



Information Systems Review

제22권 제4호

ISSN : 1229-5078(Print) 2713-8143(Online)

머신러닝 기반의 뷰티 커머스 고객 세그먼트 분류 및 활용 방안 : 언택트 서비스 중심으로

윤상혁, 최윤진, 이소현, 김희웅

To cite this article : 윤상혁, 최윤진, 이소현, 김희웅 (2020) 머신러닝 기반의 뷰티 커머스 고객 세그먼트 분류 및 활용 방안 : 언택트 서비스 중심으로 , Information Systems Review, 22:4, 75-92

① earticle에서 제공하는 모든 저작물의 저작권은 원저작자에게 있으며, 학술교육원은 각 저작물의 내용을 보증하거나 책임을 지지 않습니다.

② earticle에서 제공하는 콘텐츠를 무단 복제, 전송, 배포, 기타 저작권법에 위반되는 방법으로 이용할 경우, 관련 법령에 따라 민, 형사상의 책임을 질 수 있습니다.

www.earticle.net

머신러닝 기반의 뷰티 커머스 고객 세그먼트 분류 및 활용 방안: 언택트 서비스 중심으로

A Study of the Beauty Commerce Customer Segment Classification and Application based on Machine Learning: Focusing on Untact Service

윤상혁 (Sang-Hyeak Yoon) 연세대학교 정보대학원 박사/스마트미디어랩 매니저
최윤진 (Yoon-Jin Choi) 연세대학교 정보대학원 박사과정
이소현 (So-Hyun Lee) 연세대학교 정보대학원 연구교수
김희웅 (Hee-Woong Kim) 연세대학교 정보대학원 교수, 교신저자

요 약

인구 및 세대 구조가 변화면서 점차 대면 관계를 꺼리는 고객의 태도 변화가 정보기술의 발달과 스마트폰의 확산으로 더욱 커지고 있다. 이는 정보기술에 익숙해진 현대 고객들의 소비패턴인 효율성 및 신속성과도 부합되는 것으로, 오프라인 망 중심의 유통회사들이 판매 및 서비스 방식을 언택트로 전환하려는 움직임이 활발해지고 있다. 최근 다양한 분야에서 언택트 서비스가 활성화되고 있지만, 뷰티 제품의 경우 고객의 피부타입 및 상태에 따라 제품 선택이 쉽지 않으므로 비대면을 통해 제품을 추천하기가 쉽지 않다. 이와 관련하여 온라인 뷰티 분야에서 제품 추천을 위한 추천시스템 개발 및 추천 관련 연구들이 수행되었지만, 대부분이 설문조사 방법이나 소셜 데이터를 이용하여 추천 알고리즘을 개발한 연구들이었다. 즉, 고객의 피부타입이나 제품 선호도 등의 실제 사용자 정보를 기반으로 세그먼트를 분류한 연구는 부족하였다. 그리하여, 본 연구에서는 뷰티 분야에서의 언택트 서비스 중의 하나인 모바일 애플리케이션의 고객 정보와 검색 로그 데이터를 기반으로 머신러닝 기법의 K-prototypes 알고리즘을 이용하여 고객 세그먼트를 새롭게 분류하고, 이를 기반으로 언택트 마케팅 전략 방안을 제안한다. 본 연구는 머신러닝 기법을 이용하여 새롭게 고객 세그먼트를 분류함으로써 관련 기존 문헌의 범위를 확장하였다. 더불어, 언택트 서비스라는 새로운 소비 트렌드를 반영하여 고객 세그먼트를 분류하고, 이를 기반으로 뷰티 분야의 언택트 서비스에 활용할 수 있는 구체적인 방안을 제시했다는 실무적 의의가 있다.

키워드 : 뷰티, 언택트, K-prototypes, 고객 세그먼트, 머신러닝

† 본 연구는 중소벤처기업부 중소기업기술정보진흥원이 지원하는 제품서비스기술개발사업의 지원을 받아 진행되었음 (S2754537).

I. 서론

최근 불필요한 소통 대신 편한 단절을 택하는 사람들이 증가하면서 정보기술을 이용한 비대면 거래의 셀프서비스인 언택트 서비스가 퍼지고 있다. 언택트(untact) 서비스는 직원이 고객과 접촉하지 않고 정보를 제공하는 것으로 직원들의 과도한 응대 대신 편하게 쇼핑할 수 있도록 지원하는 것을 말한다(Fitzsimmons *et al.*, 2008; Meuter *et al.*, 2000). 정보기술이 발달하고, 특히 모바일과 디지털 기기 이용에 친숙해지면서 이러한 비대면 서비스가 활성화되고 있다. 편리함을 극대화하고 스트레스를 최소화한 언택트 소비 트렌드는 새로운 마케팅의 유통 방식으로 4차산업혁명 시대의 인공지능, 빅데이터, 사물인터넷 등의 주요 기술과 접목되어 더욱 고도화될 것으로 예상된다(Lee and Lee, 2020). 또한, 고객을 직접 접촉하지 않기 때문에 소비자들의 다양한 요구에 대한 폭넓은 데이터를 수집하여 활용할 수 있다.

이러한 기술의 발전을 넘어 고객들이 타인과의 연결, 접촉을 불편해하고 피곤하게 여기는 소비패턴 변화의 이해도 필요하다. 시장조사 전문기업인 엠브레인의 트렌드모니터(2018)의 비대면 서비스 관련 설문조사에 따르면 매장에서 직원이 말을 거는 것보다는 혼자 쇼핑할 수 있는 것이 더 좋다는 의견이 859%로 가장 높았으며, 특히 뷰티 매장에서 혼자 쇼핑하는 것을 선호하였다. 2018년 소비트렌드 중 하나로 언택트 마케팅이 제시되면서 타인과 대면 거래 및 서비스를 불편하게 생각하는 사람들이 늘어나리라 전망하였다(김난도 등, 2018). 이는 인간관계에서 받는 스트레스가 늘어나면서 사소한 관계조차도 귀찮게 느껴져 접촉을 최소화하고자 하는 심리가 내포된 것으로 볼 수 있다. 이러한 고객들의 심리가 기술과 결합하여 언택트 마케팅이 활성화되고 있으며, 대부분 산업에서 이를 통한 새로운 서비스들을 소개하며 수익 증진에 노력하고 있다.

최근 금융거래, 외식업, 뷰티 커머스 등 다양한 분야에서 비대면 서비스의 언택트 마케팅이 주목

받고 있다. 사례로 스타벅스의 ‘사이렌오더’는 모바일 앱을 통해 미리 주문과 결제를 하고 매장에서 주문한 음료만 받아가는 언택트 서비스로 2017년 하루 이용 6만 건에서 2018년 7만 8,000건, 2019년 상반기 11만 건까지 증가하였다.¹⁾ 그 밖에, 맥도날드, 버거킹 등의 패스트푸드 매장에서는 소비자가 키오스크로 메뉴를 선택하고 결제까지 진행하도록 하고 있다. 실제 수익과 관련해서 현대카드와 현대캐피탈에서 언택트 서비스를 제공하는 주요 매장 15곳의 자료를 분석한 결과, 전체 매출이 2017년 67억 원에서 2019년 359억 원으로 5배 이상 급증한 것을 확인하였다. 특히, 20대의 결제금액의 비중이 28.4%에 달한 것으로 확인되어 20대의 소득수준이 기성세대에 비해 낮은 것을 고려하면 20대의 젊은 층이 언택트 문화를 주도하는 것을 확인할 수 있다.²⁾

이러한 언택트 마케팅에 관한 관심과 이를 활용한 업계의 매출 증가로 20대의 밀레니얼 세대(Millennial Generation)가 주 고객층인 뷰티샵에서도 언택트 서비스를 제공하고 있다. 하지만, 뷰티 제품의 경우 고객의 피부타입 및 상태에 따라 다른 제품을 사용해야 하므로 비대면을 통해 제품을 추천하는 것에 한계가 있다. 특히, 사용해보지 않은 새로운 제품을 테스트 없이 비대면으로 구매하는 것은 고객에게 쉽지 않은 결정이다. 또한, 자신의 피부타입에 맞는 제품이 무엇인지를 정확히 알고 뷰티 제품을 구매하는 소비자가 많지 않다. 최근 뷰티 매장에서 직원을 대신하여 스마트거울 등의 최신 기술을 통해 고객의 피부 상태를 파악하고 제품을 추천하고 있지만, 이에 대한 만족이나 실제 매출에 대한 검증이 이루어지지 않은 상태이다. 또한, 온라인 뷰티 커머스 분야에서 제품 추천을 위

- 1) 아시아경제, “‘편한단절’ 선호...도움 없는 쇼핑터 점포 등 ‘언택트’ 소비 인기”, 2019. 10. 25., Available at <https://www.asiae.co.kr/article/2019073001262737508>.
- 2) MOBIINSIDE, “[IT 트렌드 바로읽기] 유통가의 새바람! 언택트(untact)”, 2019. 07. 23., Available at <https://www.mobiinside.co.kr/2019/07/23/it-untact/>.

한 추천시스템 개발 및 제품 추천의 연구들이 수행되었지만(Okuda *et al.*, 2017; 김성연 등, 2017; 신혜란 등, 2019; 윤여수, 박현준, 2019; 이은주 등, 2018; 조영성 등, 2012), 대부분이 설문이나 소셜 데이터를 이용하여 추천 알고리즘을 개발한 것이었다. 즉, 피루타입, 제품 선호도 등의 실제 사용자 정보를 기반으로 세그먼트를 분류하여 추천 모델을 개발한 연구는 부족했다.

따라서, 본 연구에서는 언택트 서비스 중의 하나인 모바일 애플리케이션의 고객 정보와 검색 로그 데이터를 기반으로 머신러닝 기법을 이용하여 고객 세그먼트를 새롭게 분류하고, 이를 기반으로 언택트 서비스 전략 방안을 제안한다. 즉, 모바일 애플리케이션을 통해 수집한 고객 데이터를 기반으로 피루타입 및 제품 선호도에 따라 소비자 유형을 새롭게 분류하고, 이를 언택트 서비스에 활용하는 방안을 제시한다. 특히, 본 연구는 언택트 문화를 대표하는 20대가 주요 대상인 뷰티 분야에 머신러닝 알고리즘인 K-prototype을 이용하여 새롭게 고객 세그먼트를 분류했다는 학술적 의의가 있다. 더불어, 언택트 서비스라는 새로운 소비 트렌드를 반영하여 고객 세그먼트를 분류하고, 이를 언택트 마케팅에 활용할 수 있는 구체적인 방안을 제시했다는 실무적 의의가 있다.

II. 개념적 배경

2.1 언택트 마케팅

최근 스마트폰과 SNS의 확산과 함께 언택트 마케팅(untact marketing)이 화두로 떠오르고 있다. 언택트 기술은 셀프(self), 자동화(automation), 무인(unmanned) 등의 여러 기술적인 개념들을 포괄하지만, 고객으로서 체감하는 핵심 컨셉은 사람 간 접촉이 없는 것이다.³⁾ 이러한 언택트 기술을 기반

으로 한 언택트 마케팅은 접촉을 의미하는 콘텐츠에 언(un)이 붙어 ‘접촉하지 않는다’라는 뜻으로 사람과의 접촉을 최소화하는 비대면 형태로 정보를 제공하거나 물품을 거래하는 마케팅 기법을 말한다(오수연, 2018). 최근 다양한 분야에서 관심이 고조되고 있는 언택트 마케팅은 1인 가구의 급증과 같은 인구 및 세대 구조가 변화하면서 점차 대면 관계를 꺼리는 고객의 태도 변화가 정보기술의 발전에 힘입어 더욱 확산한 것으로 볼 수 있다. SNS, 블로그 등을 통한 제품의 정보 획득으로 점원과의 대화의 필요성을 느끼지 못하고, 매장 직원이 주는 무언의 압박과 스트레스에서 벗어나 편안한 쇼핑을 즐기기를 원하는 고객들이 증가하고 있는 것이다(Jaakkola *et al.*, 2015; Lee, 2019). 이는 정보기술에 익숙해진 현대 고객들의 소비패턴인 효율성 및 신속성에도 부합된다. 특히, 연령대가 낮을수록 비대면 형태의 서비스 이용에 대한 선호가 더 크게 나타나고 있으며, 기업으로서도 비용 절감 차원에서 효율적이다. 이로 인해, 오프라인 망 중심의 유통회사들이 판매 및 서비스 방식을 언택트로 전환하는 움직임이 활발해지면서 산업 전반에 언택트 서비스가 시도되고 있다.

다양한 분야에서 기계로 메뉴를 주문하는 키오스크나 가상현실(VR) 쇼핑, 인공지능(AI) 챗봇 등 첨단기술을 활용해 판매 직원이 소비자와 직접 대면하지 않고 상품이나 서비스를 제공하고 있다.⁴⁾ 특히, 뷰티 업계에서는 이러한 흐름에 따라 다양한 형식의 언택트 서비스를 선보이고 있다. 대표적으로 한국 신세계백화점의 뷰티 매장 시코르(CHICOR)는 고객이 요청하기 전까지는 매장 직원이 쇼핑에 관여하지 않는 ‘뷰티 놀이터’를 진행하고 있으며, 올리브영에서도 고객이 필요할 때만 응대하도록 서비스하고 있다. 중국 항저우의 이니스프리 매장에서는 스마트 스크린을 도입하여 매장에 설치된 스크린에서 제품 소개, 피부 상태 측정, 가상 메이

3) 전자신문, “‘2020년, 비대면 언택트(Un+Contact) 마케팅이 뜬다!’” 2019.10.08., Available at <https://m.etnews.com/20191008000126>.

4) 동아닷컴, “‘사람과 마주치기 싫어요’...‘언택트 소비’ 뜬다”, 2020.02.26., Available at <http://www.donga.com/news/article/all/20200225/99860823/1>.

크업 등을 직원들을 대신하여 수행하고 있다.⁵⁾ 여기에는 피부 상태를 점검하고 피부에 맞는 제품을 추천해 주는 기기부터, 진행 중인 세일 행사를 알려주고 최신 유행 화장품을 안내해 주는 기기 등으로 고객이 제품을 선택하는 것을 사람을 대신해서 도와주고 있다. 이렇듯, 뷰티 분야에서 다양한 언택트 서비스들이 제공되고 있으며 여기에는 사람과의 접촉이 없는 ‘언택트’의 특성만을 살린 서비스뿐만 아니라 기술과 결합한 지능화된 서비스들이 포함된다. 최근, 이러한 기술을 기반으로 뷰티 브랜드에서는 무인 점포를 개점하여 매장에 직원이 없는 대신 디지털 장치를 통해 제품을 사용하거나 구매할 수 있도록 하고 있다. 하지만, 언택트 서비스의 온라인에서는 직접 제품을 테스트해보지 못하는 단점이 있고, 오프라인에서는 제품 테스트는 가능하지만, 실제 뷰티 전문가의 조언을 얻지는 못하는 한계점이 있다. 이는 경험재인 뷰티 제품은 본인이 사용해 본 제품이 아니라면 직접 테스트해보는 것이 필요하고, 스마트 스크린과 같은 제품 추천 기능이 포함된 디지털 기기가 제공되기도 하지만 비용의 문제와 더불어 고객 만족의 검증이 이루어지지 않은 상태이다. 또한, 관련 기존 연구에서 개발한 뷰티 추천시스템은 대부분 설문이나 소셜 데이터를 이용한 것에 한계점이 있다.

그리하여, 본 연구에서는 뷰티 분야에 최근 소비트렌드를 반영한 언택트 마케팅 활성화에 이바지하기 위하여 언택트 서비스의 모바일 애플리케이션을 통해 광범위한 고객 데이터를 기반으로 고객 세그먼트를 새롭게 분류하고자 한다. 특히, 본 연구에서 사용하고자 하는 모바일 애플리케이션에서는 언택트 서비스를 사용하는 고객들이 뷰티 제품의 가격표만 카메라로 찍으면 자동으로 가격 비교, 제품특징, 리뷰 등의 정보를 제공해주는 서비스로 고객의 피부 상태, 선호 제품 등의 정보를 수집할 수 있다. 즉, 언택트 서비스를 통해 수집한

고객 데이터를 기반으로 머신러닝 기법을 이용하여 피부 상태 및 제품 선호도에 따라 새롭게 고객 유형을 분류하였다. 본 연구에서 분류한 유형을 기반으로 뷰티 제품 추천 모델을 개발하거나 뷰티 커머스 분야의 언택트 마케팅에 활용할 수 있다.

2.2 뷰티 제품 추천 관련 선행연구

뷰티 제품군은 실생활에서 매일 사용되는 소비재로, 그 브랜드와 종류가 매우 다양하다. 또한, 다양한 뷰티 제품 중에서 사용자에게 추천하기 위해서는 제품 인기뿐 아니라 고객의 피부 타입이나 상태도 고려해야 한다. 다시 말해, 뷰티 제품 추천을 일반적인 추천 알고리즘에 적용하게 되면 피부 문제와 같은 부작용이 발생할 수 있다(김성언 등, 2017). 이에, 뷰티 제품 추천과 관련된 연구들은 기존 추천 연구들과 다른 형태로 발전되고 있다.

뷰티 제품 추천과 관련된 선행연구를 살펴보면, 소비자의 시각 감성을 활용한 화장법(makeup style) 추천(Chung, 2014), 퍼지 추론과 감성 사전 기반 뷰티 제품 추천 시스템(김성언 등, 2017). 개인의 피부 유형에 따른 빅데이터 기반 뷰티 제품 추천(Okuda *et al.*, 2017; 신혜란 등, 2019; 이은주 등, 2018), 이미지 인식 기반의 화장품 추천(윤여수, 박현준, 2019) 등의 연구들이 있다(<표 1> 참조).

구체적으로 Chung(2014)은 화장법(Facial Makeup-style) 추천시스템을 협업 필터링(Collaborative Filtering) 기반으로 구축하였다. 해당 시스템은 사용자가 선호하는 화장법과 관련된 질문에 답하게 되면 PC 서비스로 구현된 화면을 통해 화장법을 제안한다. 총 978명의 이용자로부터 데이터를 수집해 사용자의 시각적 감성(Visual Sensibility)을 현대적(Contemporary), 성숙(Mature), 개성(Individual)으로 분류하였다. 마지막으로 연구에서 제안한 추천시스템이 사용자의 시각적 감성에 영향을 미친다는 것을 실증적으로 증명하였다.

김성언 등(2017)은 뷰티 제품 성분 분석과 뷰티 제품과 관련된 고객 댓글을 이용하여 뷰티 제품을

5) The polestar, “간섭은 거부한다! 소리 없이 강한 언택트 마케팅”, 2019.01.03., Available at <https://blog.naver.com/thepolestar01/221433455301>.

〈표 1〉 뷰티 추천 관련 선행연구

저자	데이터 수집	추천 알고리즘	연구 목표
Chung(2014)	사용자 설문	협업 필터링(Collaborative filtering)	화장법 추천이 시각적 감성에 미치는 영향 분석
김성언 등(2017)	소셜 데이터	퍼지 추론과 감성 사전 구축	뷰티 제품별 추천점수 제공
Okuda <i>et al.</i> (2017)	소셜 데이터	협업 필터링(Collaborative filtering)	사용자 유사도 기반 추천 시스템 개발
이은주 등(2018)	사용자 설문	통계기반 클러스터링 기법	모바일 기반 뷰티 관련 서비스 개발
신혜란 등(2019)	소셜 데이터	회귀 분석, 의사 결정 트리, 나이브 베이즈 모델, K-Means 등	개인 맞춤형 추천 시스템 개발
윤여수, 박현준(2019)	사용자 사진/영상	이미지 인식 기술	사용자 피부 색상 기반한 화장품 색상 추천시스템 개발

추천해 주는 시스템을 개발하였다. 제품 성분 점수를 산출하고 고객의 댓글을 텍스트 마이닝 해서 감성 점수를 계산하였다. 또한, 두 데이터의 기준의 모호함을 해결하기 위해 퍼지 추론을 활용하였다. 퍼지 추론을 통해 제품별 추천 정도(Recommendation Degree)를 계산하는 알고리즘을 개발하였다. 알고리즘 성능 분석을 위해 실제 소포몰의 상품과 상품평을 수집해 시뮬레이션 결과를 제시하였다. Okuda *et al.*(2017)는 포털 사이트의 리뷰를 수집하여 뷰티 제품 추천시스템을 제안하였다. 본 시스템은 리뷰 데이터를 기반으로 사용자 군집 별 선호 뷰티 제품을 파악한 후, 비슷한 성향의 사용자에게 뷰티 제품을 추천한다. 즉, 사용자 기반 협업 필터링을 활용하여 사용자의 뷰티 제품 선호에 맞는 맞춤형 추천시스템을 제시하였다. 유사한 방식으로 신혜란 등(2019)도 소셜 미디어 데이터와 사용자 정보를 결합하여 개인 맞춤형 뷰티 제품 추천시스템을 개발하였다. 추천시스템에 활용된 알고리즘은 회귀 분석, 의사 결정 나무, 나이브 베이즈 모델, K-means 등이다. 이은주 등(2018)은 사용자의 피부 유형 및 유전자를 이용하여 맞춤형 뷰티 제품 추천 서비스를 제안하였다. 피부 유형 구분을 위해 모바일 기반 자가진단 프로그램을 활용하고 유전자 정보는 전문기관의 결과를 수집하여 개인별 뷰티 제품 추천 서비스를 개발하였다.

최근 연구 흐름은 영상/이미지 인식이나 딥러닝 기술을 활용한 뷰티 제품 추천시스템 개발로 한

차원 발전하고 있다. 윤여수, 박현준(2019)는 사용자의 얼굴이 포함된 영상을 수집하여 사용자의 피부색을 기계를 통해 정확하게 추출하고 이를 바탕으로 사용자에게 적합한 파운데이션 색상을 제안하는 시스템을 개발하였다. 제안된 시스템을 평가하기 위해 다양한 인종과 나이대별 사용자의 영상으로 실험한 결과 사용자에게 적합한 파운데이션 색상을 추천해 주는 것을 확인하였다.

선행연구를 분석한 결과, 기존 연구들은 대부분 설문데이터나 소셜 데이터를 기반으로 추천 알고리즘을 개발하는 것에만 초점을 두고 있다. 즉, 실제 사용자들의 검색 기록을 이용하여 사용자 클러스터 분류하고 추천한 연구는 부족하였다. 그리하여, 본 연구에서는 뷰티 매장에서 뷰티 제품 검색 데이터를 기반으로 뷰티 제품 취향 유형을 새롭게 분류하고, 이를 통해 뷰티 커머스 분야의 언택트 서비스에 활용 방안을 제시하고자 한다.

Ⅲ. 연구방법론

3.1 클러스터링 기법

클러스터링(군집화)은 일련의데이터 객체를 클러스터(군집)로 분류하는 과정으로, 유사한 데이터는 같은 클러스터 내에 있고, 유사하지 않은 데이터는 다른 클러스터에 속하도록 하는 머신러닝 기법이다(Han *et al.*, 2011; Huang, 1998; Jain *et al.*, 1999;

Jain and Dubes, 1988; Ji *et al.*, 2015). 대표적인 클러스터링 방법론으로 밀도 기반(density-based) 알고리즘, 분할적(partitional) 클러스터링을 뽑을 수 있다. 밀도 기반 알고리즘은 데이터가 모여있어서 밀도가 높은 부분을 군집으로 구성한다. 대표적인 밀도 기반 알고리즘으로 DBSCAN(Density-Based Spatial Clustering of Applications with Noise)가 있다(Ester *et al.*, 1996).

분할적 클러스터링 알고리즘은 비용함수를 최적화하면서 주어진 수만큼의 군집으로 나눈다(Ji *et al.*, 2015). 분할적 클러스터링 중에서 가장 널리 사용되는 클러스터링 기법은 MacQueen(1967)이 개발한 K-means이다. K-means 알고리즘은 각 군집의 중심과 데이터 객체와의 거리의 제곱 합을 비용함수로 하여 비용함수를 최소화하는 방향으로 소속 군집을 업데이트하는 방식으로 클러스터링하는 알고리즘이다. 비교적 간단하게 효율적으로 클러스터링할 수 있다는 장점이 있지만, 입력 데이터 측면에서 가지는 한계점이 있다. K-means는 기본적으로 연속형 변수를 입력값으로 갖는다. 따라서 연속형이 아닌 범주형 데이터의 경우 전처리를 하여 연속형으로 변환한 뒤 입력해야만 유클리디안 거리를 계산하여 군집화가 가능하다. 그러나 변수 대다수가 범주형이면 최적의 결과를 도출하기에 한계점이 있을 수 있다(Huang and Ng, 2003). 이에, Huang(1998)은 범주형 변수 데이터를 활용한 군집화에 최적화된 K-modes 기법을 제안하였다. K-means는 연속형 변수, K-modes는 범주형 변수에 각각 효율적인 클러스터링 알고리즘이다. 또한, Huang은 연속형, 범주형 변수가 모두 있는 데이터를 효율적으로 처리하기 위하여 K-means와 K-modes를 결합하여 K-prototypes 알고리즘을 제안했다(Huang, 1997). 본 연구에서 사용한 뷰티 추천 데이터의 경우에는 사용자의 성별, 피부 타입, 제품 성분, 제품 카테고리과 같은 범주형 변수가 다수 존재하며, 동시에 나이, 구매 횟수와 같은 연속형 변수도 함께 입력변수로 사용 되어야 한다. 따라서 본 연구에서는 범주형 변수와 연속형 변수가 함께 효율적으로 고려될 수 있는 K-prototypes 알고리즘을

사용하였다.

K-means 알고리즘은 식 (1)과 같이 연속형 변수 데이터를 군집화하는 방법으로 군집의 중심값은 군집의 평균값(mean)으로 정한다. 군집의 중심값과 객체 간의 거리 계산에는 유클리드 거리를 사용한다.

$$d_{k-means}(X, C) = d_{euclid}(X_i, C_j)^2 \quad (1)$$

$$= \sum_{l=1}^r |x_{il} - c_{jl}|^2$$

K-modes 알고리즘은 식 (2)와 같이 범주형 변수 데이터를 군집화하는 방법으로 군집의 중심값은 군집의 최빈값(modes)으로 정한다. 군집의 중심값과 객체 간의 거리 계산에는 객체와 중심값이 같으면 거리는 0, 같지 않으면 1을 주는 방식의 단순 매칭(simple matching)법을 사용한다.

$$d_{k-modes}(X, C) = d_{simple\ matching}(X_i, C_j) \quad (2)$$

$$= \sum_{l=1}^r \delta(x_{il}, c_{jl})$$

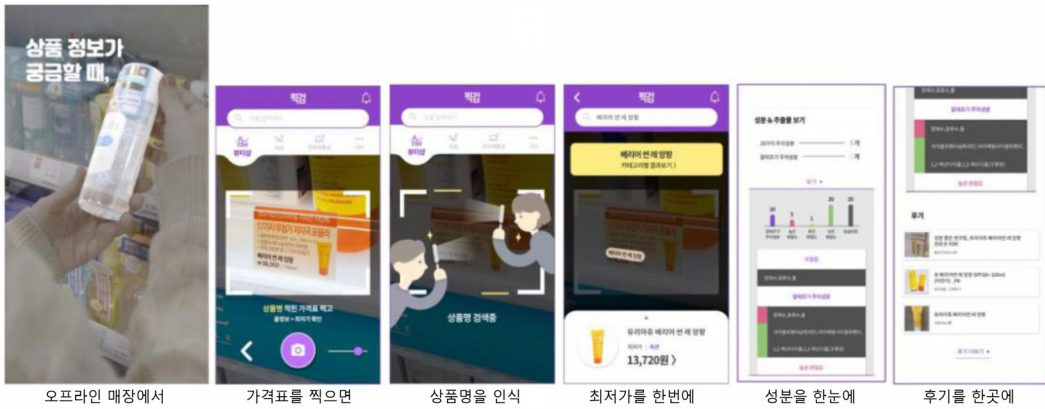
$$\delta(x_{il}, c_{jl}) = \begin{cases} 0, & \text{when}(x_{il} = c_{jl}) \\ 1, & \text{when}(x_{il} \neq c_{jl}) \end{cases} \quad (3)$$

K-prototypes 알고리즘은 연속형 변수와 범주형 변수가 혼합된 데이터를 클러스터링 하는 방법으로 연속형 변수에는 K-means를, 범주형 변수에는 K-modes를 각각 적용하여 결합한다. 군집의 중심값은 연속형 변수의 평균값과 범주형 변수의 최빈값으로 정한다. 군집의 중심값과 객체 사이의 거리를 범주형 변수 거리에 가중치를 부여하여 연속형 변수 거리와 합해서 구한다.

$$d_{k-prototypes} = d_{k-means} + \lambda d_{k-modes} \quad (4)$$

$$= d_{euclid}(X_i, C_j)^2 + \lambda d_{simple\ matching}(X_i, C_j)$$

본 연구는 연속형과 범주형 데이터가 혼합된 데이터의 특성에 가장 적합한 K-prototypes 알고리즘을 적용하여 사용자의 뷰티 제품 검색 데이터를



〈그림 1〉 찍검 소개

통해 사용자 뷰티 제품 선호를 분석하고 사용자 유형화하고자 한다.

3.2 데이터 수집

본 연구를 위해 뷰티 상품정보 검색서비스인 ‘찍검’⁶⁾의 협조를 받아 사용자 데이터와 사용자의 뷰티 제품 검색 로그 데이터를 사용하여 분석을 진행하였다. 찍검은 2018년 10월에 시작한 언택트 마케팅 서비스로, OCR 기술 기반으로 뷰티 제품명을 인식하여 제품 정보를 제공한다. 즉, 사용자가 뷰티 매장에서 관심있는 화장품이 있을 때 자신의 스마트폰 카메라로 찍으면 뷰티 제품 10개 카테고리별 맞춤 상세정보를 제공하는 서비스이다(〈그림 1〉 참조). 매장에서 직원의 도움 없이 혼자 쇼핑하는 것을 선호하는 이용자로부터 관심을 받고 있다. 2020년 4월 기준 찍검의 이용자는 4만 명이며, 회원가입 시 인구통계정보, 피부 고민, 피부타입 정보 등을 수집한다.

본 연구에서는 찍검 가입자 중 10,000명을 무작위 추출하여 사용자 정보를(비식별 코드, 성별/나이, 피부타입) 대상으로 분석을 하였다. 가입자 중 사용자 정보가 제대로 기재되지 않은 이용자, 직원

id, 아웃라이어를 제외하고 최종적으로 분석에 활용한 사용자는 총 9,263명이다. <표 2>는 수집된 이용자의 인구통계정보와 검색 건수를 요약한 자료이다. 사용자 중 남성 비중은 8.8%, 여성 비중은 91.2%로 여성 사용자의 비중이 높다. 연령대는 평균 20.18세이며 표준편차는 7.98로 10대와 20대 사용자의 비중이 높다(전체의 86.44%).

〈표 2〉 기술통계표

변수		빈도	비중
성별	남성	660	8.8%
	여성	8,603	91.2%
나이 (평균: 20.18, 표준편차: 7.98)	10~< 20	5,572	60.15%
	20~< 30	2,435	26.29%
	30~< 40	796	8.59%
	40~< 50	356	3.84%
	> = 50	104	1.12%
검색 건수 (평균: 2.13, 표준편차: 3.31)	1~< 2	6,839	73.83%
	3~< 4	1,390	15.01%
	5~< 10	779	8.41%
	10~< 30	225	2.43%
	> = 30	30	0.32%
합계		9,263	100.00%

분석에 사용된 검색 로그는 총 49,307건이었다. 검색 로그는 사용자가 검색한 뷰티 제품 이미지를

6) <https://play.google.com/store/apps/details?id=io.rocketview.lowcost&hl=ko>.

뷰티 카테고리별로 분류하였다. 분류 방식은 찍검에서 활용되고 있는 카테고리 방식을 활용하였다. 분류된 뷰티 카테고리는 스킨케어, 마스크팩, 베이스, 립아이, 썬케어, 클렌징, 바디, 헤어, 향수, 네일이다. 각 이용자별로 뷰티 카테고리를 검색한 횟수를 변수로 활용하였다.

3.3 연구설계 및 분석 절차

본 연구의 데이터 분석 절차는 <그림 2>와 같이 총 4단계로 진행되었다. 찍검의 사용자 데이터와 검색 데이터를 바탕으로 (1) 데이터 수집, (2) 데이터 전처리, (3) 군집 알고리즘 및 군집 K의 결정 그리고 마지막으로 (5) 고객 세그먼트 도출 및 언택트 마케팅 도출의 순서로 진행했다. (1) ‘데이터 수집’ 단계에서는 사용자 구분자(FCM_token)를 기준으로 뷰티 제품 검색 로그와 사용자 정보를 추출하였다. 분석에 활용된 변수는 총 15개로 스킨케어 검색 건수, 마스크팩 검색 건수, 베이스 검색 건수, 립아이 검색 건수, 썬케어 검색 건수, 클렌징 검색 건수, 바디 검색 건수, 헤어 검색 건수, 향수 검색 건수, 네일 검색 건수, 총 검색 횟수, 이용자 나이, 성별, 그리고 이용자 피부 타입이다.

(2) ‘데이터 전처리’ 단계에서는 총 4가지의 전처리를 수행하였다. 첫 번째로는 명목형 변수 데이터인 성별, 피부타입을 원핫인코딩(One-hot-encoding)을 통해 더미 변수로 변환하였고, 총 10개 뷰티 제품 카테고리별 검색 건수와 검색 로그 합계를 활용하였다. 또한, 뷰티 제품 탐색 이용자의 특성상 남성 이용자는 전체의 7%에 불과하며, 남성과 여성 제품에는 구분이 있는 경우가 많아서 남성 이용자들은 별도의 세그먼트로 분리한 뒤 나머지 8,603명의 여성 이용자를 대상으로 클러스터링을 진행하기로 하였다. 세 번째로 수치형 데이터는 데이터 범위를 맞추주기 위하여 모든 값을 0에서 1사이로 Min-Max 스케일링하였다.

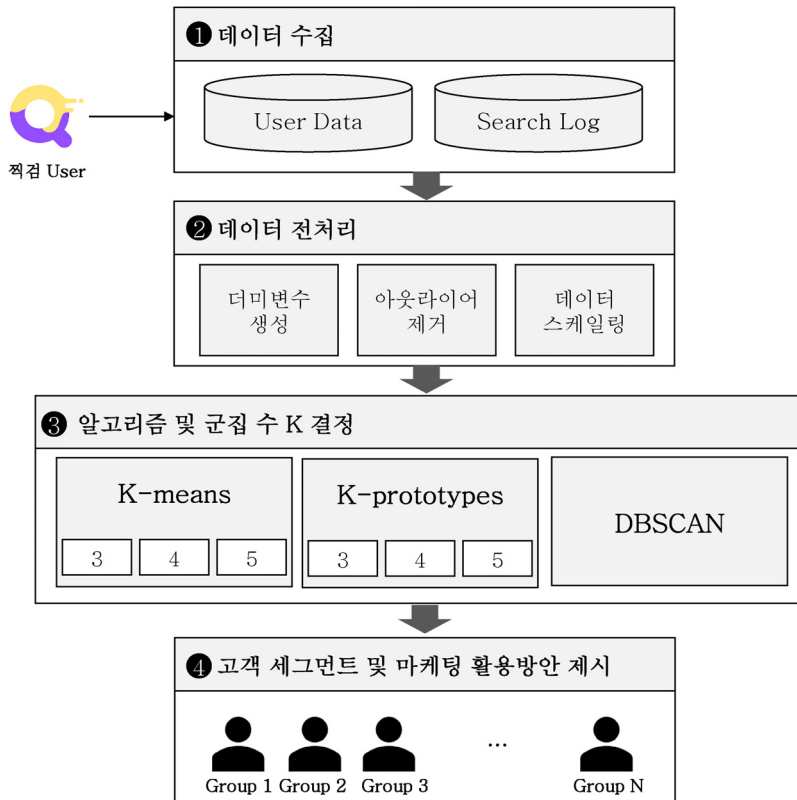
데이터 전처리 후에, (3) ‘알고리즘 및 군집 수 K 결정’ 단계에서 실루엣 평가를 하였다. 실루엣

평가는 그룹 내 데이터 거리와 다른 그룹 간의 거리를 계산하여 군집 분석이 잘 되었는지 비교하는 방법이다(Rousseeuw, 1987). 실루엣값은 -1에서 1사이의 값이 나오며, 1에 가까울수록 클러스터링 분석이 잘 된 것이다. 객관적 평가 기준으로 알고리즘을 평가하기 위해, 분할적 알고리즘인 K-means, K-prototype과 밀도 기반 알고리즘인 DBSCAN을 실루엣 계수로 비교하였다. 또한, 밀도 기반 알고리즘의 경우 군집 수 K값을 사전에 설정할 필요가 없지만, 분할적 알고리즘은 군집의 수 K값에 따라 군집의 결과가 다르므로, K값을 3부터 5까지 바꿔가면서 분석하였다. 알고리즘 수행은 파이썬(Python) 3.5와 파이썬 분석 패키지인 pypi.org와 scikit-learn을 이용하였다. 각 알고리즘은 10회 반복 수행하여 가장 좋은 결과를 선택하였다. 그 결과 K-prototypes 기법으로 군집의 수가 4개 일 때 가장 우수한 결과(0.506)가 나왔다. 각 알고리즘의 군집별 실루엣 계수는 <표 3>에 정리하였다.

<표 3> 알고리즘별 실루엣 계수 비교

군집의 수	분할적 알고리즘		밀도 기반 알고리즘
	K-means	K-prototype	DBSCAN
3	0.497	0.494	0.071
4	0.500	0.506	
5	0.426	0.464	

마지막으로 (4) ‘고객 세그먼트 및 언택트 마케팅 활용 방안’ 단계에서는 분석된 세그먼트들의 특성들을 파악해보고, 뷰티 분야의 언택트 마케팅 활용 방안을 제시하였다. 클러스터링 기법은 비지도 학습으로 평가와 검증이 어려운 분석 기법이다(Pfzner et al., 2009). 따라서 클러스터링 결과에는 사람의 평가가 필요하다(Feldman and Sanger, 2007). 해당 분야 전문가의 주관적 시각에서 클러스터링 결과의 평가가 필요하므로(Weiss, 2005) 뷰티 업계 종사자와 찍검 서비스 운영자로 구성된 3인의 전문



〈그림 2〉 연구 절차

가 그룹을 구성하여 클러스터링 결과를 확인하고, 가장 잘 군집화된 결과를 선정하였다. 클러스터링 결과 품질 평가에는 각 고객 세그먼트 별로 검색된 제품들의 검색패턴과 제품 선호 유사도가 기준이 되었다. 전문가 3인이 모두 동의하는 클러스터링 결과를 최종 결과로 삼아 각 세그먼트 별 탐색적 데이터를 분석하였다. 본 연구 결과를 기반으로 뷰티 분야의 제조사/유통사/플랫폼사 차원에서의 구체적인 활용 방안을 제시한다.

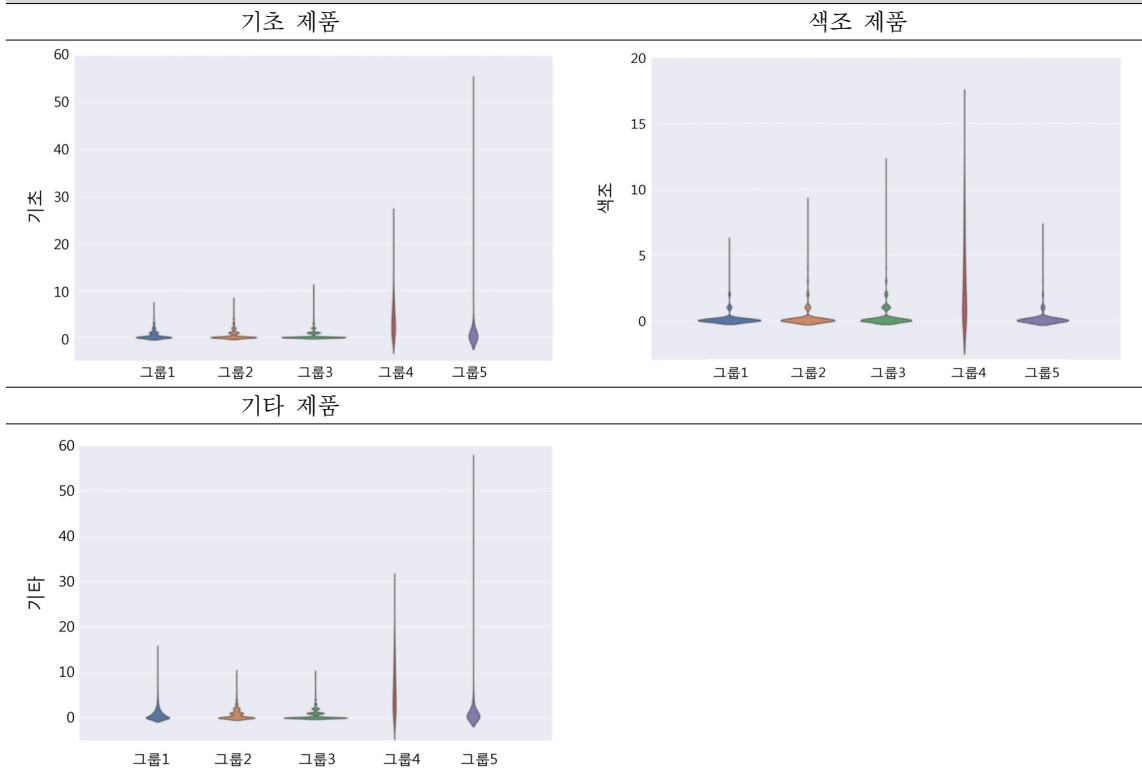
IV. 연구 결과

본 연구를 통해 도출된 총 5개의 고객 세그먼트는 <표 4>와 같다. 남성 고객을 별도로 분리하여 임의 배치한 세그먼트 그룹5 외에 나머지 세그먼트

트 그룹 1~그룹4는 K-prototypes 알고리즘을 통해 도출한 고객 세그먼트이다. 클러스터링 결과 해석을 통해 세부 세그먼트별 특성과 전체 세그먼트의 특성을 파악할 수 있다(박지혜, 2017). 전체 세그먼트에 걸친 공통된 특징은 제품군 중에서도 ‘기타’ 품목군(예: 클렌징, 썬케어, 헤어, 바디, 네일 등)에 대한 언택트 탐색이 색조(예: 립스틱, 아이섀도, 베이스 등), 기초(예: 스킨, 로션, 에센스 등)와 같은 다른 품목군과 달리 더 적극적이라는 것이다. 직접 발라보면 제품의 특징을 바로 알 수 있는 색조, 기초 제품과 달리 매장에서의 직접 체험이 어려운 기타 품목군에서는 언택트 탐색이 전반적으로 더 높았다. 이는 사용 후기, 성분정보 등을 통해 추가 정보가 필요한 품목군에서 언택트 탐색이 많다는 것을 말한다.

<표 4> 고객 세그먼트

세그먼트	그룹1	그룹2	그룹3	그룹4	그룹5	
주요 특징	기초 제품 주사용의 30대 여성	기초 제품 주사용의 20대 여성	전 제품 사용의 10대 여성	색조 제품 주사용의 10대 여성	기초 제품 주사용의 20대 남성	
비율	6.95%	24.32%	60.24%	1.36%	7.13%	
성별	여성	여성	여성	여성	남성	
언택트 서비스 검색 빈도 (mean= 2.18, std= 3.66)	1.82	1.90	1.84	20.80	2.92	
주요 연령대 (mean= 20.75, std= 8.48)	30대	20대	10대	10대	20대	
피부타입	지복합성	복합성	복합성	건복합성	건복합성	
주요 검색 및 구매 제품	기초화장품	32.43%	31.79%	29.13%	30.21%	45.53%
	색조화장품	10.24%	16.56%	28.35%	20.97%	7.00%
	기타	57.43%	51.65%	42.52%	48.82%	47.47%



기타 제품군은 모든 그룹에서 검색 및 구매 빈도가 높았으므로, 이를 제외하여 세그먼트 별 주요 특징을 나타내었다(<표 4> 참조). 그룹1은 언택트

서비스를 통해 기초화장품을 주로 검색하는 지복합성 피부의 30대 여성 그룹(6.95%)이다. 다른 그룹들과 비교하여, 그룹 1은 연령대가 가장 높고,

언택트 서비스 검색 빈도는 가장 낮다. 그룹 2는 언택트 서비스를 통해 기초화장품을 주로 검색하는 복합성 피부의 20대 여성 그룹(24.32%)이다. 그룹 2는 그룹 1과 함께 기초화장품을 주로 검색하는 그룹으로, 다른 그룹들에 비해 색조화장품의 검색 빈도가 낮다. 그룹 3은 언택트 서비스를 통해 기초화장품과 색조화장품을 다양하게 검색하는 복합성 피부의 10대 여성 그룹(60.24%)이다. 특히, 다른 그룹들에 비해 색조화장품의 검색 및 구매 비율이 높다. 그룹 4는 기초화장품과 색조화장품을 모두 검색하는 건복합성 피부의 10대 여성 그룹(1.36%)이다. 이들은 총 20.80회가량의 언택트 탐색을 오프라인 매장에서 하며, 이는 다른 그룹보다 월등히 높은 수치로 언택트 탐색에 가장 적극적인 그룹이다. 마지막으로, 그룹 5는 언택트 서비스를 통해 기초화장품을 주로 검색하는 건복합성 피부의 20대 남성 그룹(7.13%)이다. 그룹 5는 다른 그룹들에 비해, 색조화장품 검색 비율은 낮고 기초화장품 검색 비율은 높다.

V. 결과 토의 및 시사점

5.1 연구 결과 토의 및 활용 방안

본 연구는 최근 활성화되고 있는 뷰티 분야에서, 언택트 서비스 로그 데이터를 기반으로 고객 세그먼트를 새롭게 분류하고, 이를 기반으로 언택트 마케팅 전략 방안을 제안한다. 언택트 서비스의 모바

일 애플리케이션을 통해 수집한 고객 검색 및 로그 데이터를 머신러닝 기법의 K-prototypes 알고리즘을 이용하여 새롭게 고객 유형을 분류하고, 이를 뷰티 분야의 추천 모델로 활용할 수 있다. 기존 관련 연구들(Chung 2014; Okuda *et al.*, 2017; 김성연 등 2017; 신혜란 등 2019; 윤여수, 박현준 2019; 이은주 등 2018)은 설문 또는 소셜 데이터를 기반으로 추천 알고리즘 개발에 중점을 두었다. 즉, 언택트 서비스 환경에서 고객의 오프라인 행동 데이터 기반의 연구가 부족하였는데, 본 연구에서는 검색 로그 데이터를 기반으로 머신러닝 기법을 이용하여 고객 세그먼트를 새롭게 분류한 것에 차별성이 있다. 이를 기반으로 뷰티 분야의 제조사/유통사/플랫폼사 차원에서 본 연구에서 도출한 고객 세그먼트의 활용 방안을 <표 5>와 같이 제안한다.

첫째, 뷰티 제품 제조사 측면에서는 고객 세그먼트별 고객의 특징 및 취향을 기반으로 새로운 제품을 개발할 수 있다. 언택트서비스 환경에서 가장 많이 검색되는 제품군은 피부 타입에 크게 영향을 받지 않는 바디로션, 헤어 에센스, 핸드크림 등의 ‘기타’ 제품들이었다. 즉, 전문가의 조언이나 안내 없이 애플리케이션에서 추천해 주거나 고객 스스로 테스트해보고 쉽게 구매할 수 있는 ‘기타’ 제품군이 언택트 환경에서 가장 많이 판매되고 있었다. 그러므로, 언택트 서비스 환경에서는 ‘기타’ 제품군의 설명 또는 리뷰를 체계적으로 관리하여 해당 제품군의 판매를 늘릴 수 있다. 또한, 고객들의 구

<표 5> 제조사/유통사/플랫폼사 차원의 활용 방안 제안

구분	고객 세그먼트 활용 방안	
	활용 방안	내용
제조사	그룹 타겟팅 제품 개발	고객 세그먼트별 특징 및 취향을 기반으로 한 제품 개발
유통사	제품 진열 및 구성	고객 세그먼트별 판매 제품 진열 및 구성
	제품 프로모션 기획	고객 세그먼트별 제품 묶음 판매 및 할인판매 전략
플랫폼사	추천 알고리즘 개발	뷰티 제품 추천 알고리즘을 활용하여 고객들에게 최적의 뷰티 제품 추천
	온라인 맞춤형 제품 광고	고객 세그먼트별 맞춤형 제품 광고
	언택트 서비스 애플리케이션 구축	고객 세그먼트 기반의 애플리케이션 구축 및 서비스 향상

매 이력 및 취향을 기반으로 다양한 ‘기타’ 제품군의 뷰티 제품을 개발하여 판매할 수 있다.

둘째, 유통사 측면에서는 제품의 진열 및 구성뿐만 아니라 제품 프로모션 기획에 활용할 수 있다. 언택트 서비스의 경우, 오프라인 매장에서는 매장 내 제품을 통해서도 정보가 부족한 제품이 자주 검색된 것을 확인할 수 있었다. 그러므로, 매장 내 진열된 제품의 체험과 함께 각 제품에 대한 충분한 정보를 제공할 필요가 있다. 특히, 본 연구의 고객 세그먼트를 활용하여, 사용자가 속한 그룹의 사용 후기를 함께 제공할 수 있다. 온라인에서도 제품의 사용 후기를 충분히 보여주고, 구매하는 제품에는 샘플을 함께 보내어 미리 체험할 수 있도록 할 수 있다.

다음으로, 온라인 또는 오프라인 매장에서 판매하는 뷰티 제품을 진열하는데 고객 세그먼트별로 함께 자주 판매되는 제품을 고려하여 진열 및 배열할 수 있다. 온라인에서는 제품 배열을 가깝게 하는 것뿐만 아니라 특정 제품을 검색했을 때 같은 세그먼트에 속하는 소비자가 해당 제품과 같이 구매하는 제품을 화면에서 추천할 수 있다. 그 밖에, 세그먼트별로 함께 자주 판매되는 제품들을 묶어서 판매할 수도 있을 뿐만 아니라 특정 제품의 재고 문제가 발생할 때 해당 제품과 함께 자주 판매되는 제품을 묶어서 할인된 가격으로 판매할 수 있다.

셋째, 플랫폼사 측면에서는 본 연구를 통해 분류되는 고객 세그먼트를 기반으로 추천 알고리즘을 개발하여 언택트 서비스를 이용하여 뷰티 제품을 구매하고자 하는 고객들에게 최적의 뷰티 제품을 추천할 수 있다(김형수 등, 2019). 더불어, 각 그룹의 연령대, 피부타입, 언택트 서비스 검색 정도 등의 특징을 기반으로 만족할 수 있는 제품을 추천할 수 있다. 대표적으로, 고객 세그먼트의 그룹 3은 다른 그룹들에 비해 색조화장품의 검색 및 구매 빈도가 높은 10대 여성 그룹이다. 이들은 10대라는 나이 특성상 색조화장품에 대한 관심은 많지만, 사용 경험이 적을 가능성이 크므로, 자연스럽게 메이

크업할 수 있는 색조화장품을 추천할 수 있다. 더불어, 해당 화장품의 사용 방법과 이를 활용하여 메이크업할 수 있는 정보를 함께 제공할 수 있다.

또한, 온라인에서 타겟팅 광고에 활용할 뿐만 아니라 언택트 서비스의 애플리케이션 구축에 활용될 수 있다. 지금까지 온라인 뷰티 커머스 분야에서 제품 추천을 위한 추천 시스템과 모델 관련 연구들이 수행되었지만, 대부분이 설문조사 또는 소셜 데이터를 이용한 추천 알고리즘 개발로 한계점이 있었다. 본 연구에서는 실제 사용자의 검색 및 로그 데이터를 기반으로 고객 세그먼트를 분류하여 고객들이 만족을 느낄 수 있는 제품을 계속 추천할 수 있다. 다음으로 최근 온라인 커머스 이용률이 증가하고 있으므로, 본 연구에서 도출한 고객 세그먼트를 온라인 맞춤형 광고에 활용할 수 있다. 이는 온라인에서 특정 제품을 검색하는 소비자가 속하는 세그먼트의 특성 및 선호에 따른 관련 서비스 및 제품을 광고하는데 활용할 수 있다. 즉, 광고하고자 하는 제품 또는 서비스를 세그먼트의 특성을 기반으로 그룹별 맞춤 광고를 할 수 있다. 예를 들어, 피부 타입에 건성이 포함되고 주요 검색 제품이 기초 제품인 그룹에는 수분크림과 같은 수분 성분이 많이 포함된 기초 제품 위주로 광고할 수 있다.

마지막으로 언택트 서비스 애플리케이션의 구축 및 서비스 향상에 이바지할 수 있다. 본 연구에서는 언택트 서비스의 모바일 애플리케이션을 통해 고객 데이터를 수집하고, 이를 기반으로 새로운 고객 세그먼트를 분류하였다. 본 연구에서는 나이, 성별, 피부타입, 제품 검색 및 구매 이력 등 고객 정보뿐만 아니라 로그 데이터를 함께 분석한 것으로 다각적으로 고객 특성을 분석하여 세그먼트를 분류한 것이다. 그러므로, 이를 언택트 서비스 애플리케이션에 활용한다면 뷰티 전문가의 조언 없이, 제품을 직접 사용해보지 않더라도 고객들에게 만족할 수 있는 제품을 추천할 수 있는 언택트 서비스 애플리케이션을 구축하거나 서비스 향상에 이바지할 수 있다.

5.2 연구의 한계 및 향후 연구 방향

본 연구의 한계점과 이를 보완한 향후 연구 방향은 다음과 같다. 첫째, 본 연구에서는 성별과 특정 연령대에 데이터가 편중되었다는 한계점이 있다. ‘찍검’의 사용자 데이터 탐색 결과, 여성 이용자가 91%이고, 30대 미만이 86.9%로 편중되어 있었다. 이는 본 연구에서 사용한 데이터의 ‘찍검’ 서비스는 뷰티 분야의 모바일 애플리케이션으로 20-30대 여성 이용자가 대부분이라 이러한 편중 현상이 발생한 것으로 판단된다. 향후 연구에서는 연구 결과를 일반화하기 위하여 인구통계학적 특성을 고려하여 남성과 30대 이상 이용자의 데이터를 추가로 수집하여 연구를 진행할 수 있다.

둘째, 본 연구에서는 뷰티 분야의 언택트 서비스로 ‘찍검’의 모바일 애플리케이션을 기반으로 이용자 데이터를 수집하였다. 최근 언택트에 관한 관심이 커지면서, 뷰티 분야에서도 오프라인 및 온라인 기반의 다양한 언택트 서비스들이 활성화되고 있다. 다양한 언택트 서비스를 통해 관련 데이터를 수집할 수 있다면 추가적인 사용자 특성을 포함한 좀 더 세부적인 고객 세그먼트 도출이 가능할 것이다. 데이터 수집의 어려움 측면에서 본 연구의 기여가 크지만, 향후 다른 유형의 언택트 서비스를 통한 데이터 수집이 가능하다면 더 의미 있는 결과 도출이 가능할 것으로 판단된다.

셋째, 본 연구는 뷰티 분야에서 새롭게 고객 세그먼트를 분류하였지만, 제시한 고객 세그먼트를 실제 서비스에 적용하여 검증하지는 못하였다. 향후 연구에서는 본 연구를 통해 도출한 고객 세그먼트를 실제 뷰티 분야의 언택트 서비스에 적용하고, 이를 검증하는 연구가 필요하다. 대표적으로, 실제 서비스에 적용 전과 후의 상품 클릭 건수, 상품 만족도 등의 변화를 통해 본 연구의 결과 타당도를 높일 수 있다. 즉, 비교 분석 결과는 연구의 가치를 높이고 구체적인 언택트 마케팅 방안 설정에 도움을 줄 수 있을 것이다.

마지막으로, 본 연구에서는 언택트 서비스의 사

용 정도에 따른 고객 세그먼트는 도출하였지만, 언택트 서비스 활성화를 위한 직접적인 동기 요인은 파악하지 못하였다. 즉, 사용자들의 언택트 탐색 로그 데이터를 분석하였지만, 사용자들이 정말 어떤 의도와 동기 요인에 의해 언택트 탐색을 했는지는 파악하지는 못하였다. 향후 연구에서는 객관적 데이터로 파악하기 어려운 소비자의 주관적 인식 파악을 위한 설문 및 인터뷰를 통해 분석할 수 있다. 추가로, 언택트 탐색에 적극적인 사용자와 언택트 탐색에 일반적인 사용자를 비교한다면 언택트 탐색 동기 요인을 비교하여 그 차별성을 파악할 수 있을 것이다.

5.3 이론 및 실무적 시사점

본 연구는 언택트 서비스 검색 데이터를 바탕으로 고객 세그먼트를 도출하고 이를 활용할 방안을 제안하는 연구로 학술적 시사점은 다음과 같다. 첫째, 본 연구는 최근 활성화되고 있는 뷰티 분야에서 언택트 서비스의 고객 데이터를 기반으로 고객 세그먼트를 분류하고 이를 해석하였다. 매장 직원에게 느끼는 무언의 압박감에서 벗어나 편안한 쇼핑을 추구하는 소비자가 증가하고 있으나(Jaakkola *et al.*, 2015; Lee, 2019) 기존 연구에서는 오프라인 탐색 로그 데이터를 직접 활용한 연구가 부족했다. 온라인에서는 고객 이용 데이터를 기반으로 디바이스 유형을 고려한 멀티채널 마케팅 효과를 분석해 활용하는 등 다양한 분석과 활용이 활발하다(신하정, 남기환, 2018). 반면, 오프라인의 경우 오프라인 쇼핑의 특성상 고가의 고객 행동 감지 센서를 설치하거나 NFC 태그 등의 방법을 사용(최명희 등, 2013)하지 않는 이상 고객이 오프라인 소비 공간에서 어떤 제품을 구매 고려를 위해 탐색했는지 파악하는 것은 매우 어렵다. 본 연구에서는 실제 서비스 데이터를 활용하여 쉽게 구하기 어려운 오프라인 언택트 탐색 기록이 담긴 실제 로그 데이터를 바탕으로 분석을 하였다는 데 의미가 있다.

둘째, 고객 세그먼트 도출을 위하여 연속형과 범주형 데이터가 혼합된 데이터에 가장 적합한 K-prototypes 알고리즘을 활용하였다. 머신러닝의 대표적인 비지도 학습 기법인 클러스터링은 유사한 데이터끼리 같은 그룹에 속하게 묶는 기법으로, 마케팅, 사회과학, 심리학, 의료 등 다양한 분야에서 활용되고 있다. 클러스터링 기법은 고객 세그먼트 도출과 해석 및 활용 관련 연구에서도 널리 사용되는 기법이다(김담희, 안가경, 2018; 이영찬, 2018). 그러나 분석 대상 데이터의 성격에 따라 더 적합한 알고리즘이 존재하지만, 특히 사회과학, 마케팅 분야의 클러스터링 활용 연구에서 입력 데이터의 유형이 고려된 경우는 드물었다. 가장 널리 쓰이는 K-means 알고리즘은 기본적으로 연속형 변수만을 입력값으로 갖는다. 이에 대다수의 입력 데이터가 범주형이면 최적의 결과물 도출을 위해 제안된 알고리즘이 K-modes 알고리즘이며(Huang and Ng, 2003), 데이터 세트 내에 연속형과 범주형이 혼합하여 존재하는 경우에 K-means와 K-modes를 결합하여 적용하는 기법은 K-prototypes이다(Huang, 1997). 일반적인 실제 서비스 환경에서는 연속형과 범주형이 혼합된 경우가 대부분이다. 이에 본 연구에서 적용한 K-prototypes 클러스터링 알고리즘은 고객 세그먼트 도출 시 입력 데이터 유형에 따라 가장 적합한 클러스터링 알고리즘을 선택하는 데 도움을 줄 수 있다.

셋째, 본 연구는 최근 관심이 고조되고 있는 언택트 서비스 분야의 연구 중 하나로, 뷰티 분야뿐만 아니라 언택트 관련 연구 문헌 확장에 이바지한다. 기존에 비하면 마케팅의 개념으로 관련 연구들(Williams et al., 2011; 이무형, 2019)이 수행되었지만, 실제 소비자 데이터를 분석 및 활용한 연구는 드물었다. 또한, 최근 관심이 증가하고 있는 언택트 트렌드를 반영한 서비스의 검색 기록을 이용하여 전략을 제시한 연구는 부족하였다. 본 연구는 코로나 19로 인해 언택트 서비스가 빠르게 확산되고 있는 현 시점에서, 새롭게 개발 및 활용되고 있는 언택트 서비스의 소비자 검색 데이터를 이용하

여 고객 세그먼트 및 활용 방안을 제시하였다. 또한, 본 연구는 기존에 인문사회학에서 주로 다루던 소비문화 변화의 개념을 이해하고, 이를 머신러닝 기법과 결합한 융·복합적 연구로 볼 수 있다. 소비자의 소비문화 변화로 새롭게 나타난 언택트 소비 현상을 이해하며, 관련 연구의 문헌 확장에 이바지한다.

실무적 시사점으로는 본 연구를 통해 도출한 고객 세그먼트 도출 결과를 기반으로 제조사/유통사/플랫폼사 차원으로 각 이해관계자 차원의 구체적인 활용 방안을 제안하였다(<표 4> 참조). 본 연구는 뷰티 분야의 검색 데이터를 기반으로 고객 세그먼트를 도출했을 뿐 아니라 관련 산업 이해관계자들이 결과를 어떻게 활용 가능할지에 대해 세부적인 전략 방안을 제시하였다. 이를 통해 새롭게 떠오르고 있는 고객 행동 패턴인 언택트 트렌드에 발맞춰 고객 세그먼트를 새롭게 분류하고, 언택트 탐색 선호 고객군의 특성을 파악하여 향후 효율적인 마케팅 전략 수립에 활용할 수 있다. 특히, 직접 제품 테스트 및 체험을 통해 특성을 파악하기 어려운 섹터, 바디 제품 등의 제품군 탐색과 10대에서 언택트 탐색이 적극적인 것으로 밝혀졌으므로, 이를 언택트 서비스 전략 구축에 활용할 수 있다.

참고 문헌

- [1] 김난도, 김서영, 최지혜, 서유현, 이수진, *트렌드 코리아 2018*, 미래의 창, 2018.
- [2] 김담희, 안가경, “머신러닝을 이용한 고객세분화에 관한 연구”, *융복합지식학회논문지*, 제6권, 제2호, 2018, pp. 115-120.
- [3] 김성언, 김은경, 김용기, “퍼지 추론과 감성사전 구축을 통한 화장품 추천 시스템”, *한국지능시스템학회 논문지*, 제27권, 제3호, 2017, pp. 253-260.
- [4] 김형수, 박예린, 이정민, “머신러닝을 이용한 공연문화예술 개인화 장르 추천 시스템”, *Information Systems Review*, 제21권, 제4호, 2019,

- pp. 31-45
- [5] 박지혜, “초연결시대의 협력: IT 기업 간 협력 네트워크와 성과에 관한 연구”, *Information Systems Review*, 제19권, 제2호, 2017, pp. 21-35.
- [6] 신하정, 남기환, “다바이스 유형을 고려한 온라인 멀티 채널 마케팅 효과”, *Information Systems Review*, 제20권, 제4호, 2018, pp. 59-78.
- [7] 신혜란, 임유정, 홍유진, 임종태, 박재열, 이현병, 신보경, 복경수, 유재수, “소셜 미디어 환경에서 기계 학습을 활용한 화장품 추천 시스템의 설계 및 구현”, *한국콘텐츠학회 종합학술대회 논문집*, 2019, pp. 289-290.
- [8] 오수연, “언택트 마케팅 바람”, *마케팅*, 제52권, 제2호, 2018, pp. 60-64.
- [9] 윤여수, 박현준, “사용자의 피부 색상에 기반한 파운데이션 색상 자동추천”, *한국지능시스템학회 논문지*, 제29권, 제4호, 2019, pp. 280-284.
- [10] 이무형, “외식기업 비대면 채널 마케팅이 관계 효익과 재구매 의도에 미치는 영향”, *호텔리조트연구*, 제18권, 제1호, 2019, pp. 293-318.
- [11] 이영찬, “군집분석과 연관규칙을 활용한 고객 분류 및 장비구니 분석: 소매 유통 빅데이터를 중심으로”, *지식경영연구*, 제19권, 제4호, 2018, pp. 59-76.
- [12] 이은주, 송재오, 김이나, 유재수, “화장품 추천을 위한 개인의 피부 유형 및 유전자를 이용한 빅데이터 분석 기반 모바일 서비스”, *한국콘텐츠학회 종합학술대회 논문집*, 2018, pp. 495-496.
- [13] 조영성, 구미숙, 류근호, “RFM 기법과 k-means 기법을 이용한 개인화 추천시스템의 개발”, *한국컴퓨터정보학회논문지*, 제17권, 제6호, 2012, pp. 163-172.
- [14] 최명희, 전정호, 강희구, 이경진, “다양한 유틸리티 태그를 활용한 컨벤션 서비스 시스템 적용 사례: 2013 오송 화장품 · 뷰티 세계박람회를 중심으로”, *Information Systems Review*, 제15권, 제3호, 2013, pp. 111-128.
- [15] 트렌드 모니터, “비대면 서비스 관련 인식 조사”, *트렌드 모니터*, 2018. 7. 13, Available at <http://www.trendmonitor.co.kr/tmweb/trend/allTrend/detail.do?bIdx=1684&code=0201&trendType=CKOREA>.
- [16] Chung, K. Y., “Effect of facial makeup style recommendation on visual sensibility”, *Multimedia Tools and Applications*, Vol.71, No.2, 2014, pp. 843-853.
- [17] Ester, M., H. P. Kriegel, J. Sander, and X. Xu, “A density-based algorithm for discovering clusters in large spatial databases with noise”, *Kdd*, Vol.96, No.34, 1996, pp. 226-231.
- [18] Feldman, R. and J. Sanger, *The Text Mining Handbook: Advanced Approaches in Analyzing Unstructured Data*, Cambridge University Press, 2007.
- [19] Fitzsimmons, J. A., M. J. Fitzsimmons, and S. Bordoloi, *Service Management: Operations, Strategy, Information Technology*, McGraw-Hill New York, 2008.
- [20] Han, J., M. Kamber, and J. Pei, “Data mining concepts and techniques third edition”, *Morgan Kaufmann*, 2011.
- [21] Huang, Z., “Clustering large data sets with mixed numeric and categorical values”, *Proceedings Of 1st Pacific-Asia Conference on Knowledge Discovery And Data Mining*, 1997.
- [22] Huang, Z., “Extensions to the k-means algorithm for clustering large data sets with categorical values”, *Data Mining and Knowledge Discovery*, Vol.2, No.3, 1998, pp. 283-304.
- [23] Huang, Z. and M. K. Ng, “A note on k-modes clustering”, *Journal of Classification*, Vol.20, No.2, 2003, p. 257-261.
- [24] Jaakkola, E., A. Helkkula, L. Aarikka-Stenroos, and K. Verleye, “The co-creation experience from the customer perspective: Its measurement and determinants”, *Journal of Service Management*,

- Vol.26, No.2, 2015, pp. 321-342.
- [25] Jain, A. K. and R. C. Dubes, *Algorithms for Clustering Data*, Prentice-Hall, Inc, 1988.
- [26] Jain, A. K., M. N. Murty, and P. J. Flynn, "Data clustering: A review", *ACM computing surveys (CSUR)*, Vol.31, No.3, 1999, pp. 264-323.
- [27] Ji, J., W. Pang, Y. Zheng, Z. Wang, Z. Ma, and L. Zhang, "A novel cluster center initialization method for the k-prototypes algorithms using centrality and distance", *Applied Mathematics & Information Sciences*, Vol.9, No.6, 2015, pp. 2933-2942.
- [28] Lee, D., "Effects of key value co-creation elements in the healthcare system: Focusing on technology applications", *Service Business*, Vol.13, No.2, 2019, pp. 389-417.
- [29] Lee, S. M. and D. Lee, "Untact: A new customer service strategy in the digital age", *Service Business*, Vol.14, No.1, 2020, pp. 1-22.
- [30] MacQueen, J., "Some methods for classification and analysis of multivariate observations", *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, Vol.1, No.14, 1967, pp. 281-297.
- [31] Meuter, M. L., A. L. Ostrom, R. I. Roundtree, and M. J. Bitner, "Self-service technologies: Understanding customer satisfaction with technology-based service encounters", *Journal of Marketing*, Vol.64, No.3, 2000, pp. 50-64.
- [32] Okuda, A., Y. Matsunami, M. Ueda, and S. Nakajima, "Finding similar users based on their preferences against cosmetic item clusters", *Proceedings of the 19th International Conference on Information Integration and Web-based Applications & Services*, 2017, pp. 156-160.
- [33] Pfitzner, D., R. Leibbrandt, and D. Powers, "Characterization and evaluation of similarity measures for pairs of clusterings", *Knowledge and Information Systems*, Vol.19, No.3, 2009, pp. 361-394.
- [34] Rousseeuw, P. J., "Silhouettes: A graphical aid to the interpretation and validation of cluster analysis", *Journal of Computational and Applied Mathematics*, Vol.20, 1987, pp. 53-65.
- [35] Weiss, S. M., *Text Mining: Predictive Methods for Analyzing Unstructured Information*, Springer, 2005.
- [36] Williams, R., K. Ribisl, and C. Jo, *Response to Advance Notice of Proposed Rulemaking on Non-Face to-Face Sale and Distribution of Tobacco Products and Advertising, Promotion, and Marketing of Tobacco Products*, Chapel Hill, NC: University of North Carolina at Chapel Hill, 2011.

A Study of the Beauty Commerce Customer Segment Classification and Application based on Machine Learning: Focusing on Untact Service

Sang-Hyeak Yoon* · Yoon-Jin Choi** · So-Hyun Lee*** · Hee-Woong Kim****

Abstract

As population and generation structures change, more and more customers tend to avoid facing relation due to the development of information technology and spread of smart phones. This phenomenon consists with efficiency and immediacy, which are the consumption patterns of modern customers who are used to information technology, so offline network-oriented distribution companies actively try to switch their sales and services to untact patterns. Recently, untact services are boosted in various fields, but beauty products are not easy to be recommended through untact services due to many options depending on skin types and conditions. There have been many studies on recommendations and development of recommendation systems in the online beauty field, but most of them are the ones that develop recommendation algorithm using survey or social data. In other words, there were not enough studies that classify segments based on user information such as skin types and product preference. Therefore, this study classifies customer segments using machine learning technique K-prototypes algorithm based on customer information and search log data of mobile application, which is one of untact services in the beauty field, based on which, untact marketing strategy is suggested. This study expands the scope of the previous literature by classifying customer segments using the machine learning technique. This study is practically meaningful in that it classifies customer segments by reflecting new consumption trend of untact service, and based on this, it suggests a specific plan that can be used in untact services of the beauty field.

Keywords: *Beauty, Untact, K-prototype, Customer segment, Machine learning*

* Doctor, Graduate School of Information, Yonsei University / Manger, SmartmediaRep

** Research Professor, Graduate School of Information, Yonsei University

*** Doctor's Student, Graduate School of Information, Yonsei University

**** Corresponding Author, Professor, Graduate School of Information, Yonsei University

○ 저자 소개 ○



윤상혁 (scottyoon@yonsei.ac.kr)

연세대학교 정보대학원에서 정보시스템 박사학위를 취득한 후, 현재 스마트 미디어랩(SMR)에서 데이터 분석 업무를 맡고 있다. 주요 연구분야는 머신러닝, 디지털 마케팅, 비즈니스애널리틱스 등이다. 관련 연구들은 Electronic Commerce Research, IVEY Publishing 등에 논문이 게재되었다.



최윤진 (yoonjin_choi@yonsei.ac.kr)

연세대학교 정보대학원 정보시스템 박사과정 재학중이다. TBWA KOREA에서 AE로, 주식회사 몬데이터의 창업자이자 대표이사로 근무했다. 주요 연구분야는 Technology, Media & Telecommunication(TMT)으로 데이터분석, 머신러닝, 딥러닝, 디지털 비즈니스, 디지털 미디어, 비즈니스 인텔리전스 등이다.



이소현 (dlthgus1010@gmail.com)

연세대학교 정보대학원에서 정보시스템 박사학위를 취득한 후, 현재 연세대학교 정보대학원 연구교수로 재직 중이며, 중국 시안교통대학교 임용예정이다. 주요 연구분야는 디지털 비즈니스, 소셜미디어 등이다. 관련 연구들은 Information Systems Research, Communications of the ACM, Information & Management, Internet Research, International Journal of Information Management 등에 논문이 게재되었다.



김희웅 (kimhw@yonsei.ac.kr)

National University of Singapore 정보시스템학과에서 근무한 후, 현재 연세대학교 정보대학원 교수로 재직 중이다. 주요 연구분야는 디지털 비즈니스, 정보시스템 관리 및 활용 등이다. 관련 연구들은 MIS Quarterly, Information Systems Research, Journal of Management Information Systems 등에 60여 편의 논문이 게재되었다. JAIS, IEEE TEM의 편집위원으로 활동했고, KrAIS 회장을 역임했다.

논문접수일 : 2020년 06월 01일

게재확정일 : 2020년 09월 03일

1차 수정일 : 2020년 08월 14일