호관계를 분석하고자 하는 것이다. 만족/보통/불만족 세 영역에서 가장 흥미로운 영역이라고 할 수 있는 불만족 평가영역의 의미연결망을 나타내고 있는 Figure 6과 Table 6으로 제시된 불만족의 의미연결망 분석 결과를 살펴보면, 만족과 보통의 의미연결망 분석과는 다르게 '피부', '제품', '트러블', '구매', '사용' 등의 주제어가 각 매개중심성, 연결중심성, 그리고 근접중심성의 상위에 도출되어 있다.

불만족 영역에서 가장 높은 연결중심성을 보이는 것은 '제품'으로 9개의 연결중심성을 가지고 있는 것으로 도출되었다. 이는 전체 의미연결망 분석에서 가장 높은 수치를 가진 강력한 연결중심성 허브 주제어로 나타나 '제품'을 중심으로 '크림', '피부', '스킨', '트러블', '얼굴', '건성', '생각', '효과', '성분' 등의 주제어와 함께 불만족과 연계되고 있음을 보여주는 중요한 지표의 역할을 나타내고 있으며, 불만족이 발생하게 되는 중요한 역할을 하고 있어 유념해서 관찰하여야 할 주제어이다.

의미연결망 분석에서 일반적으로 중심성이 높다는 것은 생존율이 높거나 기업 성과가 좋다고 표현하지만(An, 2009), 불만족 분석에 나타난 중심성이 높다는 것은 부(負)의 성격을 가지고 있어 반대로 브랜드 전환과 고객 이탈에 높은 영향력을 미치는 우선순위 요인으로 해석하여야 할 것이다.

이러한 분석 결과는 도출된 Figure 3과 Table 3의 워드클라우드 분석 결과와 무관하지 않은 것으로 보이며, 확인적 요인분석을 위한 관측변수로 도입할 주제어로만 볼 것이 아니라 불만족 고객이 남긴 상위 주제어인 '피부', '제품', '트러블', '구매', '사용'에 대한 전체 담론도 확인하여 고객 만족을 위한 대응 전략을 수립하여야 할 것이다.

Table 7. Measurement item

NO	Code	Observation variable
SAT (Satisfaction)	SAT 1	Overall, I am satisfied with the cosmetics that I am currently using.
	SAT 2	I will recommend the cosmetics brand that I am currently using.
	SAT 3	I am satisfied with the moisturizing effect of the cosmetics currently used.
	SAT 4	I am satisfied with the skin cosmetics that I am currently using.
MOD (Moderate)	MOD 1	I am satisfied with the soothing effect of the cosmetics that I am currently using.
	MOD 2	I am satisfied with the absorption effect of the cosmetics that I am currently using.
	MOD 3	I am satisfied with the price of the cosmetics that I am currently using.
	MOD 4	I am satisfied with the effect of the cosmetics that I am currently using.
DIS (Dissatisfaction)	DIS 1	I will not repurchase the cosmetic brand that I am currently using.
	DIS 2	I am dissatisfied with the cosmetic products that I am currently using.
	DIS 3	I am dissatisfied because the cosmetics that I am currently using cause skin trouble.
	DIS 4	I am dissatisfied because the fragrance of the cosmetics that I am currently using does not meet my expectations.

Selected discourse review texts with high centrality for each domain derived through wordcloud and semantic network analyses as latent variables for confirmatory factor analysis.

3. 확인적 요인분석 결과

확인적 요인분석은 탐색적 요인분석이 탐색적 형식으로 관측 문항이 그들의 내재된 요인들과 어느 정도 연결되어 있는가를 결정하기 위해 진행하는 것과는 다르게 연구자가 설정한 연구모형을 분석하기 전 단계에서 분석을 위한 관측 문항들이 실제 자료와 부합하는지를 확인 또는 검증하는 통계기법으로(Kim & Song, 2009; Park *et al*., 2014) 모든 잠재변수에 대해 수행되어야 하며, 잠재변수별로 개별적으로 평가할 수도 있으나 일반적으로 전체 잠재변수가 하나의 모델로 구성된 통합 측정모델을 대상으로 한꺼번에 수행한다.

확인적 요인분석을 통해 잠재변수의 단일차원성, 신뢰도, 타당도를 평가할 수 있으며, 신뢰도와 타당도를 평가하기에 앞서 단일차원성에 대한 평가를 먼저 수행한다. 측정모델을 구성하는 관측변수는 오직 하나의 구성 개념인 잠재변수만을 측정해야 하는 데 측정모델의 이러한 특성을 단일차원성이라고 하며, 측정모델의 단일차원성은 각각의 잠재변수가 단일요인 모델에 의해 잘 적합 되는지로 평가한다. 모델의 전반적인 적합도는 다양한 적합도 지표를 통해 평가할 수 있다(Kwahk, 2019).

만족/보통/불만족의 의미연결망 분석 결과에서 도출된 각 중심성은 통계 분석에서 훌륭한 독립변수나 종속변수가 될 수 있어(An, 2009) 본 연구의 워드클라우드 분석과 의미연결망 분석에서 중심성이 높은 주제어가 포함된 담론 리뷰를 연구 목적에 맞게 정제 과정을 거쳐서 만족/보통/불만족의 세 영역에 대한 관측변수로 도입하여 확인적 요인분석을 실시하였다.

Figure 7의 세 잠재변수를 보면 각각 네 개의 관측변수에 의해 측정되며, 도출된 지표에 의한 신뢰도는 무작위 측정오차에 의해 영향을 받게 된다. 그리고 각각의 관측변수들은 그들과 관련된 요인에 의해 회귀 되고 서로 상관되어 있다. 확인적 요인분석 모델에서 추정되어야 할 주요한 모수들은 회귀계수인 요인적재치와 오차분산 그리고 설명력(R2)이 된다. 구조방정식 모델을 분석하기 전에 모델에서 추정되어야 할 모수의 수를 계산하는 것은 중요하며, 이러한 정보는 검증하고자 하는 모델이 통계적으로 식별 가능한가를 파악하는 데 중요하다.

확인적 요인분석을 통한 측정모델의 신뢰도 및 타당성 검증은 잠재 개념 간 구조관계 분석에 대한 논리적인 근거가 된다. 확인적 요인분석의 주요 적합도 기준치를 살펴보면 χ^2값은 작을수록 바람직하며, p값은 유의수준 ≥ 0.05가 바람직하며, 적합지수(Goodness-of-Fit Index, GFI)는 0.90 이상이면 양호하고, 조정적합지수(Adjusted Goodness-of-Fit Index, AGFI)는 0.90 이상이면 양호하고, 근사오차자승평균이중근(Root Mean Square Error of Approximation, RMSEA)은 0.05 이하이면 좋으며, 표준부합지수(Normed Fit Index, NFI)는 0.90 이상이면 양호하며, 비교적합지수(Comparative Fit Index, CFI)는 0.90 이상이면 양호하고, 증분적합지수(Incremental Fit Index, IFI)는 0.90 이상이면 양호하며, 비표준적합지수(Non-Normed Fit Index, NNFI)도 0.90 이상이면 모델 적합도가 양호하여 전체적인 측정모형을 분석하는 데에는 무리가 없는 것으로 권장된다(Schumacker & Lomax, 1996; Kim & Song, 2009).

본 연구에서는 확인적 요인분석을 위해 LISREL 8.80 프로그램을 이용하여 각 요인과 그들의 측정을 위한 변수들 간의 관련성 측정 분석을 실행하였으며, 그 결과는 Figure 7 그리고 Table 9와 같이 나타났다. Table 9에 제시된 바와 같이 확인적 요인분석의 적합도는 권장수준을 만족하고 있으며, 많은 국내·외 연구논문에 제시된 확인적 요인분석의 적합도 수준을 수렴하고 있는 것으로 확인된다.

도출된 결과에서 χ^2는 78.98이며 자유도(df)는 51로 나타났으며, 이에 대한 p=0.007로 나타났다. 좀 더 살펴보면 적합지수(GFI)는 0.94, 표준적합지수(NFI)는 0.88, 비표준적합지수(NNFI)는 0.94, 근사오차자승평균이중근(RMSEA)은 0.049, 비교적합지수(CFI)는 0.95, 증분적합지수(IFI)는 0.95, 최초표준화부합도지수(GFI)는 0.94, 그리고 조정적합지수(AGFI)는 0.91로 나타났다. 따라서 확인적 요인분석의 전체 모형은 모집단 자료 분석에 적합하다고 판단할

Table 8. Correlation coefficient between constituent concepts

	x1	x2	x3	x4	x5	x6	x7	x8	x9	x10	x11	x12
x1	1											
x2	-0.001	1										
x3	0.168	-0.138	1									
x4	0.446	-0.040	0.118	1								
x5	0.382	-0.063	0.319	0.316	1							
x6	0.382	-0.063	0.319	0.316	0.524	1						
x7	0.427	-0.037	0.228	0.358	0.426	0.426	1					
x8	0.410	-0.039	0.268	0.344	0.487	0.487	0.388	1				
x9	-0.041	0.126	-0.164	-0.048	-0.090	-0.091	-0.068	-0.066	1			
x10	0.111	-0.091	-0.072	0.097	0.038	0.038	0.066	0.064	0.117	1		
x11	0.012	-0.141	-0.178	0.075	-0.068	-0.068	-0.038	-0.045	-0.158	0.020	1	
x12	-0.030	-0.181	-0.137	-0.067	-0.028	0.028	0.010	-0.047	-0.124	-0.026	0.129	1

Correlation analysis of the observed variable texts in Table 7.

http://dx.doi.org/10.20402/ajbc.2022.0072

수 있다.

Figure 7의 화살표 상의 값은 경로계수 또는 요인 적재치 혹은 요인 부하량이라고도 한다. 이는 잠재변수에 미치는 영향력이다. 예컨대 만족(SAT)에서 경로계수가 가장 높게 나타난 것은 x1으로 경로계수가 0.67로 나타났다. 이는 만족의 관측변수 중 '현재 사용하고 있는 화장품에 대해 전반적으로 만족한다.'라는 관측변수가 만족이라는 잠재변수에 가장 높은 영향력을 미치고 있음을 보인다. 이는 만족요인에서 이 관측변수인 전반적 만족도 수준을 1단위 높이게 되면 만족이 0.67만큼 증가하는 것으로 해석할 수 있다.

부(負)의 부호로 나타난 x2의 경로계수는 −0.07로 나타났다. 이는 만족요인에서 관측변수 x2인 '현재 사용하고 있는 화장품 브랜드를 다른 사람에게 추천할 것이다.'라고 하는 만족으로 인한 추천의도의 수준 1단위를 높이게 되면 만족에서 다른 관측변수와는 다르게 0.07만큼의 영향력이 줄어들며, 다른 나머지 세 개의 관측변수의 영향력은 상대적으로 늘어나는 것으로 해석할 수 있다. 이러한 결과는 밀레니얼 화장품 이용 소비자는 자신이 사용하고 있는 화장품 브랜드를 다른 사람에게 추천하는 것을 상대적으로 그렇게 중요하게 생각하지 않는 데 원인이 있으며, 화장품이라는 상품만이 가지고 있는 특성으로 유추 해석할 수 있다.

보통의 잠재변수에서 '현재 사용하고 있는 화장품 가격에 대해 만족한다.'와 '현재 사용하고 있는 화장품 사용 효과에 대해 만족한다.'가 동일하게 0.71로 나타났다. 이들은 12개의 관측변수 중 가장 높은 영향력을 가지고 있는 것으로 나타내고 있으며, 화장품 가격과 사용효과의 수준을 1단위 높이면 보통의 잠재변수에 대해 그만큼의 영향력을 가지는 것으로 해석할 수 있다.

잠재변수 불만족에서 가장 낮은 영향력을 보이는 관측변수는 x11으로 0.11의 영향력을 가지고 있는 '현재 사용하고 있는 화장품으로 인해 피부에 트러블이 발생하여 불만족한다.'로 나타났다. 이는 불만족요인에서 피부 트러블 개선 수준을 1단위 높이면 불만족이 0.11만큼 증가하는 것이 아니라 감소함을 나타낸다. 즉, 세 영역 중 불만족 영역이 의미하는 것은 다른 영역과는 다르게 부(負)로 나타나기 때문에 그만큼 "−"가 되는 것으로 해석되어야 한다.

이 불만족 잠재변수에서 부의 경로계수로 나타난 세 관측변수 x9, x10, x12는 각각 −0.48, −0.10, 그리고 −0.13으로 측정되었다. 즉 '현재 사용하고 있는 화장품 브랜드는 다음에 재구매하지 않을 것이다.', '현재 사용하고 있는 화장품 제품에 대해 불만족한다.', 그리고 '현재 사용하고 있는 화장품 향은 나의 기대에 미치지 못해 불만족한다.'라고 하는 이들 각 관측변수의 수준을 1단위 높이게 된다면, 화장품 이용 소비자의 불만족이 나타난 경로계수만큼 영향력이 줄어드는 것으로 해석 가능하며, 관측변수에서 재구매를 나타내고 있는 x9의 경우, 도출된 불만족 관측변수를 긍정적으로 1단위 수준을 높이는 노력을 한다면 불만족이 −0.48만큼 줄어들어 미약하나마 고객 이탈 의도와 브랜드 전환의도를 낮출 수 있는 것으로 해석할 수 있다.

잠재변수 간의 측정에서 만족과 보통은 0.88, 만족과 불만족은 0.15, 그리고 보통과 불만족은 0.25로 나타나 만족과 보통은 높은 상호관계를 가지고 있으며 유사성이 높아 차별성이 낮으며, 만족과 불만족 그리고 보통과 불만족의 경우에는 아주 낮은 상호관계를 보이고 유사성이 낮음을 알 수 있어 잠재변수 간의 차별성이 두드러진다고 할 수 있다. 이와 같은 확인적 요인분석의 결과는 유의미한 관계가 되는 것으로 볼 수 있다.

일반적으로 가설검증을 위한 유의수준은 0.01이나 0.05 수준에서 정해진다. 이 유의수준보다 크게 발생하면 영가설은 기각되지 않고 이 유의수준보다 작게 발생하면 영가설은 기각되고 대립가설이 채택된다. 유의수준 0.01에서 임계값은 2.58이며, 유의수준 0.05에서의

Table 9. Confirmatory factor analysis for the measurement model

Latent variable	Observation variable	Factor loading	R^2	Error variance
SAT (ξ1)	SAT 1	0.67	0.45	0.55
	SAT 2	-0.07	0.01	0.99
	SAT 3	0.38	0.14	0.86
	SAT 4	0.57	0.33	0.67
MOD (ξ2)	MOD 1	0.71	0.50	0.50
	MOD 2	0.71	0.50	0.50
	MOD 3	0.62	0.39	0.61
	MOD 4	0.68	0.46	0.54
DIS (ξ3)	DIS 1	-0.48	0.21	0.85
	DIS 2	-0.10	0.01	1.00
	DIS 3	0.11	0.01	1.10
	DIS 4	-0.13	0.02	1.06
Model fit	χ^2=78.98 (df=51, p=0.007); Goodness of Fit Index (GFI)=0.94; Adjusted Goodness of Fit Index (AGFI)=0.91; Root Mean Square Error of Approximation (RMSEA)=0.049; Non-Normed Fit Index (NNFI)=0.94; Normed Fit Index (NFI)=0.88; Comparative Fit Index (CFI)=0.95; Incremental Fit Index (IFI)=0.95,			

임계값은 1.96이다. 본 연구와 같은 구조방정식 모델의 확인적 요인분석의 적용에서 분석 결과가 전반적인 모델 적합도 지수를 만족시키고 있을 뿐만 아니라 Chi-Square값 역시 만족시키고 있으며, P값 역시 0.007로 나타났기 때문에 유의수준 $p<0.05$, $p<0.01$ 범위에서 회귀계수가 0이라는 영가설을 기각할 수 있다. 따라서 대립가설을 채택할 수 있는 모수치들 사이에 확률적으로 차이가 있으며, 연구모델은 적합하다는 결론을 내릴 수 있다.

Conclusion

본 연구는 밀레니얼 소비자를 대상으로 이들이 구매한 화장품에 대해 감성적 담론의 형식으로 작성한 개별정서 반응 리뷰를 텍스트 마이닝 분석 기법으로 만족/보통/불만족의 세 영역으로 나누어 워드클라우드 분석과 의미연결망 분석을 실시하여 각 영역 간의 유사점과 차이점을 분석하고, 이들 평가영역을 잠재변수로 설정하여 각 영역에서 중심성이 높게 도출된 주제어가 포함된 텍스트 문장을 연구의 목적에 맞게 정제 과정을 거친 후 관측변수로 도입하여 구조방정식 모델 분석 방법으로 확인적 요인분석을 실시하였다.

화장품 기호가치 소비에 대한 개별감성 반응 분석 과정을 전개함으로써 분석 방법에 대한 어떤 유의미한 차이가 있으며, 텍스트 마이닝 기법으로 도출된 결과를 이용하여 확인적 요인분석이 가능한가? 그리고 선행연구와는 어떤 다른 시사점을 제공할 수 있을까? 라는 의문점에서 수행한 연구 결과는 다음과 같다.

첫째, 소셜미디어에 담론 형태로 존재하고 있는 밀레니얼 소비자의 개별정서 반응 빅데이터를 크롤링하여 소비자 자신이 구매한 화장품 브랜드 선택에 대해 어떻게 평가하고 있는가를 분석함으로써 설문조사와 직접 면담을 통한 심층 인터뷰보다 절차 과정이 다소 어렵게 진행되었지만, 외부 통제없이 자유스러운 환경에서 소비자 스스로 작성한 빅데이터를 이용한 분석은 다른 실증 연구보다는 훨씬 효과적이고 의미 있는 연구가 될 것으로 판단된다.

둘째, 소셜미디어에 존재하는 개별정서 반응을 소비자 관점으로 만족/보통/불만족의 세 영역으로 분류하여 워드클라우드 분석과 의미연결망 분석으로 얻게 되는 이들 영역 간의 유사점과 차이점의 결과는 빅데이터 분석이라는 학문적 의의뿐만 아니라 소비자에게 제공하는 서비스 품질의 질적 향상을 기대할 수 있으며, 각 영역의 분석 결과에서 빈도가 높게 도출된 주제어와 중심성이 높게 도출된 주제어는 밀레니얼 화장품 이용 소비자들이 지각하는 주요한 트랜드와 이슈의 평가지수가 되어 화장품 구매 태도에 영향을 미치는 브랜드 선택 속성으로 전략적 시사점을 제공하고 있어 화장품 기업에게 전략 수립을 위한 우선순위 체계를 제공할 수 있을 것이다.

셋째, 불만족 평가영역에서 가장 높은 연결중심성을 보이는 주제어는 '제품'으로 다른 9개의 주제어 노드와 연결되는 것으로 나타났다. 이는 전체 평가영역에서 가장 높은 수치로 나타났으며, 의미연결망 분석에서 강력한 허브 주제어로 나타나 '제품'을 중심으로 '크림', '피부', '스킨', '트러블', '얼굴', '건성', '생각', '효과', '성분' 등의 주제어와 함께 불만족과 연계되고 있음을 나타내는 중요한 지표의 역할을 하는 것으로 보인다. 의미연결망 분석에서 일반적으로 중심성이 높다는 것은 생존율이 높거나 기업 성과가 좋다고 표현하지만, 불만족 영역 분석에 나타난 중심성이 높은 것은 반대로 유념해서 관찰하여야 할 주제어로 해석하여야 하며, 또한 고객 만족을 위해 다른 마케팅 4P's 믹스 요인을 보완하여 경쟁력 강화를 위한 노력을 하여야 할 것이다.

넷째, 구매 후 담론 형식으로 소셜미디어에 남긴 개별정서 반응 리뷰에 대해 워드클라우드 분석과 의미연결망 분석으로 중심성이 높게 도출된 주제어를 만족/보통/불만족의 잠재변수를 측정하는 관측변수로 도입하여 확인적 요인분석을 실시함으로써 개별정서로 이루어진 중심성이 높은 각 주제어와의 상관관계와 높은 유사성을 알 수 있었으며, 잠재변수에 대한 관측변수로 도출된 측정변수는 유의수준 0.01과 0.05 수준에서 영가설이 기각되고 대립가설로 채택되는 조건을 갖추었기에 다른 유사한 인과분석에서 유의미한 연구가설 측정항목으로 선정 가능한 기초자료로서의 활용이 가능해진다.

연구의 시사점으로는

첫째, 밀레니얼 소비자에 의해 각 영역에서 얻게 되는 개별정서 반응의 연구 결과에 따라 불만족 영역에 나타난 주제어 요인들은 브랜드 자산 강화를 위한 개선 활동을 통하여 제거할 기회도 가질 수 있어 화장품 메타 마켓에서 대고객 서비스 향상을 위한 폭넓은 실무적 시사점도 제시할 수 있을 것이다.

둘째, 향후 연구에서는 특정 주제의 연구를 대상으로 본 연구와 같이 텍스트 마이닝 기법을 이용한 확인적 요인분석과 이때 도출된 동일한 측정 문항을 적용한 실증 분석의 확인적 요인분석을 동시에 수행하여 도출된 결과를 상호비교해 본다면 학문적으로 의미가 있으며 유용한 시사점을 제공할 수 있을 것이다.

셋째, 최초로 수행한 텍스트 마이닝 기법을 이용한 구조방정식 모델의 확인적 요인분석적용을 위해 텍스트 마이닝 분석으로 도출된 높은 중심성의 주제어가 포함된 담론 텍스트를 상관분석하여 관측변수로 도입하여 이를 확인적 요인분석한 결과는 만족할 만한 성과를 이루었다고 생각한다. 향후 텍스트 마이닝 기법의 발전과 함께 자연어 처리 기법과 현재의 딥 러닝(deep learning) 기술의 한계를 뛰어넘는 발전이 지속된다면, 실험실의 질적연구나 설문 조사 방법을 하지 않고 자연어 처리 기법만으로도 잠재변수에 대한 관측변수로 도입하여 구조방정식 모델 분석을 통한 정량적 분석을 할 수 있으며, 유사한 인과분석을 위한 연구모델의 개발과 함께 측정 문항의 개발도 할 수 있어 이의 추가적인 연구가 필요하다.

Acknowledgements

This research was supported by the Ministry of Education of the Republic of Korea and the National Research

http://dx.doi.org/10.20402/ajbc.2022.0072

Foundation of Korea(NRF-2020S1A5B5A17089647).

Author's contribution

For this study, YJC and MJY performed data collection and analysis for research analysis. MSS designed this project and wrote the manuscript with assistance from YJC and MJY.

Author details

Man Seok Song (Lecturer), Byuckkang Liberal Arts College, Kyungwoon University 70, Gangdong-ro, Sandong-eup, Gumi-si, Gyeongsangbuk-do 39160, Korea; Yun-Jae Cho (Adjunct Professor), School of Start-up & Assets Convergence (Marketing), Changwon National University, Changwon, Gyeongsangnam-do 51140, Korea; Mi Ju Yim (Adjunct Professor), College of Business Administration, Dong-A University 225, Gudeok-ro, Seo-gu, Busan 49236, Korea.

References

Aggarwal CC, Wang H. Text mining in social networks. In Social network data analytics. Aggarwal CC (ed.), Springer, Boston, MA, pp353-378, 2011.

Ahn MJ. Knowledge, attitude, and prevention practice of cosmetology students toward Covid-19 infectious disease. *Asian Journal of Beauty and Cosmetology*, 19: 459-466, 2021.

Anderson JC, Gerbing DW. Structural equation modeling in practice: a review and recommended two-step approach. *Psychological Bulletin*, 103: 411-423, 1988.

An TG. Network structure between advertising agency and Event promotion agency using social network analysis. Doctoral dissertation. Dongguk University, 2009.

Bae JH, Han NG, Song M. Twitter issue tracking system by topic modeling techniques. *Journal of Intelligence and Information Systems*, 20: 109-122, 2014.

Bagozzi RP, Yi Y. On the evaluation of structural equation models. *Journal of the Academy of Marketing Science*, 16: 74-94, 1988.

Bollen KA. Structural equations with latent variables. John Wiley & Sons, Inc, New York, pp10-36, 1989.

Cha MK. Semantic network analysis of “arts management” in newspaper articles: from 1990 to 2014. *The Journal of Cultural Policy*, 29: 168-200, 2015.

Cho M, Ko EB, Kim JH, Lee JM. Effect of customized cosmetics selection attributes and safety perception on intention to purchase. *Asian Journal of Beauty and Cosmetology*, 19: 513-524, 2021.

Day RL. Research perspectives on consumer complaining behavior. In Theoretical developments in marketing. Lamb CW, Jr, Dunne PM (eds.), The American Marketing Association, Chicago, IL, pp211- 215, 1980.

de Leeuw J. Review of four books on causal analysis. *Psychometrika*, 50: 371–375, 1985.

Freedman DA. As others see us: a case study in path analysis. *Journal of Educational Statistics*, 12: 101-128, 1987.

George G, Haas MR, Pentland A. Big data and management. *Academy of Management Journal*, 57: 321-326, 2014.

Guttman L. What is not what in statistics. *Journal of the Royal Statistical Society*, 26: 81-107, 1977.

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.

Jöreskog KG. A general approach to confirmatory maximum likelihood factor analysis. *Psychometrika*, 34: 183-202, 1969.

Kim BS, Lee JM, An SK. Effect of online cosmetics information on brand trust and continuous purchasing intentions online. *Asian Journal of Beauty and Cosmetology*, 19: 395-408, 2021.

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 JJ. A study on the jewelry market trend reflecting the characteristics of millennial generation: focusing on the trend of global jewelry brands. *Journal of Korea Design Forum*, 24: 115-126, 2019.

Kim JS. A study on the perception of fashion streaming service using text mining analysis: focused on project anne. *Journal of Fashion Design*, 18: 107-118, 2018.

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 SH, Song MS. A study on the perceived risk, wating time, and communication on repurchase intention of the health care consumer. *Academy of Customer Satisfaction* Management, 11: 19-47, 2009.

Ko JY, Seo HJ. A study on the selection attribute of coffee consumer's. *Journal of Hotel & Resort*, 8: 23-41, 2009.

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.

Kwahk KY. Structural equation modeling using R: mediation/moderation effect analysis and multiple-group analysis. *Knowledge Management Research*, 20: 1-24, 2009.

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 JS, Jun BR, Kim BR, Lee JY. An analysis of the appearance management behavior of generation MZ in the post-corona era. *Asian Journal of Beauty and Cosmetology*, 19: 543-553, 2021.

Lee SK. A review of big data analysis based on marketing perspective. *Korea Journal of Business Administration*, 28: 21-35, 2015.

Lee SS. A content analysis of journal articles using the language network analysis methods. *Journal of the Korean Society for Information Management*, 31: 49-68, 2014.

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.

Park EH, Lee SJ. The influence of female consumer's cosmetics purchase experience on image-making efficacy and cosmetics pursuit benefits. *Journal of the Korea Fashion & Costume Design Association*, 17: 131-144, 2015.

Park HH, Ku YS. The influence of female university students' cosmetic purchase motivation on cosmetic attribute evaluation and brand repurchase intention. *Fashion & Textile Research Journal*, 11: 252-261, 2009.

Park HJ. The new cosmetics trend research in the new normal era through big data analysis. *Korean Society of Cosmetics and Cosmetology*, 10: 465-479, 2020.

Park IH, EOM HJ, Lee GB. Issues and concurrent solutions for the application of confirmatory factor analysis in physical education and exercise science research. *The Korean Journal of Measurement and Evaluation in Physical Education and Sport Science*, 16: 1-22, 2014.

Preedy VR, Watson RR. Handbook of disease burdens and quality of life measures. Springer, New York, pp35-57, 2009.

Sabidussi G. The centrality index of a graph. *Psychometrika*, 31: 581-603, 1966.

Schumacker RE, Lomax RG. A beginner's guide to structural equation modeling. Lawrence Erlbaum Associates, NJ, pp79-122, 1996.

Seo EH, Kim JD. A study on the Korean cosmetic purchase behaviors of the 20's woman in Beijing, China. *Journal of the Korean Society of Cosmetology*, 15: 578-586, 2009.

Shin YW, Lee CYJ. A study of brand communication through the consumer trend of the millennial generation and mirrors from a dualistic perspective. *Journal of the Korean Society of Design Culture*, 24: 355-364, 2018.

Song MS. What semiotic values do cosmetics consumers consume?: analysis of cosmetic brand reputation and brand selection attributes using text mining. *Asian Journal of Beauty and Cosmetology*, 19: 263-275, 2021.

Strauss W, Howe N. Generations: the history of America's future, 1584 to 2069. William Morrow & Co, New York City, pp337-340, 1991.

Steyvers M, Griffiths T. Probabilistic topic models. In Handbook of latent semantic analysis. Landauer TK, McNamara DS, Dennis S, Kintsch W (eds.), Psychology Press, UK, pp424-440, 2007.

Yu JP. The criticisms and considerations of structural equation modeling. *Journal of Product Research*, 34: 83-93, 2016.

Zhu D, Lappas T, Zhang J. Unsupervised tip-mining from customer reviews. *Decision Support Systems*, 107: 116-124, 2018.

http://dx.doi.org/10.20402/ajbc.2022.0072

국문초록

밀레니얼 소비자의 화장품 구매후 개별정서 반응: 텍스트 마이닝 기법을 이용한 만족/보통/불만족 속성의 평가

송만석[1*], 조윤재[2], 임미주[3]
[1]경운대학교 벽강교양대학, 경상북도 구미시, 한국
[2]창원대학교 미래융합대학, 경상남도 창원시, 한국
[3]동아대학교 경영학과, 부산, 한국

목적: 본 연구는 밀레니얼 세대에 의해 주관적으로 생성된 구매 후 만족/보통/불만족에 대한 속성 리뷰의 구전 정보를 크롤링하여 텍스트 마이닝 기법인 워드클라우드 분석과 의미연결망 분석으로 각 영역간의 유사점과 차이점을 분석하고 확인적 요인분석을 실시하고자 한다. **방법:** 본 연구를 위해 R버전 4.1.2 프로그램을 빅데이터 수집 및 분석 도구로 사용하였으며, 수집된 데이터에 대한 전처리 및 불용어 처리 과정을 거쳐서 연구 목적에 맞는 텍스트 마이닝 분석을 수행하였다. 그리고 이 결과를 이용한 확인적 요인분석은 LISREL 8.80 프로그램을 이용하여 수행하였다. **결과:** 워드클라우드 분석 결과 '만족' 평가영역에서 피부가 1위, '보통' 평가영역에서 제품이 1위, 평가영역에서 1위가 되는 것으로 나타났다. '불만족' 평가영역에서는 피부가 1위를 차지하였다. 그리고 확인적 요인분석에서는 만족/보통/불만족 3개의 잠재변수의 상관관계가 차별화가 되는 것으로 나타났다. **결론:** 소셜미디어에 존재하는 밀레니얼 소비자의 개별정서 반응을 만족/보통/불만족 영역으로 분류하여 도출된 워드클라우드 분석과 의미연결망 분석으로 얻게 되는 이들 영역간의 유사점과 차이점의 결과는 매우 의미 있는 결과를 얻었다고 볼 수 있다. 그리고 텍스트 마이닝에서 도출된 확인적 요인분석의 관측변수는 인과분석에서 수행하는 설문지의 측정 항목으로 도입될 수 있을 것으로 기대된다.

핵심어: 밀레니얼 소비자, 텍스트 마이닝, 워드클라우드 분석, 의미연결망 분석, 확인적 요인분석

This research was supported by the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea(NRF-2020S1A5B5A17089647).

참고문헌

고재윤, 서혜진. 커피이용객의 선택속성에 관한 연구. *호텔리조트연구*, 8: 23-41, 2009.

곽기영. R을 이용한 구조방정식모델링: 매개효과분석/조절효과분석 및 다중집단분석. *지식경영연구*, 20: 1-24, 2019.

김기룡, 김삼섭. 텍스트마이닝을 이용한 국내 통합교육 연구 동향 분석. 지체. 중복. *건강장애연구*, 59: 133- 157, 2016.

김동현, 정윤진, 조원일. AHP를 활용한 온라인 화장품 몰에서 제품구매 시 선택속성에 관한 연구. *경영교육연구*, 33: 387-406, 2018.

김병수, 이정민, 안성관. 온라인 화장품 정보가 브랜드 신뢰 및 온라인에서의 지속적 구매의도에 미치는 영향. *아시안뷰티화장품학술지*, 19: 395-408, 2021.

김성환, 송만석. 의료소비자의 지각된위험, 대기시간, 커뮤니케이션이 재구매의도에 미치는 인과관계 연구. *고객만족경영연구*, 11: 19-47, 2009.

김정지. 밀레니얼 세대의 특성을 반영한 주얼리 시장 경향 연구: 해외주얼리 브랜드 사례를 중심으로. *한국디자인포럼*, 24: 115-126, 2019.

김종선. 텍스트 마이닝을 통한 패션 스트리밍 서비스 인식에 관한 연구: 프로젝트 앤 사례를 중심으로. *한국패션디자인학회지*, 18: 107-118, 2018.

남궁지은, 김보람, 조원일. AHP를 활용한 로드샵 화장품 구매 선택속성에 관한 연구: 한•중 소비자 비교 중심으로. *경영교육연구*, 32: 65-84, 2017.

박은희, 이상주. 여성소비자의 화장품구매체험이 이미지 메이킹 효능감과 화장품추구혜택에 미치는 영향. *한국의상디자인학회지*, 17: 131-144, 2015.

박일혁, 엄한주, 이기봉. 체육학연구에서 확인적 요인분석 이용의 문제점과 올바른 적용. *한국체육측정평가학회지*, 16: 1-22, 2014.

박혜진. 뉴 노멀 시대의 빅 데이터 분석을 통한 화장품 트렌드에 대한 연구. *한국화장품미용학회지*, 10: 465-479, 2020.

박현희, 구양숙. 여대생의 화장품 구매동기가 화장품 속성평가와 브랜드 재구매의도에 미치는 영향. *한국의류산업학회지*, 11: 252-261, 2009.

배정환, 한남기, 송민. 토픽 모델링을 이용한 트위터 이슈 트래킹 시스템. *지능정보연구*, 20: 109-122, 2014.

서은혜, 김주덕. 중국 20대 여성들의 한국 화장품에 대한 인지도 및 선호도 연구. *한국미용학회지*, 15: 578-586, 2009.

송만석. 화장품 소비자는 어떤 기호가치로 소비를 하는가?: 텍스트 마이닝을 이용한 화장품 브랜드 평판과 브랜드 선택속성 분석. *아시안뷰티화장품학술지*, 19: 263-275, 2021.

신용환, 이캐시연주. 이중적 관점으로 본 밀레니얼 세대의 소비와 거울을 통한 브랜드 커뮤니케이션 연구. *한국디자인문화학회지*, 24: 355-364, 2018.

안미정. 뷰티전공 대학생의 Covid-19에 대한 지식, 태도, 예방실천. *아시안뷰티화장품학술지*, 19: 459-466, 2021.

안태기. 사회연결망분석을 이용한 광고회사와 이벤트기획사간 관계 구조. 동국대학교 대학원 박사학위논문, 2009.

우종필. 구조방정식모델의 비판과 고찰. *상품학연구*, 34: 83-93, 2016.

이서구. 빅데이터 분석에 관한 마케팅적 접근. *대한경영학회지*, 28: 21-35, 2015.

이수상. 언어 네트워크 분석 방법을 활용한 학술논문의 내용분석. *정보관리학회지*, 31: 49-68, 2014.

이제성, 전보람, 김벼리, 이준영. 포스트 코로나 시대의 MZ세대의 외모관리행동 분석. *아시안뷰티화장품학술지*, 19: 543-553, 2021.

이종혁, 길우영. 토픽모델링을 이용한 뉴스 의제 분류와 미디어 다양성 분석: 대통령 신년 기자회견 관련 뉴스 분석을 통해. *한국방송학보*, 33: 161-196, 2019.

조민, 고은비, 김지혜, 이정민. 맞춤형화장품 선택속성과 안전성 인식이 구매행동 의도에 미치는 영향. *아시안뷰티화장품학술지*, 19: 513-524, 2021.

차민경. 국내 언론에 나타난 '예술경영'관련 이슈의 의미연결망 분석: 1990년부터 2014년까지 국내 일간지 기사 분석을 중심으로. *문화정책논총*, 29: 168-200, 2015.

한관종. 사회과학 방법론으로서의 연결망 분석기법 적용의 의의와 연구과제: 의미와 연결망 분석(semantic network analysis)을 중심으로. *사회과교육연구*, 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.

http://dx.doi.org/10.20402/ajbc.2022.0072

中文摘要

千禧一代消费者购买化妆品后的个人情绪反应：使用文本挖掘技术评估满意/一般/不满意的属性

宋晚碩[1*]，趙允載[2]，林美珠[3]
[1]庆云大学Byunkkang教养大学，庆尚北道龟尾市，韩国
[2]昌原大学未来融合大学，庆尚南道昌原市，韩国
[3]东亚大学经营学科，釜山，韩国

目的： 本研究抓取千禧一代主观产生的购买后满意/正常/不满意的属性评论口碑信息，通过词云分析和语义网络分析等文本挖掘技术分析各领域的异同，我们打算进行验证性因素分析。**方法：** 本研究使用R version 4.1.2程序作为大数据收集分析工具，通过对收集到的数据进行预处理和停用词处理，进行适合研究目的的文本挖掘分析。并且使用该结果使用LISREL 8.80程序进行验证性因素分析。**结果：** 词云分析结果发现，皮肤在“满意”评价区排名第一，产品在“正常”评价区排名第一，在评价区排名第一。在“不满意”评价区，皮肤排在第一位。并且在验证性因素分析中，发现满意/正常/不满意三个潜变量的相关性是有区别的。**结论：** 通过将千禧一代消费者在社交媒体中的个体情绪反应分类为满意/正常/不满意区域得到的词云分析和语义网络分析得到的这些区域的异同结果得到了非常有意义的结果。可以看到。此外，期望引入文本挖掘验证性因素分析的观测变量作为因果分析问卷的测量项。

关键词： 千禧一代消费者，文本挖掘，词云分析，语义网络分析，验证性因素分析