

# 고객의 특성 정보를 활용한 화장품 추천시스템 개발

## Beauty Product Recommendation System using Customer Attributes Information

김효중 (Hyojoong Kim)	연세대학교 정보대학원 석사과정
신우식 (Woosik Shin)	연세대학교 정보대학원 박사과정
신동훈 (Donghoon Shin)	연세대학교 정보대학원 석사과정
김희웅 (Hee-Woong Kim)	연세대학교 정보대학원 교수, 교신저자
김화경 (Hwakyung Kim)	(주)로켓뷰 CEO

### 요약

인공지능 기술이 발달함에 따라 빅데이터를 활용한 개인화 추천시스템에 대한 관심이 증가하고 있다. 특히 뷰티 제품의 경우 개인의 취향과 더불어 피부 특성 및 민감도에 따라 제품 선호도가 명확히 구분되므로 축적된 고객 데이터를 활용하여 고객 맞춤형 추천서비스를 제공하는 것이 필요하다. 따라서 본 연구에서는 딥러닝 기법을 활용하여 제품 검색 기록과 개인 사용자의 피부 타입과 고민 등의 콘텍스트 정보를 함께 반영한 심층 신경망 기반의 추천시스템 모델을 제시하고자 한다. 본 연구에서는 실제 화장품 검색 플랫폼의 데이터를 활용하여 성능 평가를 실시하였다. 본 연구의 실험 결과, 고객의 콘텍스트 정보를 포함한 모델이 제품 검색 기록만을 활용한 기준의 협업 필터링 모델들 보다 우수한 성능을 보임을 확인하였다.

**키워드 :** 개인화 추천시스템, 신경망 모델, 잠재 선호도, 고객 특성 정보, 뷰티 제품, 검색 로그

### I. 서론

인터넷 보급의 확대와 정보통신기술(Information and Communication Technology)의 발달로 전자상거래(e-commerce)를 통한 온라인 쇼핑이 보편화 되었고 이에 따라 온라인 쇼핑 시장에 유통되는 제품의 다양성도 커지고 있다. 전자상거래에서 제품 종류의 증가는 서로 다른 소비자들의 기호를

충족시켜주어 고객들의 편의성을 높이는 반면에 특정 기호가 형성되지 않은 소비자들에게 제품의 다양성은 구매 의사를 결정하는데 비용을 증가시키는 문제를 발생시키고 있다. 즉, 선택의 복잡성에 직면한 소비자들이 정보 과부하로 자신에게 맞는 제품을 선택하지 못한 채 오로지 제품 정보 수집에 몰두하여 구매로 이어지지 않는 문제가 발생하고 있다(Jannach *et al.*, 2010; Persson, 2018). 여러 제품을 검색하는 것이 용이한 전자상거래의 특성상 이러한 정보 과부하 문제는 오프라인 매장 이용 시 보다 쉽게 발생하고 이는 매출에 직접적으로

† 본 연구는 중소벤처기업부 중소기업기술정보진흥원이 지원하는 빅데이터 기반 서비스 개발사업의 지원을 받아 진행되었음(G21S293578601).

영향을 미친다. 따라서 온라인 쇼핑 플랫폼에 있어 고객들의 정보 검색 비용을 낮추고 구매 의사 결정을 돋는 개인 맞춤형 서비스 제공에 대한 필요와 중요성이 급증하고 있다(송희석, 2019). 그 중에서 축적된 사용자들의 정보 및 행동 데이터를 활용하여 고객이 선호할 만한 상품을 추천하여 고객들의 탐색 비용을 줄이고 구매 의사 결정을 촉진시키는 개인화 추천 시스템이 주목을 받고 있다.

추천 시스템의 경우 선호하지 않거나 적합하지 않은 제품을 추천하는 경우와 같이 추천 성능이 좋지 못하면 오히려 구매 의사로 감소시키고 고객의 불만 요소로 확대될 수 있는 리스크를 가지고 있다(Kim and Kim, 2016). 특히, 뷰티 제품의 경우 개인의 취향과 더불어, 소비자의 피부 특성 및 민감도에 따라 특정 성분이 들어간 제품에 대한 선호도가 명확히 구분되므로 축적된 고객 정보를 융합하여 개인 맞춤형 추천을 하는 것이 추천 시스템의 성능 및 만족도에 크게 영향을 미칠 수 있다. 하지만 현재까지 뷰티 제품 관련 추천 시스템 연구들은 고객의 특성을 반영하지 않은 채 고객의 구매 및 선호 제품을 토대로 추천 시스템 모형을 제시하거나(Gholamian *et al.*, 2011), 피부특성과 같은 고객의 콘텍스트(context) 정보를 활용한 경우에는 고객들의 구매 기록을 반영하지 않고 피부타입에 따라 고객을 구분하여 화장품의 성분을 기반으로 고객이 속한 피부타입그룹에 적합한 제품을 매칭하는 것에 그치는 한계가 있다(Patty *et al.*, 2018; Yim *et al.*, 2016). 이처럼 현재까지 고객의 콘텍스트 정보와 구매 및 선호 기록을 포괄적으로 활용한 연구는 찾기 어렵다. 방법론측면에서도 기존 연구들은 사용자와 아이템 간의 유사도를 산출하는 협업 필터링 모형(Collaborative Filtering; CF)을 활용하였으나 이 기법은 신규 고객 또는 신규 아이템이 등장했을 때 추천이 어려운 콜드 스타트(cold start)가 발생하는 문제가 있다.

따라서 본 연구에서는 심층 신경망 협업 필터링 모형(Neural collaborative filtering; NCF)을 기반으로 제품 검색 기록과 더불어 사용자의 피부 특성,

성별, 연령 등의 콘텍스트 정보를 포괄적으로 반영한 심층 신경망 추천시스템모형(NCF with context information; NCF-Cl)을 제시하고자 한다. 콘텍스트 정보를 활용하여 신규 고객에 대해서도 기존 고객과의 콘텍스트 정보 유사도를 바탕으로 유의한 추천이 가능한 모델을 제시하여 기존 협업 필터링의 콜드 스타트 문제를 보완하고자 한다. 본 연구의 추천모형을 검증하기 위해 콘텍스트 정보를 제외한 심층 신경망 협업 필터링 모형과 일반화 행렬 분해를 활용한 기존 협업 필터링 모형을 활용하여 성능 비교를 시행하였다.

## II. 문헌 연구

### 2.1 뷰티 산업 동향 연구

현대 사회에서 화장(Makeup)은 개인의 개성을 표현하는 하나의 중요한 수단이며 이를 통해 자신만의 이상적 외모 추구하고 자기만족(Self-Satisfaction)을 이루는 것이 중요한 가치로 여겨지고 있다(구혜경, 나종연, 2015). 오늘날 화장에 대한 인식 변화와 소득 및 소비 수준의 향상으로 기존 성인 여성 위주의 화장품시장이 청소년 그리고 성인 남성으로도 확장되고 있다. 소비자들의 개성 표현 욕구와 소비계층의 확대는 화장품 구매로 이어지고 있으며 이에 따라 국내 화장품 시장은 지속적으로 성장하고 있는 추세이다. 신유원, 신민선(2019)에 따르면 국내 화장품산업의 총생산규모는 2018년 약 16조 원으로 전년대비 14.7% 증가하였고 2014년 이후 10% 이상의 연평균 성장률을 보이고 있다. 국내 시장의 성장과 더불어 국내 화장품 브랜드의 해외 수출도 지속적으로 증가하고 있다. 특히 동남아시아 국가들의 경우 현지 브랜드의 부재와 K팝을 비롯한 한국 방송콘텐츠에 대한 관심도가 높아지면서 한국 화장품에 대한 선호도가 증가하고 있다.

한국무역통계진흥원에 따르면 2018년 화장품 수출액은 약 63억 달러로 2017년 대비 26.5% 증가

한 것으로 조사되었다. 하지만 이러한 성장과 동시에 여러 새로운 브랜드들이 국내 화장품 시장에 출시되면서 화장품 산업 내 브랜드 간 경쟁도 치열해지고 있는 실정이다. 예를 들면, 식품안전처 화장품제조판매업에 등록된 화장품 브랜드의 수는 2015년 6,442개에서 2019년 17,260개로 4년 사이에 국내 화장품 브랜드 수만 1만여 개 증가하였다. 브랜드 간 경쟁이 불가피한 상황에서 여러 화장품 제조업체들은 고객들에게 맞춤형

서비스를 제공하여 고객의 만족도를 높이고 고객과의 장기적 관계를 형성하려는 노력을 하고 있다. 이런 맥락에서 개인화된 추천서비스는 브랜드의 경쟁력 강화에 매우 중요함에도 불구하고 화해, 찍첨 등 국내 뷰티 제품 관련 플랫폼들의 제품추천 기술 현황을 살펴보면, 대부분 제품의 성분 정보를 제공하거나 사용자들의 리뷰를 제공하고 피부 타입에 따라 군집화하여 일괄적으로 제품을 추천하는 수준이며 개인화된 추천 시스템을 제공하는 플랫폼은 찾아보기 힘들다. 따라서 고객의 콘텐츠 정보와 고객의 활동 기록 정보를 모두 반영한 개인화 추천 시스템 개발이 필요한 실정이다.

## 2.2 추천 시스템 연구

추천 시스템은 사용자의 특성과 선호도를 고려하기 위해 크게 내용 기반 추천(Content-based Recommendation)과 협업 필터링(CF)으로 발전되어 왔다. 먼저, 내용 기반 추천 시스템은 컨텐츠가 비슷한 아이템을 추천하는 방식으로, 사용자가 과거 경험했던 아이템 중 비슷한 속성의 아이템을 분석하여 유사 아이템을 제시한다. 특히, 비정형 텍스트 데이터를 기반으로 Term Frequency - Inverse Document Frequency(TF-IDF)를 활용하거나 아이템들의 특성을 기반으로 군집화하여 추천하는 방식이 주로 활용되어 왔다(Bobadilla *et al.*, 2013; Schafer *et al.*, 2007). 하지만, 과거 기록이 없는 새로운 사용자에 대한 추천이 어렵고 선호하는 항목에 대해 반복적으로 추천하는 한계점이 나타나며

협업 필터링 기반의 추천 시스템이 활발하게 연구되고 있다. 협업 필터링은 사용자-아이템 행렬(user-item matrix)을 통해 산출한 사용자 및 제품 간의 유사도를 활용하는 방식이다. 이처럼 다수의 사용자 데이터를 기반으로 특정 사용자와 유사한 선호 패턴을 가지는 다른 사용자들의 정보를 활용하여 내용 기반 추천의 한계를 보완하며, 학계뿐만 아니라 산업계에서도 활발하게 활용되어 오고 있다(Cremonesi *et al.*, 2012).

하지만 전통적인 메모리 기반(Memory-based) 협업 필터링은 아이템의 종류와 사용자의 수가 급증함에 따라 유사도 계산을 위한 알고리즘 구현에 더욱 많은 계산량이 요구되는 한계점이 존재한다. 특히, 새로운 사용자나 아이템에 대한 과거 기록이 없으면 선호 예측 및 제품 추천이 어려운 콜드 스타트 문제가 제기되었다(Su and Khoshgoftaar, 2009). 콜드 스타트 문제는 데이터 희소성과 관련이 있으며 콜드 스타트 상황에서 사용자가 만족할 수 있는 추천을 제공하는 것이 추천시스템의 핵심과제로 떠올랐다(Park and Chu, 2009; Zhang *et al.*, 2010).

이러한 콜드 스타트 문제를 해소하기 위해 새롭게 제안된 방식은 모델 기반(Model-based) 협업 필터링으로서, 대표적으로 행렬 분해법(Matrix Factorization)과 딥러닝 기반 협업 필터링이 높은 성능을 나타내며 주목받아 왔다. 행렬 분해법은 메모리 기반 협업 필터링에서 활용되었던 사용자-아이템 행렬을 사용자 잠재 벡터와 아이템 잠재 벡터로 분해하여 손실함수를 통해 최적화되어 새롭게 산출된 행렬을 기반으로 아이템을 추천한다(Koren *et al.*, 2009). 이와 더불어 행렬 분해법이 지닌 강점을 인공신경망 형태로 구현된 대표적인 딥러닝 기반 협업 필터링 방식인 심층 신경망 협업 필터링 모형 또한 새롭게 제안되었다(Gao *et al.*, 2019; He *et al.*, 2018). 심층 신경망 협업 필터링 모형은 원핫인코딩(One-hot Encoding)으로 구성된 사용자와 아이템 데이터를 임베딩을 통해 사용자 잠재 벡터와 아이템 잠재 벡터로 표현한 뒤, 다층 신경망(Multi-Layer Perceptron)에 통과시켜 사용

자와 아이템 간의 비선형적 상관관계를 학습하여 기존의 메모리 기반 협업 필터링의 문제점을 보완하는 데 기여하였다.

### 2.2.1 신경망 기반의 협업 필터링

본 연구는 모델 기반의 협업 필터링 중 신경망 기반의 협업 필터링 모형을 기반으로 설계되었다. He *et al.*(2017)은 기존의 일반화 행렬 분해(Generalized Matrix Factorization, GMF) 모형의 선형성 한계를 지적하고, 이를 딥러닝 기반의 협업 필터링 알고리즘을 통해 극복하고자 하였다. 즉, 기존의 일반화 행렬 분해 모형에서 다층신경망 모형을 결합한 심층 신경망 협업 필터링 모형을 개발한 것이다. 이는 고객의 잠재요인을 선택 여부에 따른 이항 출력값(Binary output)을 기반으로 모형화한 것으로 구매 가능성 혹은 예상 선호도가 높게 산출된 항목을 추천하는 목적으로 개발된 방법이다(이주오 등, 2020).

He *et al.*(2017)이 제안한 심층 신경망 협업 필터링 모형은 입력층(Input layer)에서 고객과 상품의 잠재 요인을 포함하는 임베딩층(embedding layer), 일반화 행렬 분해와 다층 신경망 모형을 포함하는 은닉층(Hidden layer), 그리고 고객의 상품 예상 선호도를 산출하는 출력층(Output layer)으로 구성되어 있다. 은닉층에는 기존의 일반화 행렬 분해 모형이 가진 선형성의 한계를 극복할 수 있도록 Sigmoid 등의 활성화 함수를 사용하는 것이 특징이다. 최종적으로 일반화 행렬 분해의 출력 결과와 고객 및 제품의 잠재요인 임베딩층의 출력 결과를 결합(Concatenate)하고 다수의 FC Layer를 거쳐 출력층으로 이어지는 구조로 설계되었다.

심층 신경망 협업 필터링 모형의 학습은 확률적 경사 하강법을 통해 이루어진다. 이는 출력값과 실제값의 비교를 통해 로그 손실(log loss) 목적함수를 최소화하도록 진행된다. He *et al.*(2017)가 개발한 심층 신경망 협업 필터링 모델은 다른 비교 모델(Baseline Model)에 비해 성능이 탁월하게 우수하였다. 이는 고객의 제품 선호에 대한 잠재요

인을 비선형 신경망 모델링을 통해 기존의 일반화 행렬 분해의 단점을 보완하였다는 점에 기인한다고 볼 수 있다.

### 2.2.2 고객의 특성 정보를 반영하는 협업 필터링

Ricci *et al.*(2011)은 사용자와 아이템의 단순 상호관계 파악뿐만 아니라 고객의 특성 정보도 포함하는 포괄적인 추천 모델을 제안하였다. 여기서 의미하는 고객 특성 정보에는 고객의 성별, 나이 등 정적인(Static) 요소부터 제품 구매 주기, 앱 접속 빈도 등 활동적인(Dynamic) 요소까지 포괄하는 개념이다.

Ricci *et al.*(2011)가 제안한 모델은 기존의 협업 필터링 모형에서 N개의 특성 정보(Contextual Information) 만큼 차원을 확장한 아키텍처로 볼 수 있다. 즉, 사용자와 아이템의 2차원 행렬에서 특성 정보의 개수만큼 차원을 확장하여, 입력층에 투입하게 되는 것이다. 고객 정보의 반영은 고객의 제품 선호도를 정교하게 반영할 수 있도록 하여 추천시스템 모델의 성능 향상에 기여할 수 있다.

Covington *et al.*(2016)은 유튜브 방송 클립을 고객에게 추천하는 모델을 개발하는 과정에서, 고객의 특성 정보를 반영하는 것이 모델의 성능 향상 측면에서 상당한 기여 한다는 것에 대해 실증적으로 입증하였다. 본 연구에서 제안하는 아키텍처는 고객의 제품 선호도에 영향을 미칠 수 있는 고객의 특성 정보가 입력층에서 받을 수 있도록 설계되었다.

## 2.3 뷰티 제품 관련 추천 시스템 연구

뷰티 제품에 대한 사용자 데이터가 누적됨에 따라 뷰티 제품에 대한 추천시스템 연구는 활발하게 진행되어왔다. 국외의 경우 사용자 리뷰를 활용하거나 제품 카테고리의 특성에 대한 TF-IDF를 활용하여 제품의 속성을 추출하여 추천 시스템에 적용한 연구가 진행되었다. Matsunami *et al.*(2017)

은 뷰티 제품에 대한 사용자의 리뷰를 기반으로 한 사용자 유사도 측정 방법을 제시하였다. 사용자 리뷰 데이터를 기반으로 동시발생 키워드 사전을 활용한 자동 스코어링 방법을 구축하였다. 또한 산출된 평점으로 뷰티 제품을 군집화하고 사용자 유사도를 계산하였다. 이처럼 리뷰 데이터를 기반으로 데이터의 희소성 문제를 해소하며 유사한 사용자에 따라 제품을 추천하는 방식의 추천시스템을 제안하였다. Patty *et al.*(2018)은 40개의 제품 카테고리를 기반으로 TF-IDF를 활용한 컨텐츠 기반 추천시스템을 제안하였다. TF-IDF 분석을 위해 제품의 컨텐츠를 화장품 태입, 사용시기, 피부타입 등의 정보로 구성하였다. 그리고 앞서 분류한 40개의 제품 카테고리별과의 코사인 유사도를 산출하여 제품군을 추천하는 방식의 추천 시스템을 구축하였다. Yoon and Joung(2020)은 뷰티 제품의 속성과 평가 및 후기 데이터를 추출하기 위해 웹크롤링을 통해 텍스트 데이터를 수집하였고 TF-IDF 방식을 활용하여 제품 간의 유사도를 판별하는 추천 시스템을 구축하였다.

국내에서도 뷰티 제품에 대한 추천 시스템 연구가 활발하게 진행되었다. Yim *et al.*(2016)은 OCR(Optical Character Recognition)을 활용하여 각 뷰티 제품의 성분 정보를 추출하였고 사용자의 선호도 정보와의 자카드 유사도를 계산하여 추천하는 방식의 연

구를 진행하였다. 또한 K-최근접 이웃 알고리즘을 기반으로 특정 제품의 성분과 유사한 성분을 함유한 유사한 제품을 분류하여 추천하는 방식을 활용하였다. 또한, 윤여수, 박현준(2019)은 딥러닝 기술을 활용하여 사용자의 얼굴 이미지를 인식하고자 하였다. 사용자의 피부 색상에 대한 특성을 추출하고 이에 맞는 파운데이션 제품을 자동 추천하는 알고리즘을 구축하여 뷰티 제품 추천을 위한 데이터 범위를 확대하는데 기여하였다.

뷰티 제품의 특성상, 피부는 개인마다 상태의 차이가 있으며 피부 고민 또한 상이하기 때문에 선호도를 비롯한 사용자의 특성 정보를 반영해야 한다(이은주 등, 2018). 선행 연구들은 리뷰 데이터, 제품 성분 정보, 피부 색상 등을 활용하여 데이터 희소성 문제를 해소하며 다양한 정보를 반영한 추천 시스템 개발을 시도하였다. 하지만 선행연구 중 일부는, 제품의 성분이 유사하거나 사용자가 선호해왔던 제품과 유사한 제품을 매칭하는 컨텐츠 혹은 최근접 이웃 기반 추천시스템 방식을 활용하였다. 따라서 사용자별 피부 타입, 피부 고민, 성별, 연령 등의 고객 콘텍스트 정보를 반영하지 못하고 사용자가 선호하는 특성을 가진 항목을 반복적으로 추천하는 한계점이 존재한다. 또한, 피부 타입, 사용자 리뷰와 같은 고객의 특성 정보를 반영하기 위한 다양한 시도도 있었지만, 주로 활

〈표 1〉 뷰티제품 관련 추천 시스템 선행연구

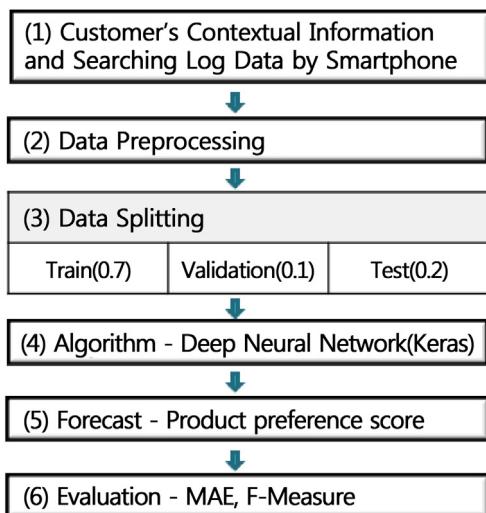
References	Method	Context, Feature and Key Contribution
Matsunami <i>et al.</i> (2017)	Correlation coefficient (Similarity)	리뷰 데이터를 기반으로 데이터의 희소성 문제를 해소하며 사용자 간 유사도 기반 추천 시스템 개발
Patty <i>et al.</i> (2018)	Content based Filtering with TF-IDF	제품 카테고리 별 코사인 유사도를 산출하여 유사도 기반 제품군을 추천하는 방식의 시스템을 구축
Yoon and Joung(2020)	Content based Filtering with TF-IDF	웹크롤링을 통해 제품의 속성과 평가 및 후기 데이터를 추출하기 위해 텍스트 데이터를 수집하고 TF-IDF 방식을 활용하여 제품 간의 유사도를 판별하는 추천 시스템을 구축
윤여수, 박현준(2019)	Deep Learning(OpenCV)	딥러닝 기술을 활용하여 사용자의 얼굴 이미지를 인식하여 피부 특성을 추출하고 이에 맞는 파운데이션 제품을 추천하는 알고리즘을 구축
Yim <i>et al.</i> (2016)	Content based Filtering with KNN	OCR을 활용하여 제품의 성분 정보를 추출하고 사용자의 선호도와의 유사도를 기반으로 추천 시스템 개발

용된 협업 필터링의 특성상 제품에 대한 선호도와 사용자의 특성 정보를 모두 반영한 추천 시스템을 구축하기에는 한계가 있다. 뷰티제품 관련 추천 시스템 선행연구 내용을 정리한 <표 1>을 통해 종합적으로 확인할 수 있다. 따라서, 본 연구에서는 사용자와 제품 선호도 사이의 비선형 관계를 모델링하고, 더불어 고객의 콘텍스트 정보를 투입 변수의 차원 확장(Input Dimension Extension) 형태로 반영하는 추천시스템 모델을 제안하고자 한다.

### III. 추천 시스템 모형

#### 3.1 연구 절차

본 연구는 고객이 스마트폰을 통해 H&B(Health & Beauty) 오프라인 매장에서 제품 검색한 로그와 고객 특성 정보 데이터를 활용한다. 검색 로그 데이터에는 고객별 관심 제품에 대한 잠재적 선호(Implicit Data)가 포함되어 있다고 볼 수 있다. 한편 고객 특성 정보에는 고객의 성별, 피부 타입, 피부 고민 등이 있다. 전처리 과정과 학습, 검증, 평가 데이터 분리 단계를 거쳐 실제 신경망 모델



<그림 1> 연구 절차

을 통해 고객 선호를 예측할 수 있도록 한다. 마지막으로 모델 예측력을 MAE와 F-measure를 통해 검증하고자 한다. 이때 실험 결과는 평가 데이터 셋의 평가지표를 토대로 기술하고자 한다. 아래 <그림 1>은 연구 단계를 나타내는 도식이다. 연구 단계는 고객의 검색 로그 데이터 및 특성 정보를 수집하는 것에서 시작하여, 고객의 상품별 누적 검색 횟수를 테이블 형태로 변환하는 데이터 전처리 작업을 진행한다. 이후 학습-검증-평가 데이터 셋으로 구분한 후 예측 모델링 구축과 평가를 수행한다.

#### 3.2 데이터 수집

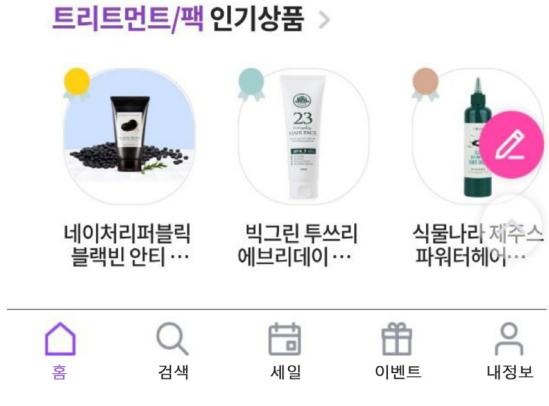
본 연구는 로켓류 주식회사에서 운영하고 있는 찍검 서비스에서 생성된 고객의 검색 로그 데이터를 기반으로 진행되었다. 찍검은 화장품 검색 플랫폼으로써, H&B 스토어 오프라인 매장에서 손쉽게 카메라로 상품명을 찍으면 자동으로 관련 정보를 제공하는 서비스이다. 고객의 검색 기록을 바탕으로 고객별 화장품 선호를 도출할 수 있으며, 이를 바탕으로 화장품 제조업체 혹은 H&B 스토어 등에 생산과 마케팅 전략을 제시함으로써 수익을 창출하는 비즈니스 모델을 가지고 있다. 아래 <그림 2>를 통해 찍검 서비스를 접속하였을 때 마주하는 첫 화면과 추천시스템 기반 상품 추천 서비스 제공 방법(예시)에 대해 확인할 수 있다.

고객의 과거 검색 기록을 토대로 고객이 향후 높은 선호 혹은 관심을 나타낼 것으로 예상되는 제품을 선제적으로 추천하는 모델을 개발하여, 고객에게 제품 검색 편의를 제공하고 동시에 찍검 서비스의 활성화에 기여할 수 있을 것으로 기대한다. 본 연구는 고객의 제품 검색 데이터를 기반으로 미검색 제품에 대한 예상 선호를 도출하는 모델을 구축하여, 예상 선호가 높은 상위 제품군을 추천하는 서비스를 개발하는 데 실무적 목적을 가지고 있다.

고객 제품 검색 데이터는 2020년 1월 2일부터



찍검 서비스 첫 화면



추천 상품 제시 방법

〈그림 2〉 찍검 서비스 첫 화면과 추천 상품 제시 방법

7월 23일까지 수집된 것을 활용하였다. 해당 데이터는 7,800명의 57,300건의 로그 데이터가 포함되어 있으며, 1,429개의 제품을 대상으로 구성되어 있다. 검색 로그 데이터는 페이지 뷰 내에서 랜덤워크 형식으로 부여되는 URL 형태로 수집되며, 상이한 사용자의 같은 제품의 검색로그는 동일한 URL이 부여되어 제품 식별과 선호도 파악이 가능한 데이터로 볼 수 있다. 정제된 제품 조회 데이터에 대한 예시는 <표 2>와 같다.

## 〈표 2〉 제품 조회 데이터 예시

고객 토큰	제품 ID	조회수
410	493	1
410	5514	2
410	1658	2
410	66	1
410	93	2
...	...	...
3446	4118	2
3446	2745	1
8125	18	1
4015	3308	1
4015	153	1

하기 <표 3>은 입력층에서 투입된 고객 특성 변수에 관해 정리한 것이며, 고객의 성별, 피부타입, 그리고 피부고민 부문에서 고객이 회원가입 시 선택할 수 있는 항목에 대한 정보가 포함되어 있다. 수집된 데이터에 대한 기초통계정보와 상관관계분석에 대해서 <표 4>에 수록하였다. <표 4> 을 통해서 건성, 민감성, 복합성과 같은 피부타입과 미백, 건조함과 같은 피부고민이 제품조회와 양의 상관관계를 보이는 것을 확인할 수 있다.

### 〈표 3〉 고객 특성 정보표

특성	카테고리
성별	남성, 여성
피부타입	지성, 건성, 복합성, 민감성
피부고민	주름&탄력, 미백&다크서클, 트러블&피부진정, 건조함&각질, 유분기&블랙헤드

### 3.3 모형 개요

본 연구에서는 고객의 개인 선호도를 반영한 화장품 추천시스템 모형을 제안하고, 기존의 연구 모형인 일반화 행렬 분해 모형과 선풍, 신경망 혼합

〈표 4〉 기술통계 및 상관계수표

변수	평균	표준 편차	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
(1) 제품조회	1.321	1.429	1										
(2) 지성	0.214	0.41	-0.012	1									
(3) 건성	0.386	0.487	0.001	-0.413	1								
(4) 복합성	0.237	0.425	0.009	-0.29	-0.441	1							
(5) 민감성	0.164	0.37	0.002	-0.231	-0.351	-0.246	1						
(6) 성별 (여성)	0.91	0.286	-0.065	-0.121	0.02	0.038	0.065	1					
(7) 주름 탄력	0.315	0.465	0.074	-0.091	0.053	0.029	-0.002	-0.067	1				
(8) 미백, 다크서클	0.514	0.5	0.017	-0.028	0.035	0.019	-0.037	0.096	0.181	1			
(9) 트러블, 피부진정	0.61	0.488	0.001	0.027	0.033	-0.134	0.081	0.105	-0.027	0.038	1		
(10) 건조함, 각질	0.489	0.5	0	-0.188	-0.046	0.19	0.05	0.055	0.208	0.128	0.08	1	
(11) 유분기, 블랙헤드	0.523	0.5	-0.018	0.125	0.131	-0.203	-0.078	0.048	0.056	0.157	0.265	0.124	1

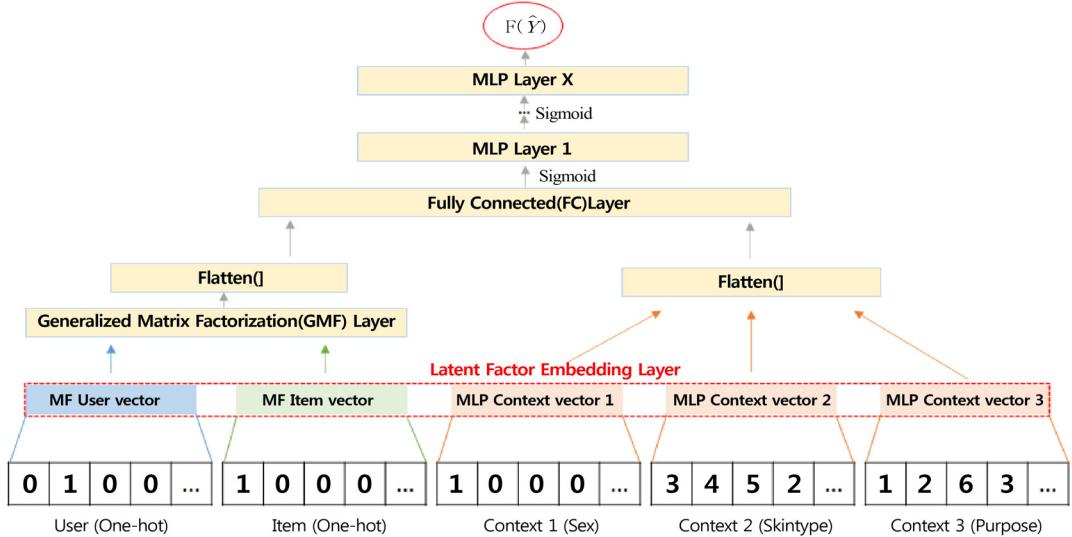
필터링 모형과의 성능 비교를 진행하고자 한다. 우선 본 연구에서 제안하는 모델의 특징은 고객의 특성 정보와 고객의 뷰티 제품 선호도를 반영하고 기존의 메모리 기반 협업 필터링의 한계점 중 하나인 콜드 스타트 문제를 보완할 수 있는 모델 기반의 협업 필터링으로 설계한 점이다. 고객 특성 정보는 입력 변수의 차원 확장(Input Dimension Extension) 형태로, 고객의 뷰티 제품 선호도는 잠재 요인 임베딩 레이어(Latent Factor Layer), 행렬 분해 및 내적 연산, 그리고 신경망을 활용한다. 아래 〈표 5〉는 제안 모델의 특징을 정리한 도식 자료이다.

한편, 제안하는 모형은 고객의 제품 선호에 담긴 잠재요인을 토대로 구축한 협업 필터링 모형이

며, 이는 기존의 일반화 행렬 분해 모형의 선형성 한계를 극복한 심층 신경망 협업 필터링 모델을 응용하여 개발하였다. 즉, 고객 특성 정보를 포함하는 신경망 모형(Neural Collaborative Filtering with Contextual Information, NCF-CI)을 제안하고자 한다. 신경망을 기반으로 추론하는 고객 선호도 예측 모델이라는 큰 틀 속에 고객 개인별 특성 정보(Contextual Information)를 입력 변수로 반영하도록 구성하였다. 이러한 특성 정보를 심층 신경망 모델의 투입 변수로 활용하는 효익은 콜드 스타트 이슈를 해결하고, 데이터 수집이 미약할 때 이를 콘텍스트 정보가 보완할 수 있다는 점에 있다(Covington et al., 2016). Covington et al.(2016)은 고객의 특성을 파악할 수 있는 각종 콘텍스트

〈표 5〉 추천시스템 주요 이슈 요약

가능 이슈	모델 내 해결 방안
고객 특성 정보	입력 변수의 차원 확장(Input Dimension Extension)
제품 선호도	잠재 요인 임베딩 레이어(Latent Factor Layer)
콜드 스타트(Cold start)	모델 기반 협업 필터링(Model based Collaborative Filtering)

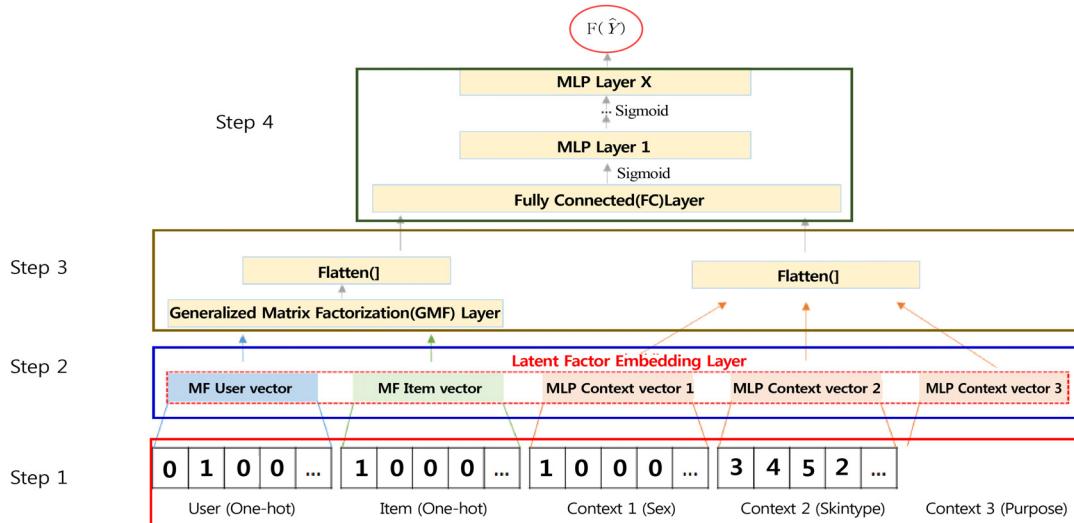


〈그림 3〉 제안 모델 아키텍처

정보를 통해 초기 신호가 약한 모델 및 신규 데이터의 유입 등을 보완하는 역할로 활용하고 있다고 밝히고 있다. 해당 정보 반영을 통해 고객의 제품 선호도에 미치는 영향 요인에 대한 설명력을 제고 할 수 있을 것으로 기대한다. 제안 모형은 <그림 3>과 같다.

### 3.4 제안 모형

본 연구에서 제안하는 모형은 고객의 제품 검색 로그 데이터를 기반으로 고객에게 맞춤형 제품을 추천해줄 수 있는 모델로 기획하였다. 아래 <그림 4>는 제안 모형의 단계별 구축 과정을 담고 있다.



〈그림 4〉 제안 모형 절차

(STEP 1) 먼저 고객과 아이템을 각각 원핫인코딩(One-hot encoding)으로 입력한다. 또한 고객의 특성 정보(Contextual information)는 고객별 라벨 인코딩(Label encoding)으로 입력한다. 본 연구는 화장품 검색 로그 데이터를 기반으로 제품 검색 추천을 목적으로 진행되는 연구로써, 고객이 회원 가입 시 입력한 성별, 피부타입, 그리고 피부고민(앱 사용목적)을 특성 정보로 활용하였다.

(STEP 2) 잠재요인 임베딩층을 구성하는 단계이다. 우선, 고객과 아이템은 각각 잠재요인 K개의 정보를 포함하는 임베딩층으로 구성한다. 여기서 K값은 연구자의 재량을 통해 설정할 수 있으며, 본 연구에서는 다양한 K값(200~400 범위)의 성능 지표 산출 실험을 통해 최적의 K를 300으로 설정하였다. 한편, 고객의 특성 정보는 각각의 특성 분류 개수에 부합하도록 잠재요인 K값을 지정하였다. 성별은 K=2, 그리고 피부타입과 피부고민은 회원가입 시 입력 받을 수 있는 최대 개수인 각각 K=9, K=27로 설정하였다.

(STEP 3) 일반화 행렬 분해 모형과 고객 특성 정보의 출력값을 각각 산출하는 단계이다. 고객과 아이템의 잠재요인 임베딩층은 일반화 행렬 분해 모형으로 출력값을 산출한다. 이는 일반화 행렬 분해가 잠재요인 기반의 고객 선호도 예측력이 뛰어나, 잠재요인 임베딩층의 연산 과정에 유의한 역할을 하기 때문이다(He et al., 2017). 한편, 고객 특성 정보는 고객과 아이템의 잠재요인 임베딩층과 달리 일반화 행렬 분해 모형을 거치지 않고 출력된다. 이는 고객과 아이템의 잠재요인 임베딩층에 비해 비중이 약하고, 연산 비용 측면에서 비효율적(Overhead cost)이기 때문이다.

(STEP 4) 일반화 행렬 분해 모형과 고객 특성 정보의 출력값을 결합하고 신경망 모형으로 확장하는 단계이다. 일반화 행렬 분해 출력값과 고객 특성 정보의 비연산 출력값을 단순 결합(Concatenate)하고 다층 신경망 모형을 구성한다. 아래 식 (1)과 같이, 일반화 행렬 분해 출력값과 고객의 특성 정보 임베딩 층의 출력값을 결합한

행렬을 입력으로 받은 후 각 은닉층을 거쳐 출력값( $\phi^{MLP}$ )을 산출하는 함수를 적용하였다. 이때  $a_L$ 은 각 층의 활성화 함수이며, W는 가중치, 그리고  $b_L$ 은 편향(Bias)을 의미한다.

$$\phi^{MLP} = \alpha_L \left( W_L^T \left( a_{L-1} \left( \dots a_1 \left( W_1^T \begin{bmatrix} p^{MLP} \\ q^{MLP} \end{bmatrix} + b_1 \dots \right) \right) \right) + b_L \quad (1)$$

아래 식 (2)에서,  $a_{out}$ 에 비선형 관계를 모델링 할 수 있는 활성화 함수를 적용한 것을 확인할 수 있다. 본 모형에서의 활성화 함수로는 시그모이드(Sigmoid)를 활용하였다. 해당 함수는 Relu 등 다른 활성화 함수를 활용한 실험 결과와 비교하여 가장 우수한 성능을 보여 선정하였다.

$$\hat{y}_{ui} = a_{out}(h^T(p_u \circ q_i)) \quad (2)$$

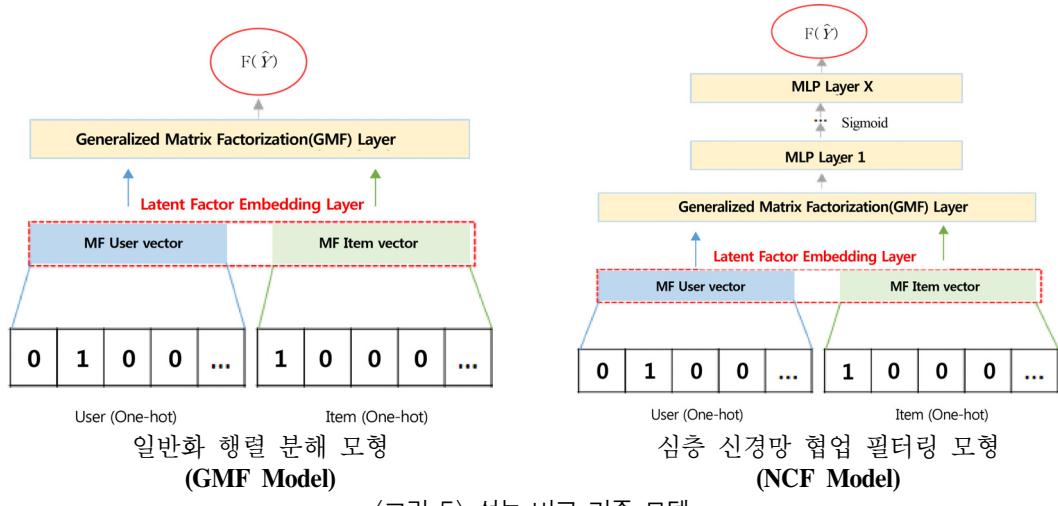
이는 선형 모델의 한계를 각 출력층에서 비선형 활성화 함수 활용으로 보완한 모형을 제안한 연구를 참조하였다(He et al., 2017). 모델의 학습은 확률적 경사 하강법(Stochastic Gradient Descent) 방식을 사용하였으며, 목적함수(MSE)를 최소화하도록 학습을 진행하였다.

## IV. 실험

### 4.1 분석 결과

본 연구가 제안하는 모형은 고객의 검색 로그 데이터를 기반으로 구축한 추천 아키텍처로, 전통적인 추천시스템과 추천 성능을 비교하기 위하여, 상위 N개 상품을 추천하고자 하였다. 비교하고자 하는 모델은 일반화 행렬 분해 모형과 고객 특성 정보를 반영하지 않은 심층 신경망 협업 필터링 모형이다. 아래 <그림 5>는 본 연구의 비교 모델인 일반화 행렬 분해와 심층 신경망 협업 필터링 모형 모형의 아키텍처에 관한 자료이다.

상위 N개 상품을 추천 성과 측정을 하기 앞서,



&lt;그림 5&gt; 성능 비교 기준 모델

본 모형을 거쳐 산출된 예측 값과 실제 선호도와의 MAE를 산출하여 모델 예측력을 비교하고자 한다. MAE는 실제 선호도 평가 수치를 나타내는  $p$ 와 예측된 수치  $q$  사이의 오차 절댓값에 대한 평균을 의미한다. 아래 식 (3)은 MAE를 측정하는 수식이다.  $N$ 은 평가 대상의 숫자를 뜻한다.

$$MAE = \frac{\sum_{i=1}^{n|p-q|}}{N} \quad (3)$$

본 연구에서 제안하는 모델인 고객 특성 정보 포함 신경망 모형(Neural Collaborative Filtering with Contextual Information, NCF-CI)과 일반화 행렬 분해와 심층 신경망 협업 필터링 모형의 MAE 및 MSE 수치는 아래 <표 6>과 같다. 이때, 평가의 신뢰성 제고를 위해, K겹(K=5) 교차검증 방식을 활용하여 각각의 MAE와 MSE 값을 도출하였다. 일반화 행렬 분해 모델은 MAE 3.012에 그쳤으며,

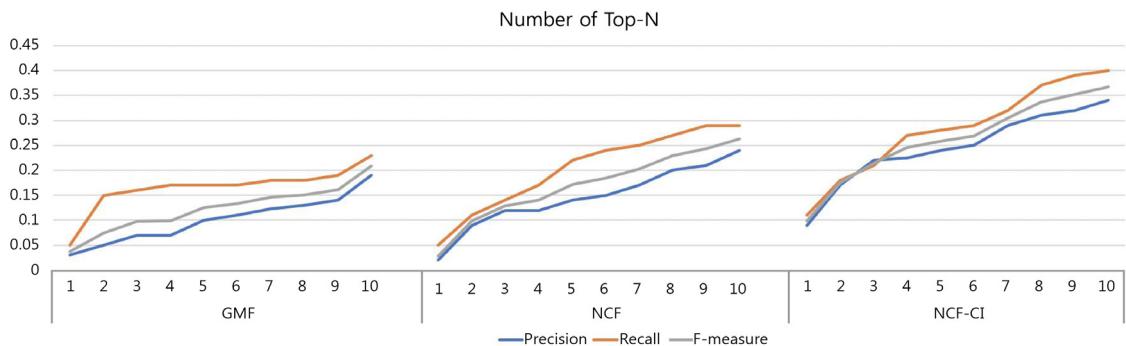
심층 신경망 협업 필터링 모델은 2.783, 그리고 NCF-CI는 2.051을 기록하면서, 본 연구에서 제안한 예측 모델의 성능이 우수하다고 평가할 수 있다.

MAE 수치에 대한 성능 비교에서 그치지 않고, TOP-N 추천 모델 성능 평가를 진행하였다. TOP-N 추천 모델 성능 평가 방식에서  $N$ 은 추천한 제품의 개수를 의미하며, 평가 데이터를 기반으로 내림차순 정렬된 추천 목록을 생성한 후  $N$ 개 내고객의 제품 선택 여부에 따른 정밀도, 재현율, F-measure 값을 측정하여 비교하는 방식이다(이주오 등, 2020). 아래 <그림 6>은  $N$ 의 개수 변화를 달리 적용하였을 때의 각 모델별 정밀도, 재현율, F-measure 값을 나타낸 그래프 자료이다.

<그림 6>에서, 일반화 행렬 분해 모형보다는 심층 신경망 협업 필터링 모형이, 심층 신경망 협업 필터링 모형보다는 NCF-CI 모델이 정밀도, 재현율, F-measure 측면에서 우수한 성과를 나타내는 것을 확인할 수 있다. 구체적으로, F-measure는 정

&lt;표 6&gt; 모델 성능 비교평가

	CV Mean GMF	CV Std GMF	CV Mean NCF	CV Std NCF	CV Mean NCF-CI	CV Std NCF-CI
MAE	3.012	0.05029	2.783	0.034694	2.051	0.021187
MSE	0.8226	0.005161	0.5281	0.006529	0.4469	0.0044872



〈그림 6〉 TOP-N 성능 비교표

밀도와 재현율에 대한 정보를 모두 반영하고 있기 때문에, 이를 중심으로 각 모델의 성능을 살펴볼 수 있다. 결론적으로, 전자의 두 모델에 비해 본 연구에서 제안한 NCF-CI 모델이 화장품 관련 고객 선호를 예측하여 검색 아이템을 추천하는 모델 측면에서 탁월하다고 평가할 수 있다.

한편, N의 값이 10개로 증가함에 따라 더 좋은 성능을 보이고 있다. 이는 고객의 예상 선호도 출력값이 고객의 선호를 제대로 반영하지 못한다는 한계점을 보인 것으로 해석할 수 있다. 즉, 출력값이 늘어날수록 고객의 차기 행동이 포진해 있다는 의미이기 때문이다. 하지만, 일반화 행렬 분해와

심층 신경망 협업 필터링 모형에 비해서는 비교적 나은 성능을 보였다는 점에서 의의가 있다고 평가할 수 있다. 모델 서빙은 <그림 7>과 같이, 고객의 검색로그 및 특성이 전산에 입력되면, 추천 상품이 품목 번호 형태로 산출되어 해당 품목과 연계된 실제 상품 이미지가 고객 어플 화면에 전송되는 방식으로 이루어질 수 있다. 또한, 아래 예시의 경우 고객에게 추천된 상품의 예시는 <표 7>과 같다. 예시 데이터는 트러블&피부진정, 건조함&각질의 피부고민을 지닌 건성 타입의 여성고객이다. 해당 고객에 대한 Top-5 추천 제품들은 보습 성분이 포함된 클렌징, 바디, 스킨케어 제품으로

```
def recommend_top_k(id, k, skin_type, gender, skin_worry):
    new_id = id2idx[id]

    user_ids = np.array([new_id]*4360)
    product_ids = np.array([x for x in range(4360)])
    skin_types = np.array(skin_type)*4360
    genders = np.array([gender]*4360)
    skin_worries = np.array([skin_worry]*4360)
    pred = model.predict([user_ids, product_ids, skin_types, genders, skin_worries]) + mu
    pred = np.ravel(pred, order='C')
    df_new = pd.DataFrame({'product_id':product_ids, 'click':pred})

    lst = []
    for i in df_new.sort_values(by='click', ascending=False).head(k).index.tolist():
        lst.append(idx2product[i])

    return lst

recommend_top_k('dxNbcbKWIffg:APA91bHFe0egBjJH0a21_UMNlp5qr0uxCdtHyYX-xarh93vqT0bp39uGPa5BNuH-6pa0feLMTNkHCuhsvn1T5vReyysjy27DFgElGXFI2tkXThCsBVW1yiir6sk2AF8fyiQiAwBkmnb', 5, 1, 0, 10)
[21.0, 2816.0, 9443.0, 22608.0, 9040.0]
```

〈그림 7〉 추천 결과 예시 코드

〈표 7〉 추천 결과 예시

특성	카테고리	제품 예시
성별	여성	
피부타입	건성	
피부고민	트러블&피부진정, 건조함&각질	

구성되었다. 이를 통해 추천 모델에서 고객의 피부타입 및 고민에 적합한 특성을 지닌 제품이 추천되는 것을 확인할 수 있다.

## V. 토의 및 시사점

### 5.1 연구 토의 및 시사점

본 연구는 고객의 상품 선호도 예측을 통한 상품군을 추천하는 서비스 개발과 관련된 것으로, 고객의 선호를 정교하게 반영할 수 있는 방법에 관해 고민하였다. 제안 모델에서는 크게 두 가지 특장점을 가지고 있다. 첫째, 고객과 상품 예상 선호 간 비선형 관계를 모델링한 것이다. 기존의 일반화 행렬 분해는 고객과 상품 선호 간 이차원 형태로 모델링함에 따라 고객의 잠재 선호 요인(Latent Factor)을 충분히 반영하지 못한 측면이 있다. 하지만 본 모델은 신경망 기반의 비선형 모델링을 통해 고객의 잠재 선호 요인을 충실히 반영하고자 하였다. 두 번째는 고객별 다양한 특성 요소를 모델에 반영한 것이다. 고객의 성별, 나이, 피부 타입 등 개인별 특성이 다르기 때문에 추천하는 상품 또한 그에 맞춰 맞춤형으로 제공될 필요가 있다. 그러한 측면에서 본 연구에서의 제안 모델은 고객의 특성을 투입 변수에 반영함으로써, 고객 맞춤형 추천 서비스가 가능하도록 설계되었다.

한편, 실험 결과 및 성능 평가 부분에 주목할 필요가 있다. 제안한 모델의 성능 평가는 일반화 행렬 분해 모형과 심층 신경망 협업 필터링 모형을 기반으로 이루어졌다. 우선, 일반화 행렬 분해 모형

을 활용한 이유는 이차원 형태의 일반화 행렬 분해 모델에 비해 비선형 관계를 모델링한 본 연구의 의의를 살펴보기 위함이다. 심층 신경망 협업 필터링 모형을 활용한 것은 고객 특성 요인을 반영한 부분의 효과성을 검증하고자 한 것이다. 두 모델과의 비교를 통해, 본 연구에서 고객-상품 선호 간 비선형 모델링 한 것과 고객 특성 요소를 모델에 직접 반영한 것은 효과적인 것으로 평가할 수 있다.

본 연구가 지닌 학술적 시사점은 다음과 같다. 첫째, 기존 연구는 고객의 특성을 반영하지 않은 채 제품에 대한 고객 선호도나 제품 관련 컨텐츠 중심의 추천시스템 연구(Gholamian *et al.*, 2011; Matsunami *et al.*, 2017; Yim *et al.*, 2016)가 진행된 것에 비해, 본 연구는 고객의 콘텍스트 정보를 포함하는 포괄적인 추천시스템 모델을 제안한 것이다. 둘째, 고객에게 직접 제품 평가(평점) 데이터를 제공받아 진행된 연구(Yim *et al.*, 2016)에 비해 고객의 제품 누적 클릭 수 등 선호도 잠재 데이터(Implicit data)를 활용하여 평점 데이터가 갖는 희소성, 주관성 등의 한계점을 보완한 것이다.

본 연구가 가진 실무적 시사점은 다음과 같다. 첫째, 쏟아지는 뷰티 제품 정보 속에서 선택의 복잡성에 직면한 고객에게 맞춤형 상품을 추천하여 구매를 유도할 수 있다. 둘째, 뷰티 업체는 고객의 선호를 미리 파악하여, 맞춤형 프로모션과 타겟 마케팅을 통해 불필요한 마케팅 비용을 줄일 수 있다. 셋째, 고객의 피부 특성, 성별, 연령 등의 콘텍스트 정보를 바탕으로 적합한 제품을 추천함으로써, 고객 경험을 개선하고 반품 등 고객 서비스(CS) 비용을 감축할 수 있다.

## 5.2 한계점 및 향후 연구 방향

본 연구의 한계점은 크게 두 가지 존재한다. 우선 TOP-K 추천 성능 지표 검증에서 1개부터 10개의 상품을 기준으로만 검증한 것이다. 10개 이상의 추천 개수에 따른 성능 지표 변화 전부 실험하지 못하여 최적의 추천 개수를 제시하지 못하였다. 두 번째는 고객 특성 변수를 제한적으로 활용하였다. 본 연구에서 활용한 데이터는 실제 서비스를 진행하고 있는 기업 정보이기 때문에 영업 부문에 부정적인 영향을 주지 않는 선에서 정보를 수집할 수밖에 없었다. 따라서, 다양한 고객 특성 데이터를 대상으로 확장성 있는 실험을 진행할 수 없었던 것이 한계로 볼 수 있다.

향후 연구 방향은 다음과 같다. 첫째, 예상 선호도 TOP-K 추천 방식에서 최적의 K 값을 도출하는 방법에 관한 연구가 필요하다. 본 연구에는 예상 선호도 상위 1개부터 10개를 추천하였을 때 고객이 해당 제품을 선택한 비율을 토대로 실험을 진행하였지만, 고객 경험 개선과 추천 서비스의 효과성 측면에서 적절한 K 값을 찾는 것은 의미가 있기 때문이다. 둘째, 고객이 남긴 댓글과 같은 텍스트 데이터 역시 고객의 선호를 예측하는 데 도움이 된다. 따라서 텍스트 자료를 포함한 확장된 추천시스템 모델 연구를 진행한다면 학술적·실무적 관점에서 크게 이바지할 수 있을 것으로 기대한다.

## 참고문헌

- [1] 구혜경, 나종연, “여성 소비자의 화장행동 분석을 통한 화장 맥락의 이해”, *소비문화연구*, 제18권, 제3호, 2015, pp. 147-179.
- [2] 송희석, “심층신경망 기반의 뷰티제품 추천시스템”, *Journal of Information Technology Applications & Management*, 제26권, 제6호, 2019, pp. 89-101.
- [3] 신유원, 신민선, “2019 화장품산업 분석 보고서”, 한국보건산업진흥원, 2019.11.18., Available at <https://www.khidi.or.kr/board/view?linkId=48809330&menuId=MENU00085>.
- [4] 윤여수, 박현준, “사용자의 피부 색상에 기반한 파운데이션 색상 자동추천”, *한국지능시스템학회 논문지*, 제29권, 제4호, 2019, pp. 280-284.
- [5] 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.
- [6] 이주오, 이형걸, 김아연, 허승연, 박우진, 안용학, “생활지수를 이용한 협업 필터링 기반 장소 추천 시스템의 설계 및 구현”, *한국융합학회논문지*, 제11권, 제8호, 2020, pp. 23-31.
- [7] Bobadilla, J., F. Ortega, A. Hernando, and A. Gutiérrez, “Recommender systems survey”, *Knowledge-based Systems*, Vol.46, 2013, pp. 109-132.
- [8] Covington, P., J. Adams, and E. Sargin, “Deep neural networks for YouTube recommendations”, *Proceedings of the 10th ACM Conference on Recommender Systems*, 2016, pp. 1-8.
- [9] Cremonesi, P., F. Garzotto, and R. Turrin, “Investigating the persuasion potential of recommender systems from a quality perspective: An empirical study”, *ACM Transactions on Interactive Intelligent Systems (TiiS)*, Vol.2, No.2, 2012, pp. 1-41.
- [10] Gao, C., X. He, D. Gan, X. Chen, F. Feng, Y. Li, and D. Jin, “Neural multi-task recommendation from multi-behavior data”, In *2019 IEEE 35th International Conference on Data Engineering (ICDE)*, 2019, pp. 1554-1557.
- [11] Gholamian, M., M. Fathian, M. Julashokri, and A. Mehrbod, “Improving electronic customers’ profile in recommender systems using data mining

- techniques”, *Management Science Letters*, Vol. 1, No.4, 2011, pp. 449-456.
- [12] He, L. Liao, H. Zhang, L. Nie, X. Hu, and T. Chua, “Neural Collaborative Filtering”, *Proceedings of the 26th International Conference on World Wide Web*, 2017, pp. 1-10.
- [13] He, X., X. Du, X. Wang, F. Tian, J. Tang, and T. S. Chua, “Outer product-based neural collaborative filtering”, *Proceedings of the 27th International Joint Conference on Artificial Intelligence*, 2018.
- [14] Jannach, D., M. Zanker, A. Felfernig, and G. Friedrich, *Recommender Systems: An Introduction*, Cambridge University Press, 2010.
- [15] Kim, M. G. and K. J. Kim, “Recommender systems using SVD with social network information”, *Journal of Intelligence and Information Systems*, Vol.22, No.4, 2016, pp. 1-18.
- [16] Koren, Y., R. Bell, and C. Volinsky, “Matrix factorization techniques for recommender systems”, *Computer*, Vol.42, No.8, 2009, pp. 30-37.
- [17] Matsunami, Y., M. Ueda, and S. Nakajima, “How to find similar users in order to develop a cosmetics recommender system”, *International Multi Conference of Engineers and Computer Scientists*, 2017, pp. 337-350.
- [18] Park, S. T. and W. Chu, “Pairwise preference regression for cold-start recommendation”, *Proceedings of the third ACM Conference on Recommender Systems*, 2009, pp. 21-28.
- [19] Patty, J. C., E. T. Kirana, and M. S. D. K. Giri, “Recommendations system for purchase of cosmetics using content-based filtering”, *International Journal of Computer Engineering and Information Technology*, Vol.10, No.1, 2018, pp. 1-5.
- [20] Persson, P., “Attention manipulation and information overload”, *Behavioural Public Policy*, Vol.2, No.1, 2018, pp. 78-106.
- [21] Ricci, F., G. Adomavicius, B. Mobasher, and A. Tuzhilin, “Context-aware recommender systems”, *AI Magazine*, Vol.32, No.3, 2011, pp. 67-80.
- [22] Schafer, J. B., D. Frankowski, J. Herlocker, and S. Sen, “Collaborative filtering recommender systems”, *In The Adaptive Web*, 2007, pp. 291-324.
- [23] Su, X. and T. M. Khoshgoftaar, *A survey of Collaborative Filtering Techniques*, Advances in Artificial Intelligence, 2009.
- [24] Yim, Y. J., H. S. Bae, Y. J. Jeong, M. Y. Kim, A. Nasridinov, K. H. Yoo, and J. E. Hong, “A user driven cosmetic item recommendation system by character recognition”, *Proceedings of the Korea Information Processing Society Conference*, 2016, pp. 722-725.
- [25] Yoon, J. and S. Joung, “A big data based cosmetic recommendation algorithm”, *Journal of System and Management Sciences*, Vol.10, No.2, 2020, pp. 40-52.
- [26] Zhang, Z. K., C. Liu, Y. C. Zhang, and T. Zhou, “Solving the cold-start problem in recommender systems with social tags”, *EPL(Europhysics Letters)*, Vol.92, No.2, 2020, p. 28002.

Information Systems Review

Volume 23 Number 4

November 2021

## Beauty Product Recommendation System using Customer Attributes Information

Hyojoong Kim\* · Woosik Shin\*\* · Donghoon Shin\*\*\* · Hee-Woong Kim\*\*\*\* · Hwakyung Kim\*\*\*\*\*

### Abstract

As artificial intelligence technology advances, personalized recommendation systems using big data have attracted huge attention. In the case of beauty products, product preferences are clearly divided depending on customers' skin types and sensitivity along with individual tastes, so it is necessary to provide customized recommendation services based on accumulated customer data. Therefore, by employing deep learning methods, this study proposes a neural network-based recommendation model utilizing both product search history and context information such as gender, skin types and skin worries of customers. The results show that our model with context information outperforms collaborative filtering-based recommender system models using customer search history.

**Keywords:** *Recommendation System, Deep Neural Network, Latent Factor, Customer Attributes Information, Beauty Product, Searching Log*

---

\* MS. Candidate, Graduate Student, Graduate School of Information, Yonsei University

\*\* Ph.D. Candidate, Graduate School of Information, Yonsei University

\*\*\* Candidate, Graduate Student, Graduate School of Information, Yonsei University

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

\*\*\*\*\* CEO, RocketView

## ● 저자 소개 ●



김 효 중 (libertygood9@yonsei.ac.kr)

연세대학교 정보대학원 비즈니스 빅데이터 분석 트랙 석사과정이다. 현재 디지털서비스연구실에서 데이터분석 관련 연구들을 진행 중이다. 주요 연구 분야는 빅데이터분석, 딥러닝, 추천시스템 등이다.



신 우 식 (woosik.shin@yonsei.ac.kr)

연세대학교 정보대학원에서 비즈니스 빅데이터 분석 트랙 박사과정이다. 현재 디지털 서비스 연구실에서 빅데이터 분석 관련 연구를 진행 중이다. 주요 연구분야는 Digital Marketing, Cyberbullying, Business Analytics 등이다.



신 동 훈 (dh\_shin@yonsei.ac.kr)

연세대학교 정보대학원 비즈니스 빅데이터 분석 트랙 석사과정이다. 현재 디지털 서비스연구실에서 데이터분석 관련 연구들을 진행 중이다. 주요 연구 분야는 딥러닝, 추천시스템, 자연어처리 등이다.



김 회 웅 (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 50여 편의 논문이 게재되었다. JAIS, IEEE TEM의 편집위원으로 활동했고, KrAIS 회장을 역임했다.



김 화 경 (hwakyung@rocketview.io)

스웨덴왕립공과대(KTH)에서 Interactive System Engineering 전공으로 석사학위를 취득한 후, 삼성전자 무선사업부 SW 책임연구원을 역임하였다. 삼성전자 사내벤처 C-Lab에서 스피노프하여 현재 주식회사 로켓뷰CEO로 재직 중이다.

논문접수일 : 2021년 05월 21일  
1차 수정일 : 2021년 07월 05일

게재확정일 : 2021년 08월 24일  
2차 수정일 : 2021년 08월 16일