

자사 IoT 가전 제품의 고객 반응 분석을 위한 비정형 텍스트(소비자 댓글) 공부정 분석에 관한 연구

정종진, 김경원, 김구환*

한국전자기술연구원

mozzalt@keti.re.kr, kwkim@keti.re.kr, guhwan@keti.re.kr

A study on the analysis of unstructured text (consumer comments) for customer response analysis of the company's IoT home appliance

Jung Jong Jin, Kim Kyung Won, Kim Gu Hwan*

Korea Electronics Technology Institute

요 약

본 논문은 IoT 가전제품을 생산, 유통, AS후속 조치까지 수행하는 중소 가전업체들이 자사의 제품이 판매되어 소비자에게 활용되는 시점에서 자사의 제품이 소비자들에게 어떤 평가를 받는지를 분석하여 차기 모델에 반영하거나 제품 경쟁력을 향상하기 위한 항목을 도출하기 위한 텍스트 분석에 대한 연구이다. 우선 자사의 제품에 대한 고객 평가가 많은 데이터 수집 채널을 선정하고, 해당 채널로부터 소비자 반응 비정형 형태의 글을 수집하여 텍스트 분석을 수행하고, 그 결과를 다양한 방식으로 시각화 전달하여 제품에 반영할 개선사항들을 알 수 있도록 도움을 주는 연구이다.

I. 서 론

기존에는 활용하지 않고 버려지던 데이터들에 대해서도, 또는 업무 관련자가 필요한 자료를 직접 검색하고 해석하기 때문에 과거에는 별도 저장할 필요도 없던 데이터들도 빅데이터 분석을 하면 좋은 결과를 얻을 수 있다 라는 막연한 기대로 다양한 루트를 통해 수집된 많은 데이터를 별도의 비용을 들여 저장/관리하는 사례들이 늘어나고 있다. 하지만, 막상 중소기업들 대부분은 여러 사정상 빅데이터를 목적에 맞게 정제하고 정확한 분석 모델을 설계하여 분석 엔진을 개발 후 시각화 하여 잘 활용할 수 있는 전문인력 및 인프라가 부족하다. 따라서 막상 언젠가 향후 활용코자 저장/관리하던 빅데이터를 효과적으로 활용하지 못 한 채 비용 지출만 늘어나 얼마 지나지 않아 유지하지 못하고 기존대로 담당자가 필요한 자료를 인터넷 검색 등으로 확인하고 업무에 반영하는 사례도 많다. 본고에서는 중소 가전업체 제품에 대한 소비자 반응 조사 및 제품 향상을 위해 다양한 채널에 존재하는 소비자 반응 댓글을 분석하여 제품향상에 필요한 인사이트를 도출하는 연구를 진행하였다[1].

II. 소비자 반응 분석

소비자 반응 조사는 소비자들이 카페, 블로그, 쇼핑몰 등에 입력한 구매 후기 글을 수집한 뒤 텍스트 처리, 자연어 처리, 토픽 모델링 기법들을 적용하여 소비자들이 제품에 대해 어떤 주제로 평가를 내리는지를 파악하고, 그 내용이 해당 주제에 대해 긍정 인지 부정적인지를 알아 내어 이를 제품 개선 또는 차기 모델 설계에 반영 할 수 있도록 도움을 주는 과정이다. 특히나 이러한 특정 주제에 대한 공부정 결과를 일정 기간을 부여해서 재 도출 하면 기간별 소비자의 공부정 변화 트렌드까지 알 수 있게 된다.

소비자 리뷰 공부정 분석을 수행하기 전, 분석 각 과정에서 분석 모형 학습, 검증 등에 활용하기 위해 소비자 리뷰 주제어/공부정 정답셋 생성 작업이 그림 1과 같이 필요하다[2].

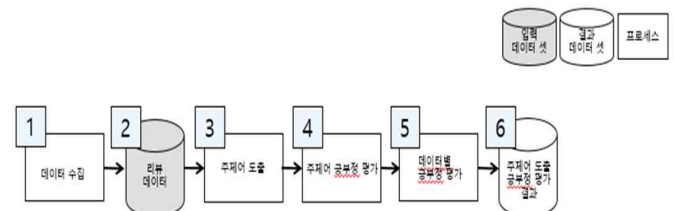


그림 1. 소비자 리뷰 주제어/공부정 정답셋 생성 절차

구축된 정답셋을 활용하여 소비자 반응을 분석하기 위해서는 기 수집된 제품별 소비자 리뷰 데이터를 활용하여 소비자 리뷰 공부정 분류 모델이 필요하다. 공부정 분석 모델 학습을 위해 네이버 쇼핑에서 수집한 제품별 소비자 리뷰 데이터를 이용하였으며, 최근 자연어 처리의 다양한 분야에서 좋은 성능을 보이고 있는 BERT 모형을 기반으로 공부정 분석 모델을 그림 2와 같이 설계하였다[3].

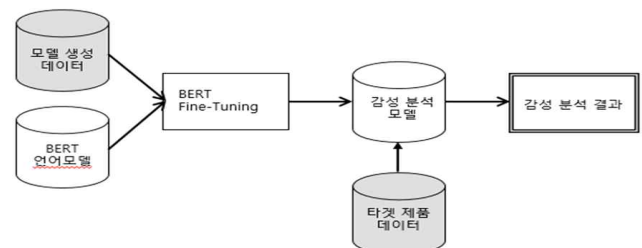


그림 2. 소비자 리뷰 공부정 모델(Bert) 구성 설계도

학습 데이터 생성을 위해 해당 제품에 대한 쇼핑 리뷰를 수집하여, 소비자 별점을 기반으로 긍/부정 레이블링을 통해 학습데이터를 구성하였다. 하

지만, 단일 제품과 관련된 소비자 리뷰는 BERT 학습을 수행하기에는 부족한 양의 데이터만이 존재하며, 더 큰 문제는 수집된 소비자 리뷰의 긍정 비율이 9:1 이상으로 긍정 리뷰가 많은 현상이 발생하였다. 신뢰할 수 있는 학습 데이터를 생성을 위해 해당 제품이 포함된 범위를 확장하여 가전제품 카테고리의 소비자 리뷰를 수집하고, 수집 데이터에서 랜덤 샘플링을 통해 긍정 비율을 5:5의 학습용 데이터셋을 구성하여 학습을 수행하였다. 모델 학습 및 검증은 통해 유의미한 결과가 도출되었으며, 해당 제품의 긍정 모델 적용 가능성을 검증하기 위해 해당 제품 리뷰 데이터 추가 검증작업이 필요하다.

III. 소비자 반응 긍정성 분석 실험

1. 실험환경

IoT 물결레 청소기 제품을 대상으로 소비자 반응이 존재하는 온라인 채널로부터 소비자 댓글을 분석하여 해당 제품에 대한 평가 및 추세를 분석하여 시각화 하는 실험을 진행하였다. 실험에 활용된 데이터 구성은 표1과 같다.

표 1 소비자 반응 긍정성분석 실험환경

	수집채널	긍정 쇼핑 리뷰	부정 쇼핑 리뷰
데이터 구성	네이버쇼핑몰, 맘카페,	Rating 1,2에 속하는 리뷰	Rating 1,2에 속하는 리뷰
데이터 수	-학습 데이터 : 457,711 (80%) -검증 데이터 : 57,213 (10%)	286,082개 - Rating 1 : 134,092 - Rating 2 : 151,990	286,057개 - Rating 4 : 42,353 - Rating 5 : 243,704

2. 실험결과

실험 성능은 아래 3가지 척도를 Accuracy, Mathew's Correlation Coefficient (MCC), ROC curve를 사용하여 평가했다.

- Accuracy : 실제값과 예측값을 비교하여 올바르게 예측한 비율이다.
- Mathew's Correlation Coefficient (MCC) :

참 긍정(TP), 참 부정(TN), 거짓 긍정(FP), 거짓 부정(FN) 4가지 항목을 모두 고려하여 성능을 평가한다. -1 ~ 1 사이의 값을 갖는다. 1과 가까울수록 분류가 잘 된 것이고, -1은 분류가 정반대로 되었을 경우, 0은 랜덤으로 분류한 값과 다름이 없다고 해석한다. 주로 분류 모델의 성능 평가에 사용한다.

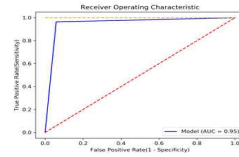
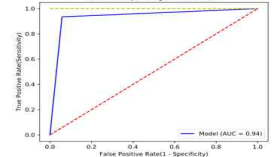
$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}$$

- ROC curve : 거짓 긍정 값이 변할 때, 참 긍정 값이 어떻게 변하는지를 보여주는 그래프다. 주로 모델 비교에 사용한다.

분석 모델 실험 결과, 모델 성능은 Accuracy 95%, MCC 0.91의 성능을 보였고, 타겟 제품 리뷰 데이터로 추가 분석한 결과는 표 2와 같이 Accuracy 88%, MCC 0.94의 성능을 보였다. 이는 가전제품 카테고리 학습된 소비자 리뷰 긍정성 모델을 해당 제품에 적용하여 분석에 활용해도 유사한 성능의 결과를 얻을 수 있음을 의미한다. 본 과정에서는 학습된 긍정성 모델을 이용하여 소비자 리뷰 기반 특정 제품 긍정성 분석 엔진을 구성하여 적용하였다. 또한, 학습된 긍정성 모델 적용시에 오분류된 소비자 리뷰를 정성적으로 분석했을 때, 오분류된 소비자 리뷰는 실제 리뷰 내용과 상이한 Rating을 부여한 데이터가 다수 있었으며, 이는 학습데이터의 긍정성 Labeling시에 소비자가 부여한 별점 Rating이 실제 리뷰내용과 다름을 의미하며, 본 학습모델의 긍정성 분류 예측이 실제로 더 정확하게 판단될 수도 있음을 반증한다. 따라서, 본 모델을 적용하여 예측된 긍정성

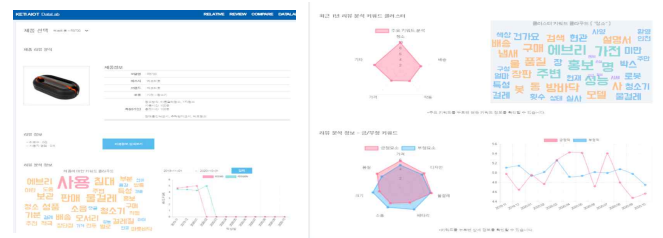
결과가 실험 결과보다 체감적으로 더 정확한 결과를 제공할 수 있을 것으로 예상된다.

표 2 모델 실험 결과

	모델 검증	추가 검증(해당 제품)
Accuracy	0.95	0.94
MCC	0.91	0.88
ROC		

3. 긍정성 결과 시각화 제공

소비자 반응 조사 시각화 제공에서는 자사 제품이 소비자들에게 어떤 평가를 받는지, 어떤 점에서 긍정 평가를 받고, 어떤 점에서 부정 평가를 받는지, 기간별로 평가가 어떻게 변화하는지를 알 수 있다.



III. 결론

본고에서는 대기업에 비해 빅데이터 분석 인프라, 인력 확보 면에서 상대적으로 열악한 중소기업체들이 빅데이터 플랫폼을 활용하여 사자의 업무 향상에 활용 수 있도록 지원하는 가전 빅데이터 분석 플랫폼을 소개하였다. 최근 빅데이터가 사회적 이슈가 되면서 빅데이터를 활용하고자 하는 마음에 데이터를 무조건 쌓고 있는 있지만 결국 분석에 활용하지 못해 방치되던 데이터들이 본고에서 소개된 플랫폼이 향후 기업 현장에서 활발한 서비스로 이어진다면 각 단계별 분석된 결과를 활용하게 되어 중소기업체들이 효율적인 기기 제조, 필요한 기능 추가, 불필요한 기능 제거 등을 수행하여 가격 경쟁력을 갖춘 우수한 제품 생산에 도움을 줄 수 있으리라 기대된다.

ACKNOWLEDGMENT

연구는 2021년도 산업통상자원부 및 산업기술평가관리원(KEIT) 연구비 지원에 의한 연구임('20000195'- 중소·중견 가전사의 IoT가전 제품개발 전주기 지원을 위한 빅데이터 상용화 플랫폼 기술 개발)

참 고 문 헌

- [1] 정종진,김경원,박종빈, “중소가전사의 제품 기획 지원용 빅데이터 분석 플랫폼 기술”, 한국통신학회지, vol. 37, No.8 2020.08
- [2] “텍스트마이닝의 이해”, (<https://iamdaisy.Tistory.com/29>), 2016
- [3] Hyo-Jun Yun et al. “Introduction of topic modeling for extracting potential information from unstructured text data: Issue analysis on news article of dementia-related physical activity”, Korean Journal of Sport Science, Vol. 30, No. 3, pp. 501-512, 2019