

타겟 제품 리뷰 데이터의 감성 분석에 관한 연구

*정승경 **김경원 ***정종진

한국전자기술연구원

cseunggyeong@keti.re.kr

A study on sentiment analysis of target product's review data

*Chung, Seung-Gyeong **Kim, Kyung-Won ***Jung, Jong-Jin

Korea Electronics Technology Institute

요약

일상의 데이터를 감성 분석에 사용할 때, 데이터 개수가 부족하거나 불균형한 문제가 발생한다. 이에 대해, 본 논문은 데이터 수집 카테고리를 확장하는 방법으로 기존의 문제들을 해결한다. 나아가 확장된 카테고리로 수집한 데이터를 이용해 감성 분석 모델을 만들고, 해당 카테고리에 속하는 타겟 제품에 대한 감성 분석을 시도한다.

1. 서론

감성 분석은 긍정 또는 부정과 같은 분류 기준별 데이터 개수가 비슷해야 학습이 제대로 이뤄진다. 개수가 불균형한 데이터로 학습한 뒤 검증해 보면 다수의 데이터가 존재하는 분류 값으로 편향되는 결과를 볼 수 있다. 따라서 감성 분석의 학습에서 분류 기준별 데이터 개수의 균형은 매우 중요한 요소임을 알 수 있다. 하지만 감성 분석에 사용할 현실의 데이터는 개수가 적거나 불균형한 특성을 보인다.

실제로 특정 제품에 대한 소비자의 반응을 감성 분석하기 위해 해당 제품의 리뷰 데이터를 수집한 결과 앞서 말한 2가지 문제점이 나타났다. 첫째는 수집한 데이터를 학습 데이터로 사용하기에 턱없이 부족했다. 우리가 타겟으로 한 제품의 경우 총 19,438개의 리뷰가 존재했다. 둘째는 부정적인 리뷰가 전체 리뷰의 1%로 분류 기준별 데이터 개수가 상당히 불균형했다. 우리가 타겟으로 한 제품의 경우 부정 데이터는 2,198개, 긍정 데이터는 18,370개 존재했다. 해당 데이터로 감성 분석해 보니 부정 데이터가 전혀 학습되지 않았고, 모두 긍정으로 분류했다.

본 논문은 위와 같은 문제점들을 해결하고, 특정 제품에 대한 소비자 반응을 감성 분석하는 방법을 제시한다. 이는 수집 카테고리의 범주를 확장하는 방법으로 전체 데이터의 개수를 늘리고, 분류 기준별 데이터 비율을 맞춰 기존의 문제점들을 해결한다. 그리고 해당 방법으로 쇼핑 사이트에서 리뷰 데이터를 수집하고, 감성 분석 모델을 생성하여 특정 제품에 대한 감성 분석을 시도했다.

2. 본론

특정 제품의 소비자 반응에 대한 감성 분석을 수행할 때, 해당 제품의 리뷰를 사용한다. 그런데 중소기업 제품의 경우 전체 리뷰 개수가 절대적으로 적고, 분류 기준인 긍정 또는 부정 값에 속하는 리뷰 수의 차이가 크다. 따라서 특정 제품의 리뷰는 감성 분석 모델을 학습시키기에 한계가 있다.

우리는 일반적으로 감정이 명사, 대명사보다 동사, 형용사에 의해 표현된다는 점에 주목했다. 따라서 하나의 카테고리 내에서 감정은 유사하게 표현되고, 감성 분석 모델은 감정을 표현하는 단어를 학습하여 분석을 수행할 것이라고 가정했다. 그리고 특정 제품이 포함되는 상위 범주의 카테고리를 이용해 데이터를 폭넓게 수집하여 데이터 개수 부족 및 불균형 문제를 해결했다. 이후, 해당 데이터를 학습하여 감성 분석 모델을 만들고, 특정 제품의 리뷰 데이터로 검증하여, 해당 모델을 특정 제품의 감성 분석에 이용할 수 있는지 알아보았다.

2-1. 데이터

감성 분석 모델을 만들기 위해 상위 범주의 카테고리 수집한 리뷰 데이터와 감성 분석을 하고자 하는 타겟 제품의 리뷰 데이터 2가지를 쇼핑 사이트에서 수집한다.

모델 생성 데이터

타겟 회사 제품의 상위 범주에 속하는 '물결레 청소기', '청소기', '가전' 3가지 카테고리로 수집한 리뷰 데이터이다. 긍정과 부정에 속하는 데이터를 동일한 개수로 A 쇼핑 사이트에서 수집했다.

데이터의 레이블은 A 쇼핑 사이트에서 사용하는 1~5 사이의 별점을 활용했다. 1과 2를 부정 데이터로 보았고, 별점 4와 5를 긍정 데이터로 보았다. 데이터는 총 572,139개로 별점 1은 134,092개, 별점 2는 151,990개, 별점 4는 42,353개, 별점 5는 243,704개로 구성되었다.

[표1] 모델 생성 데이터 구성

수집 사이트		A 쇼핑 사이트
데이터 개수	전체	572,139
	긍정	286,057
	부정	286,057

타겟 제품 데이터

타겟 회사 제품의 제품명으로 수집한 리뷰 데이터이다. 타겟 제품의 리뷰 데이터는 학습에 사용되지 않으므로 긍정과 부정에 속하는 데이터의 수를 동일하게 맞추지 않았고, A 쇼핑 사이트에서 수집했다.

데이터의 레이블은 A 쇼핑 사이트에서 사용하는 1~5 사이의 별점을 활용했다. 1과 2를 부정 데이터로 보았고, 별점 4와 5를 긍정 데이터로 보았다. 데이터는 총 2,482개로 별점 1은 554개, 별점 2는 687개, 별점 4는 186개, 별점 5는 1055개로 구성되었다.

[표2] 타겟 제품 데이터 구성

수집 사이트		A 쇼핑 사이트
데이터 개수	전체	2,482
	긍정	1,241
	부정	1,241

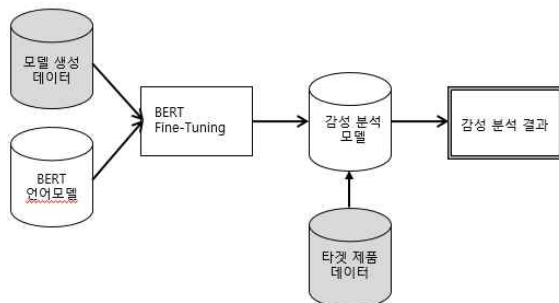
2-2. 실험 설명

본 실험은 BERT 언어모델을 감성 분석이 가능하도록 Fine-tuning 하는 방법으로 진행했다. BERT는 학습된 언어모델을 전이 학습시켜 NLP 작업을 수행한다는 점에서 장점이 있다고 판단해 선택했다. 우리는 신문기사와 백과사전으로 한국어에 대한 학습이 이뤄진 공개된 한국어 BERT 언어모델을 사용했다.

먼저, 감성 분석 모델을 만들기 위해 A 쇼핑 사이트에서 타겟 회사 제품의 상위 범주의 카테고리인 ‘물걸레 청소기’, ‘청소기’, ‘가전’으로 모델 생성 데이터를 수집한다. 모든 데이터는 nan값, 개행 문자, 특수문자가 제거되고 띄어쓰기가 보정된 형태로 처리해 사용한다. 인터넷 사이트에서 수집한 데이터는 띄어쓰기가 제대로 되지 않고, 다양한 특수문자가 존재하는 특성을 보인다. 그리고 이러한 특성은 학습 성능을 저하시키는 원인이 된다.

그리고 해당 데이터와 BERT 언어모델을 이용해 Fine-Tuning 하고, 감성 분석 모델을 생성한다. 해당 감성 분석 모델은 데이터를 긍정 또는 부정으로 분류한다. 또한, Fine-tuning 시에는 BERT 논문에서 제안된 하이퍼 파라미터 값을 사용했다. 따라서, Batch size 24, Learning rate 2e-5, epoch 2.0을 사용했다. [1]

이후, 생성된 모델은 특정 제품의 리뷰 데이터에 대한 감성 분석에 활용할 수 있는지 파악하기 위해 추가 검증 절차를 거친다. 이에, 타겟 제품의 리뷰 데이터로 감성 분석을 수행한다. 그리고 타겟 제품의 리뷰 데이터에 대한 감성 분석값이 결과로 도출된다.



[그림 1] 실험 절차

실험 성능은 아래 3가지 척도를 사용해 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

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

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

2-3. 실험 및 결과

A 쇼핑 사이트에서 ‘물걸레 청소기’, ‘청소기’, ‘가전’ 카테고리로 수집한 데이터를 학습 데이터에 80%, 검증 데이터에 10%, 테스트 데이터에 10% 비율로 나눠 사용했다. 분할된 데이터별 긍정과 부정 데이터 비율은 같다. 그리고 타겟 제품의 리뷰 데이터로 추가 검증했다.

그 결과 모델 성능은 Accuracy 95%, MCC 0.91의 성능을 보였고, 타겟 제품 리뷰 데이터로 추가 검증한 결과는 Accuracy 88%, MCC 0.94의 성능을 보였다.

[표3] 실험1 소요 시간 및 모델 성능 평가 결과

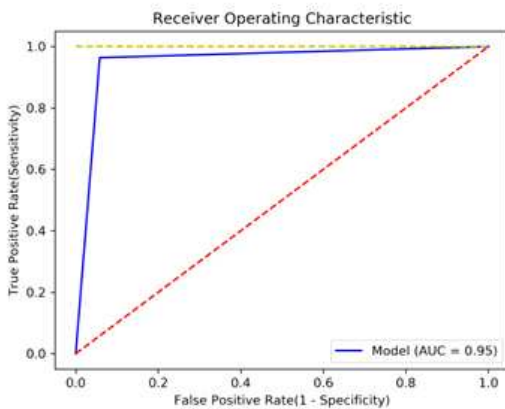
		모델	추가 검증
소요 시간		6h	1m
결과	Accuracy	0.95	0.88
	MCC	0.91	0.94

실험 결과 분석.

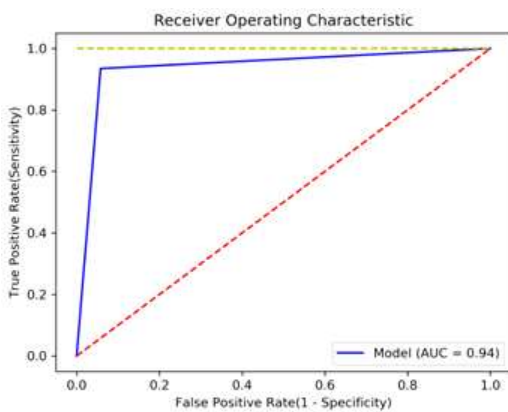
실험을 통해 감성 분석 모델과 타겟 제품 리뷰 데이터로 해당 모델을 추가 검증했을 때의 Accuracy, MCC, ROC 그래프 값이 모두 높았고, 신뢰할 만한 모델이라는 것을 파악할 수 있었다. 또한, 타겟 제품에 대한 소비자 반응을 감성 분석 할 때, 타겟 제품이 포함되는 상위 카테고리리를 이용해 수집한 데이터로 만든 감성 분석 모델을 이용할 수 있다는 사실을 확인했다.

참고 문헌

- [1] J. Devlin, et al., BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding, 2018.
- [2] A. Vaswani et al., Attention Is All You Need, 2017.
- [3] A. Minaee et al., Deep Learning Based Text Classification: A Comprehensive Review
- [4] C. Sun et al., How to Fine-Tune BERT for Text Classification



[그림 2] 실험 1 모델의 ROC curve



[그림 3] 실험 1 추가 검증 결과의 ROC curve

3. 결론 및 향후 방향

특정 제품의 소비자 반응에 대한 감성 분석 수행 시 전체 데이터의 개수와 분류 기준별 데이터 비율이 불균형한 문제가 발생한다. 이런 경우 해당 제품을 포함하는 상위 범주의 카테고리 데이터 수집하면 데이터 총 개수와 분류 기준별 데이터 비율을 맞출 수 있다. 또한, 상위 범주의 카테고리 수집한 리뷰 데이터를 이용해 감성 분석 모델을 만들고, 타겟 제품의 리뷰 데이터를 감성 분석할 수 있다는 점을 실험을 통해 확인할 수 있었다. 우리는 향후 해당 모델로 상위 범주의 카테고리에 속하는 다양한 제품 리뷰에 대한 감성 분석을 시도하고, 특정 카테고리 내의 범용적 감성 분석 모델을 만들고자 한다.

[ACKNOWLEDGEMENT]

본 논문은 2020년도 산업통상자원부 및 산업기술평가관리원(KEIT) 연구비 지원에 의한 연구임. ('20000195'-중소·중견 가전사의 IoT 가전 제품개발 전주기 지원을 위한 빅데이터 상용화 플랫폼 기술개발)