

## 감성 분석 화장품 사용자 리뷰에 대한 속성기반 감성분석

정희원<sup>o</sup>, 정영섭<sup>\*</sup>

<sup>o</sup>충북대학교 컴퓨터공학과,

<sup>\*</sup>충북대학교 컴퓨터공학과

e-mail: jhjmo0719h@naver.com, ysjay@chungbuk.ac.kr

## Aspect-based Sentiment Analysis on Cosmetics Customer Reviews

Heewon Jeong<sup>o</sup>, Young-Seob Jeong<sup>\*</sup>

<sup>o</sup>Dept. of Computer Engineering, Chungbuk National University,

<sup>\*</sup>Dept. of Computer Engineering, Chungbuk National University

### ● 요약 ●

온라인상에 인간의 감성을 담은 리뷰 데이터가 꾸준히 축적되어왔다. 이 텍스트 데이터를 분석하고 활용하는 일은 마케팅에 있어서 중요한 자원이 될 것이다. 이와 관련된 Aspect-Based Sentiment Analysis(ABSA) 연구는 한글에 있어서는 데이터 부족을 이유로 거의 선행연구가 없는 실정이다. 본 연구에서는 최근 공개된 데이터셋을 바탕으로 하여 화장품 도메인에 대한 소비자들의 리뷰 텍스트와 사전 라벨링 된 속성, 감성 극성을 기반으로 ABSA를 진행한다. Klue RoBERTa base 모델을 활용하여 데이터를 학습시키고, Python Kivipy 등으 로 전처리한 결과를 대시보드로 시각화하여 분석하기 쉬운 환경을 마련하는 방법을 제시한다.

**키워드:** 리뷰분석(review analysis), 속성기반 감성분석(aspect-based sentiment analysis), 언어모델(language model)

### I. Introduction

온라인상에서 리뷰 공유 문화가 활발해짐에 따라, 다양한 속성 (Aspect)과 감성(Sentiment)을 포함하는 데이터가 풍부해졌고, Aspect-Based Sentiment Analysis (ABSA) 전용 데이터 구축 및 연구 사례 또한 증가하는 추세이다 [1]. 특히 Hayeong Shim은 ABSA 활용이 마케팅은 물론이고 소비자에게도 이로운 전략이라 설명한다 [2].

마케터에게는 사용자의 평가가 담긴 텍스트가 만족도 분석에 있어 중요한 지표다. 이를 올바르게 분석하여 문제점이 있다면 개선하고, 장점이 있다면 극대화하는 것이 주요과업이다. 이를 통해 매출 향상 및 회사의 성장을 기대할 수 있다. 수많은 리뷰 데이터를 분석 후 빠르게 피드백을 반영하기 위해서는, 사람이 모든 리뷰를 수동적으로 검토하기보다, Natural Language Processing (NLP) 데이터 분석 기법을 사용하는 것이 불가피하지만, 한국어 데이터에 대한 ABSA 연구는 상대적으로 적은 편이다 [3]. 본 연구에서는 최근 AI-Hub에서 공개된 ‘속성 기반 감성분석 데이터’를 활용하여 화장품 리뷰에 대해 ABSA를 진행한 뒤 모델 성능을 측정한다. 이어 ABSA 결과를 한눈에 파악하고자, 테스트 데이터 셋을 활용해 모델이 예측한 결과를 기반으로 Superset 대시보드를 구축하여 분석해 본다.

화장품 리뷰를 ABSA 하기 위하여 본 연구에서는 앞서 언급한 ‘속성 기반 감성분석 데이터’의 도메인 중 ‘화장품’ 만이 모델 학습, 검증, 테스트용 데이터로 활용된다. 화장품 도메인을 카테고리리 나누면 ‘스킨케어’ 16,341건, ‘헤어/바디케어’ 15,420건, ‘메이크업/뷰티소품’ 13,485건, ‘남성 화장품’ 4,750건으로, 총 리뷰 데이터의 개수는 49,996건이다. 각 원천 데이터는 CSV 형식으로, 각 행에 한 명의 고객이 남긴 리뷰가 문장 단위가 아닌 전체 텍스트로 담겨있다. 반면 각 라벨링 데이터는 JSON 형식으로, 각 리뷰마다 전체 텍스트와 이에 대한 속성 및 감성이 존재하는 어절들만 따로 분류된 것도 제공한다. 속성의 이름과 범주는 화장품 도메인 내에서도 카테고리별로 다르다. 감성 극성은 0(neutral), 1(positive), -1(negative) 3가지가 존재한다. 라벨링 데이터 중 ABSA를 위해 사용할 값들은 Table 1과 같다.

데이터를 확인하여 미루어볼 수 있는 점은 다음과 같다. 화장품 도메인의 각 카테고리별로 속성이 상이하므로, 모델을 사용할 때 이 점을 고려해 처리해 줄 필요가 있다. 또한, Table1처럼 하나의 리뷰에 똑같은 Aspect가 2개 이상이 포함된 경우가 존재한다. 이렇게 되면 똑같은 Aspect 입에도 불구하고 서로 다른 SentimentPolarity를

가질 수도 있으므로, 이런 경우를 고려해 Aspect 중복 제거가 필요하다. 그리고 RawText에는 불용어(‘ㅎㅎ’, ‘~~’)가 존재하므로 한글만 남기도록 추가적인 전처리 작업이 필요하다.

한국어 속성 기반 감성분석을 위하여 KLUE RoBERTa-base 모델을 사용한다. 이 모델만으로 Topic Classification 을 사용해 속성 분류와 감성 분류 두 가지 작업을 모두 수행할 수 있다. Aspect 분류의 경우 하나의 리뷰 텍스트에 있는 모든 속성을 학습하기 위하여, 리뷰 RawText와, 각 속성의 존재 여부 값 (0 또는 1)을 Input으로 두고, Output은 리뷰에 해당 속성의 존재 여부 값 (0 또는 1)으로 하여 이진 분류 처리하였다. 속성이 화장품 도메인의 각 카테고리 별로 상이하므로, 속성 학습에 있어서는 각 카테고리마다 존재하는 속성의 개수만큼 모델을 돌린다. Sentiment 분류의 경우 Input은 ‘MainCategory#Aspect’의 형태로 스페셜 토큰이 RawText의 맨 뒤에 추가된 텍스트와, 이 속성에 대한 Sentiment 극성이 주어진다. Sentiment 극성은 원본 라벨링 데이터에서 positive(1) /negative(-1)/neutral(0) 이었던 값을 각각 positive(0)/ negative(1)/neutral(2)로 변환하여 사용하였다. Output으로는 이 RawText의 각 속성에 대해 예측한 감성 극성을 반환하도록 다중 분류 처리하였다.

Table 1. Data sample

Item	Value
MainCategory	메이크업/뷰티소품
RawText	향이 무지 약하네요. 향이 약하고 지속시간도 많이 짧네요.
Aspects	{Aspect: 향, SentimentPolarity: -1}, {Aspect: 지속력/유지력, SentimentPolarity: -1}, {Aspect: 향, SentimentPolarity: -1},

모델 성능은 테스트 데이터 셋으로 다음과 같은 3가지 경우를 두어 점수를 계산하는 데에 활용한다. True positive, False positive, False negative. 특히 F1-Score를 이용해 각 Epoch에서 Best 모델을 결정한다. 속성은 각 속성이 존재하는지 여부를 판단한 True, False 예측 결과값을 기준으로, 감성은 positive/negative/neutral 예측 결과값을 기준으로 정답과 대조하여 확인한다.

## II. Related Works

과거 LSTM, Naive bayes 등의 모델로 ABSA를 시도했던 방식에서 벗어나, 성능 향상을 위해 BERT 계열 언어 모델을 도입한 선행 연구가 있다. Ref. [4]는 Hugging Face에 게시된 BERT Base uncased 모델을 사용하여 pre-training과 fine-tuning을 거쳤고, 특히 하나의 레이어를 추가하여 ABSA를 수행했다. 다양한 평가 방법을 사용하여 BERT를 사용하는 것이 과거 사용했던 모델을 사용하는 것보다 성능이 좋다는 결론을 내렸다. 이 연구들은 영어 데이터에 대한 연구였으나, 한국어에 적용할 시에 좋은 결과를 낼 수 있을지는 미지수라고 볼 수 있다.

한글 BERT 언어 모델로는 xlm-roberta, KcBERT, klue의 BERT 계열 모델 모음이 있다. ‘속성 기반 감성 분석 데이터’를 제공했던 연구팀에서 ABSA 학습 및 예측 시에 klue-bert 모델을 사용했기

때문에, 본 연구에서도 klue의 BERT 언어 모델 중 하나를 활용한다.

Klue의 BERT 언어 모델에는 BERT-base, RoBERTa-small, RoBERTa-base, RoBERTa-large 등이 있다. YNAT(Topic Classification) 성능 비교를 했을 때 BERT-base가 가장 높은 점수를 냈고, RoBERTa-large가 그다음으로 높은 점수였다 [5]. YNAT 외의 분야에서는 BERT-RoBERTa-large가 가장 높은 점수를 보였지만 본 연구에서는 모델을 학습시키는 데이터의 양이 RoBERTa-large 모델을 사용할 만큼 많지 않은 관계로 RoBERTa-base를 선택해 화장품 리뷰 데이터에 대하여 ABSA를 수행한다.

본 연구는 앞서 언급한 연구와 다른 의의(novelty)가 있다. 첫째, 한글을 ABSA한 선행 연구 사례는 적은 편인데, 새롭게 배포된 대량 한글 데이터 및 BERT 계열의 모델을 사용하여 ABSA의 성능 및 효과를 입증하고자 시도했다. 둘째, 모델의 성능을 측정하는 것에서 그치지 않고, 대시보드 구축을 통해 ABSA의 결과를 시각적으로 파악하고자 하였다.

## III. The Proposed Method

### 1. Data Collection and Preprocessing

데이터는 AI-Hub에서 공개한 ‘속성 기반 감성분석 데이터’의 라벨링 데이터를 사용하여 Train, Validation, Test 데이터 셋을 6:2:2의 비율로 분리하여 활용하였다. ‘화장품’ 도메인만을 사용하였고, 이 도메인에는 4가지 Main Category(스킨케어, 헤어/바디케어, 메이크업/뷰티소품, 남성화장품)가 존재한다. 다만 앞서 Section I에서 언급한 바와 같이 다음과 같은 전처리 과정이 수반되었다.

첫 번째로, 정규표현식을 통해 RawText에서 한글만 남도록 했으며, 불용어 제거를 위해 띄어쓰기 중복을 제거했다. 두 번째로, 중복되는 Aspect는 제거하여 1개의 Aspect만 남도록 했다. Aspects 내에서 가장 먼저 등장했던 Aspect만 남도록 하였다. 마지막으로, MainCategory별로 지나치게 낮은 빈도(예: 수개, 수십개)로 등장하는 속성은 데이터에서 제거하였다.

### 2. Model

전처리된 데이터는 MainCategory별로 Klue RoBERTa-base 모델에게 주어졌고, 속성 학습 후에 감성을 학습하도록 했다. 모델 학습률은 [5]를 참고하여 1e-5로 하였고, epoch 횟수는 5로 하였다.

속성을 학습할 때는 소속 MainCategory의 속성의 개수만큼 모델을 돌렸다. 예를 들어 ‘남성화장품’에 속하는데 ‘가격’, ‘향’ 속성을 가지고 있는 리뷰 데이터가 있다고 하자. 남성화장품의 속성은 Table 2와 같이 총 20개로, 이 데이터는 ‘가격’을 1으로, 나머지 속성은 0으로 값을 설정하여 학습시킬 때와, ‘향’을 1으로, 나머지 속성은 0으로 값을 설정하여 학습시킬 때의 경우에 소속된다. 감성을 학습할 때는 각 MainCategory마다 한 번씩 모델을 돌렸으며, 특히 스페셜 토큰을 리뷰 텍스트 맨 뒤에 이어 붙이는 작업이 필요했다. 예를 들면 ‘남성화장품#가격’, ‘남성화장품#향’이라는 문자열이 리뷰 텍스트 뒤에 이어 붙게 되는 것이다. 스페셜 토큰의 형태는 ‘MainCategory#Aspect’이며, 이를 MainCategory마다 모델의 tokenizer에 스페셜 토큰으로

추가하였다. 이렇게 스페셜 토큰이 추가된 리뷰 텍스트와 함께 각 속성에 대한 감성 극성(positive(0)/ negative(1)/neutral(2))을 모델에 넘겨주면 각 속성에 대해 감성 예측을 할 수 있도록 학습시킬 수 있다.

Table 2. Aspect F1 scores of Male cosmetics category test

Aspect	F1 score
가격	0.98296
기능/효과	0.89723
디자인	0.98039
발림성	0.96470
보습력/수분감/쿨링감	0.96363
사용감	0.98924
성분	0.98630
용기	0.94680
용량	0.96470
유통기한	0.97637
윤기/피부(톤)	0.81720
자극성	0.88030
제품구성	0.90506
제형	0.95652
지속력/유지력	0.96794
편의성/활용성	0.90346
품질	0.93129
피부타입	0.94308
향	0.99004
흡수력	0.93854

Table 3. Sentiment F1 score of Male cosmetics category

Item	Value
F1 score	0.89447

#### IV. Experiments

Test 데이터 셋과 BEST 모델을 기반으로 F1 score를 측정한 결과는 Table 2, 3과 같다. 실제로는 4가지 MainCategory 데이터를 모두 본 연구에서 활용했지만, 표를 삽입할 공간 부족으로 인해 여기 신지 못하는 관계로 Figure 2, 3에 나타나는 ‘남성화장품’ 카테고리의 결과만을 적었다. 남성화장품의 경우 대부분 F1 Score가 크게 낮은 속성은 없었다. 전처리 하기 이전에는 개수가 적은 속성은 특히나 성능이 크게 떨어지는 모습을 보였는데, 제거하여 문제를 해결했다. 다만 감성 데이터와 관련하여 F1 Score는 0.9를 넘지 못했다. 예측한 값을 Section 4의 csv 결과 파일에서 확인해봤을 때, 대부분의 경우 정답 Aspects에는 없는 속성인데 추가로 존재한다고 예측하는 경우가 많아 csv의 행 개수가 정답보다 많았으며 이것이 점수에 영향을 미쳤을 것이라 생각된다.

모델 학습 및 테스트 시 성능을 F1 score로 확인했는데, 본 연구의 목적에 맞게 Dashboard를 제작해 모델의 예측 결과를 시각화하고자 하였다. 학습을 통해 저장된 BEST 모델에 Test 데이터셋을 넣어 모델이 예측한 속성 감성 극성 값과 테스트 데이터셋에 사전에 라벨링 되어있는 정답 속성, 감성 극성 값을 활용하여 비교할 수 있었다. 대시보드에 게재하기 위해 다음과 같은 추가 전처리 작업을 수행하였다. 먼저 감성 예측 결과 워드 클라우드를 위해 Python Kivipyepy

모듈을 사용하여 리뷰 텍스트로부터 보통 명사(NNG)와 고유 명사(NNP)만을 추출해냈으며, 해당 리뷰 텍스트에 대해 예측한 감성 극성을 함께 가져오도록 했다. 그리고 속성 예측 결과를 위해서는 단순히 리뷰 텍스트를 기준으로 id를 매기고 정답과 예측 값을 그대로 가져오도록 했다. 이렇게 정제된 데이터는 모두 csv로 저장되도록 하였다.

이로써 생성된 csv 파일을 시각화하기 위해 Google Sheet에 수동으로 연결했으며, 이는 Preset 대시보드에서 사용할 DB로 활용했다. MainCategory마다 별도로 생성된 csv 파일은 Preset SQL LAB에서 UNION 쿼리를 날려 Aspect Correct, Aspect Prediction, Sentiment Word Cloud 3개의 테이블로 묶었다. 생성된 테이블을 바탕으로 표 차트와 워드 클라우드 차트를 Figure 1, 2와 같이 생성하였다. Figure 1은 Preset의 대시보드 필터를 사용하여 MainCategory를 ‘남성화장품’, 리뷰 id를 257로 설정한 것이며, Figure 2는 예측한 감성 극성에 따라 워드 클라우드 차트를 3개로 나눈 것이다. ‘제품’, ‘사용’, ‘향’, ‘피부’와 같은 명사의 경우 공통적으로 자주 등장하였지만, 3가지 차트마다 나타나는 명사의 감성에 다소의 차이가 있음을 확인할 수 있었다.

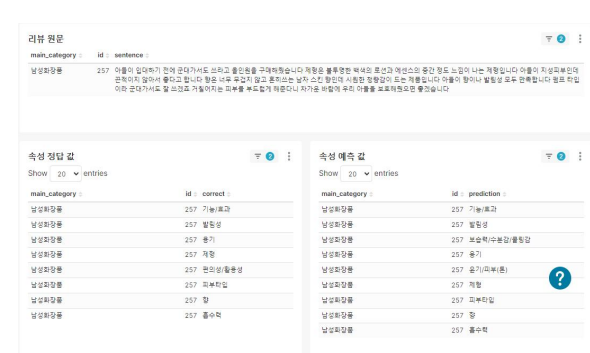


Fig. 1. Aspect dashboard



Fig. 2. Sentiment dashboard

#### V. Conclusions

본 연구는 KLUE RoBERTa base 모델을 사용함으로써 한글로 구성된 화장품 리뷰 텍스트에 나타나는 속성들을 추출하고, 이에 따른 감성을 예측하는 방법을 제시하였다. 이 연구의 한계점으로는 라벨링 데이터로부터 Train, Validation, Test 데이터 셋을 6:2:2의 비율로 분리할 때 속성과 감성이 고루 분포한 채로 나누었는지 확인하지 않고 실험이 수행되었다는 점이며, 골고루 분포하도록 조정해주면

모델의 성능이 더 향상될 것으로 예상된다. 한글 ABSA 선행 연구 사례는 적은 편이며, 특히 이 연구는 유의미한 시각적인 결과물을 도출하였으므로, 마케팅 분야의 소비자 의견 분석에 기여할 수 있을 것이다.

## ACKNOWLEDGEMENT

이 논문은 2020년도 정부(교육부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업임(No. NRF-2020R111A3053015). 이 논문은 2023년 대한민국 교육부와 한국연구재단의 지원을 받아 수행된 연구임 (NRF-2022S1A5A2A03052880)

## REFERENCES

- [1] H.H. Do et al. Deep learning for aspect-based sentiment analysis: A comparative review. *Expert Systems with Applications*, 118:272-299, 2019.
- [2] H. Shim et al. Aspect based sentiment analysis system of hotel review, reflecting user's preference. *Proceedings of the Korea Information Processing Society Conference*, 2018.05a:281-284, 2018.
- [3] Heejae Jung and Won Hee Lee. "building and evaluating korean datasets for aspect-based sentiment analysis in food review domain". In *한국정보과학회 학술발표논문집*, pages 1416-1418, 2021.
- [4] M.P. Geetha and D. Karthika Renuka. Improving the performance of aspect based sentiment analysis using fine-tuned bert base uncased model. *International Journal of Intelligent Networks*, 2:64-69, 2021.
- [5] Sungjoon Park et al. Klue: Korean language understanding evaluation, 2021.