

텍스트마이닝 기법을 활용한 올리브영 고객 리뷰 감성분석

진민준, 서혜교

경북대학교

jinmin1569@nate.com, tigraphy@gmail.com

The Sentiment Analysis of Olive Young Customer Reviews Using Text Mining Techniques

Min Joon Jin, Hye gyo Suh

KyungPook National Univ.

요약

본 논문은 올리브영에서 구매한 화장품에 대해 고객들이 남긴 리뷰를 분석하였다. 스킨토너, 에센스/세럼/앰플, 크림, 로션, 미스트/오일, 스킨케어세트 등 다양한 스킨케어 제품군에 대한 데이터를 수집하였다. ELECTRA 모델을 이용하여 감성점수를 재산출하여 재산출된 감성점수를 기반으로 TF-IDF와 Bi-gram 기법을 활용하여 제품별 핵심 키워드를 추출하였다. 이를 통해 각 제품군에 대한 핵심 특징과 감성 정보를 제공할 수 있다.

I. 서론

세계적으로 다양한 산업의 어려움속에서도 화장품 산업은 계속해서 고객들의 수요가 증가하고 있다. 특히 온라인 구매는 높은 인기를 끌며, 새로운 소비 경향과 유통 채널의 다방면에서 발전 가능성을 제시하고 있다. 특히 2019년 한국 국내의 화장품 시장규모는 대략 123억 달러로 추정되며, 세계 화장품시장의 약 3%를 차지하며 전체 8위를 기록했고, 최근 5년간 연평균성장률 16%를 기록했으며 2019년 16조 2,633억원정도 기록하였다.(한국보건산업진흥원, 2021)

컴퓨터와 스마트폰의 발달로 온라인으로 화장품을 구매하는 경우가 많아지고 있으며 소비자들은 이제 온라인 플랫폼을 통해 제품이나 서비스에 대한 평가를 실시간으로 작성하고 공휴할 수 있는 환경에 놓이게 되었다.

온라인 리뷰는 타인의 다양한 경험을 기반으로 작성되며, 실제 구매자의 구매 행동에 간접적으로 영향을 미쳐 구매 시 느끼는 위험을 보완하는 역할을한다(김민정, 2007). 또한, 선행 연구들에 따르면, 온라인 리뷰의 신뢰도는 절대적인 기준이라기보다는 수용자에 따라 달리 인식되는 상대적인 특성을 지니고 있다. 이러한 특성은 소비자의 만족도나 재구매 의도와 같은 구매의사결정 과정에 중요한 영향을 미치는 요소로 작용한다(이은재,심완섭, 2007).

본 논문에서는 올리브영 사이트에서 고객들이 남긴 리뷰를 수집하고 화장품 상품별로 가지고 있는 제품의 특성과 긍정적인 요인과 부정적인 요인들이 무엇인지 파악하고 최종적으로는 화장품의 개선요인들을 파악하여 시사점을 제시하는 것을 목표로 하였다.

II. 본론

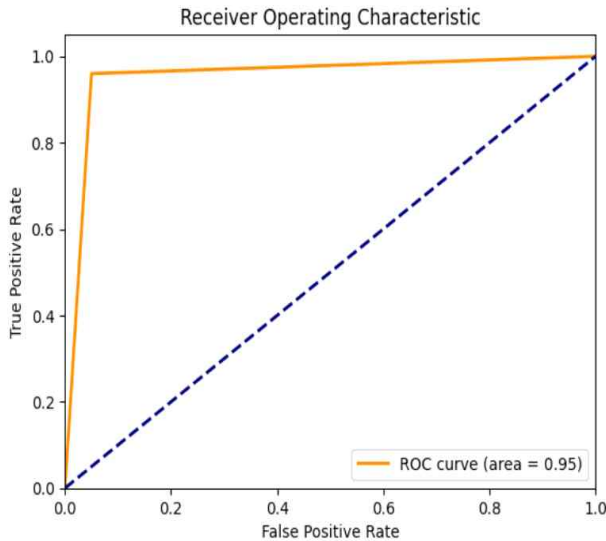
본 연구에서는 분석을 위해 올리브영에서 수집한 31,213건의 데이터를 활용하였다. [표 1]과 같이 데이터셋은 제품 정보, 고객정보, 구매 정보, 그리고 고객 리뷰 및 감성 분석 결과 등 총 22개의 변수를 포함하고 있다.

[표 1] 입력변수의 종류와 개수

구분	변수 리스트	개수
기본 정보	category, name, brand, price, ranking, sale_price	6
고객 특성	id, skin_type	2
구매 시점	date	1
제품 선택 기준	rate, select_title, select_content 등	7
리뷰 및 감성분석	txt, sentiment, predict_sentiment, positive_prob, negative_prob, score	6

2-1. 감성분석

본 논문에서는 고객 리뷰의 감성을 분석하기 위해 ELECTRA 모델을 활용하였다. 먼저, OpenAI의 GPT 모델을 이용하여 8,000개의 리뷰 데이터(긍정 4,000개, 부정 4,000개)에 대한 라벨링을 수행하였다. 이 라벨링된 데이터를 ELECTRA 모델의 학습 데이터로 사용하였다. ELECTRA 모델은 30 epochs 동안 학습되었으며, 검증 데이터셋에서 94~95%의 정확도를 보였다. 최종 테스트 데이터셋에서는 94.7% 정확도를 달성하여 높은 성능을 입증하였다. 모델의 성능은 ROC 곡선을 통해 시각화되었으며, 이는 모델의 분류 능력이 우수함을 보여주었다.



[그림 1] ROC CURVE

감성 점수 산출을 위해 모델의 출력인 로짓(logit)을 소프트맥스 함수를 통해 확률로 변환하였다. 이후 부정 확률에는 0.7의 가중치를, 긍정 확률에는 0.3의 가중치를 적용하여 최종 감성 점수를 계산하였다. 이러한 가중치 적용은 부정적인 감성을 더 민감하게 반영하기 위함이다. 분석에서 활용한 감성 점수 재산출 공식은 다음과 같다.

$$\text{감성 점수 재산출} = (\text{긍정 확률} * \text{긍정 확률 가중치} - \text{부정 확률} * \text{부정 가중치}) / (\text{긍정 가중치} + \text{부정 가중치})$$

이러한 방식으로 재산출된 최종 평점 점수는 최소 0.3에서 최대 5.298의 범위를 가졌으며, 이 점수를 기준으로 리뷰를 분류한 결과, 전체 리뷰 중 66%가 긍정적, 35%가 부정적인 것으로 나타났다.

2-2. 텍스트 마이닝

본 논문은 먼저 한국 전자 통신연구원(etri)에서 제공하는 형태소분석 api를 활용하여 형태소 분석을 실시하였다. 리뷰 텍스트중에서 긍정과 부정을 판별이 가능한 중요 의미를 파악할 수 있는 형태소인 동사, 명사, 형용사, 부사를 추출을 진행하였다. 둘째, 추출된 단어내 Bi-gram을 이용한 후 해당 키워드가 얼마나 포함되어 있는지를 확인하기 위해 단어 빈도분석을 실시하였다. 셋째, 문서 내 특정 단어의 빈도수(TF)와 여러 문서내의 특정 단어 빈도수(IDF)를 고려하여 키워드의 중요성을 파악하는 TF-IDF 분석을 실시했다.

[표 2] 제품별 긍정,부정 키워드 추출(Bi-gram)

제품	긍정	부정
로션	남자 친구, 바르 편한	피부 뒤집, 바로 쓰레기
스킨토너	자극 없이, 없이 순하	암모니아 냄새, 민감 피부
에센스/세럼/앰플	꾸준히 사용, 흡수 빠르	따끔거리 자극, 사용 중단
미스트/오일	너무 건조, 사용 너무	노즐 타입, 여간 불편

TF-IDF후 제품별 긍정 및 부정 단어를 추출한 결과는 [표 2]와 같다. 또한, 전체 제품중 평점별로 핵심 키워드를 추출하였는데 결과는 워드 클라우드로 시각화를 진행하였으며 [그림 2]는 재수정된 평점 5점 이상 키워드이고, [그림 3]은 4점 이상 5점이하 키워드이며, [그림 4]는 평점 1점이하 2점이상의 키워드이고 [그림 5]는 1점 이하의 키워드이다.



III. 결론

본 논문에서의 분석 결과는 다음과 같다. 평점이 높은 화장품의 핵심 키워드들은 주로 “자극 없이”, “없이 순하”, “꾸준히 사용”과 같이 자극이 없고 매우 순하다는 것을 알 수 있다. 부정적 키워드들은 주로 “트러블 올라오”, “피부 뒤집어지”, “사용 중단”과 같은 특정 화장품을 바르면 여드름이 생기고 피부가 뒤집어진다는 결과를 확인할 수 있다.

제품별로는 로션은 긍정 키워드는 “바르기 편하다”, 부정 키워드는 “피부가 뒤집어진다”, “쓰레기통으로 바로 버린다”이며, 나머지는 [표 2]에서 확인할 수 있다. 본 논문에서처럼 e-commerce에서 고객 리뷰를 수집해 감성분석을 수행하는 것은 화장품 회사의 마케팅 의사결정과 경영 전략 수립에 유용할 것이다.

참 고 문 헌

- [1] 김민정. 여성소비자의 정보탐색 유형별 화장품 구매성향 연구(2007)
- [2] 이은재, 심완섭. 인터넷 쇼핑 이용후기 메시지의 신뢰성평가와 구매시점 및 구매 후 행동특성연구. 국제 e-비즈니스학회, 8(3) 155-170