

BERT를 활용한 반려동물 사료제품의 감성분석 모델 개발

김영웅¹, 강다은², 이동규¹, 김건호², 윤지성¹, 김건우³, 길준민⁴

¹대구가톨릭대학교 컴퓨터공학전공

²대구가톨릭대학교 사이버보안전공

³(주)유스폴제스트

⁴대구가톨릭대학교 컴퓨터소프트웨어학부

ghkfud6545@gmail.com, ekdms5493@naver.com, ldk0450@naver.com,

kgh9941@naver.com, wltjd3169@naver.com, ceo@y-zest.com, jmgil@cu.ac.kr

A Development of Sentiment Analysis Model for Pet Feed Products using BERT

Young Woong Kim¹, Da Eun Kang², Dong Kyu Lee¹, Geonho Kim²,

Ji Seong Yoon¹, Geon Woo Kim³, Joon-Min Gil⁴

¹Major in Computer Engineering, Daegu Catholic University

²Major in Cyber Security, Daegu Catholic University

³Youthful Zest Co., Ltd.

⁴School of Computer Software Engineering, Daegu Catholic University

요 약

본 논문에서는 맞춤형 반려동물 사료제품 추천을 위해 최근의 자연어처리 모델인 KoBERT 모델에 기반하여 반려동물 사료제품에 대한 감성분석 모델을 설계하고 구현한다. 본 논문을 통해 구현된 반려동물 사료제품의 감성분석 모델은 정확도 평가에 대해서 비교적 우수한 성능을 보였으며, 학습과정에 참여하지 않은 새로운 반려동물 사료제품에 대해서 0.93 이상의 정확도를 산출하였다.

1. 서론

2020년말을 기준으로 우리나라에서 반려동물을 기르는 반려가구는 604만 가구로 전체 가구의 약 29.7%를 차지하고 있으며, 반려인은 1,448만 명으로 4명중 1명 이상이 반려동물과 함께 살아가고 있다 [1]. 이러한 반려가구의 증가로 인해 반려동물 시장도 급속한 성장을 보이고 있으며, 이제는 단순 사료 수준의 반려동물 제품보다는 기능성과 맞춤형 반려동물 제품의 출시 요구가 증대하고 있다. 그러나, 이러한 출시 요구에 따라 수많은 반려동물을 위한 사료 제품이 출시되고 있고, 소비자는 신제품의 홍수 속에서 자신의 반려동물에게 맞는 제품을 쉽게 찾기가 힘든 실정이다.

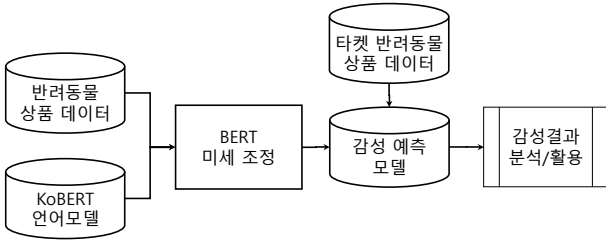
일반적으로 구매상품의 이력, 리뷰, 건강정보 등의 데이터는 온라인 커뮤니티나 쇼핑몰 사이트를 통해 실시간으로 축적되는 빅데이터의 특성을 갖고 있다 [2]. 이러한 빅데이터를 분석하여 반려동물 사료제품의 구매자 니즈 분석을 수행한다면 궁극적으로 반려동물 맞춤형의 제품 추천에 활용될 수 있을 것이다.

이러한 목적을 위해 본 논문에서는 자연어처리 기술을 활용하여 맞춤형 반려동물 사료제품 추천 모델의 설계 및 구현을 수행한다. 반려동물 커뮤니티 사이트에 게시된 리뷰글을 크롤링하며, KoBERT 모델 [3, 4]을 이용하여 반려동물 사료제품 리뷰글에 대한 감성분석을 실시하여 반려동물 사료제품의 긍정 및 부정 여부를 파악한다. 또한, 특정 반려동물 제품에 대한 실시간 감성분석을 위해 감성분석 예측 모델을 활용하여 감성분석을 실시하며, 궁극적으로 그 결과를 분석 및 가시화함으로써 맞춤형 상품 추천 서비스에 활용될 수 있을 것이다.

2. 반려동물 사료제품을 위한 감성분석

본 논문의 반려동물 사료제품을 위한 감성분석 모델은 (그림 1)과 같은 흐름에 따라 수행된다. 이를 위한 설계 및 구현을 위해 반려동물 사료제품의 리뷰글을 기반으로 감성분석을 수행하는데, 먼저 반려동물 사료제품의 리뷰글에 대한 크롤링 방법을 기술한다. 다음으로 반려동물 사료제품의 감성분석을 위한 KoBERT 언어모델에 기반한 학습 방법과 감성

예측 모델에 의한 감성분석 방법에 대해서 구체적으로 살펴본다.



(그림 1) 반려동물 사료제품의 감성분석 흐름도

2.1 크롤링

일반적으로 제품의 소비자 반응 등을 살펴보고자 할 때 인터넷 쇼핑몰이나 SNS에 게재된 해당 제품의 리뷰글을 사용한다. 본 논문에서는 우리나라의 대표적인 쇼핑몰 사이트인 네이버 쇼핑몰에서 총 28개의 반려동물 사료제품에 대해서 총 22,479개의 리뷰글을 웹크롤링을 통해 얻어내었다. 각 리뷰글은 사용자가 평가한 1에서 5점 사이의 별점이 부여되어 있다. 이 중에 부정 리뷰글은 별점 1과 2를 사용하였으며, 긍정 리뷰글은 별점 4를 사용하였다. 한편, 부정도 긍정도 아닌 중립을 나타내는 별점 3은 부정과 긍정의 명확한 학습을 위해 활용하지 않았다. 아울러, 별점 5는 이해당사자가 거짓으로 긍정 정보로 등재할 가능성이 있어서 제외하였다. <표 1>은 본 논문의 크롤링을 통해 생성된 리뷰글 데이터의 구성을 보여준다.

<표 1> 반려동물 사료제품 리뷰글 데이터 구성

항목	별점	개수	총 개수
부정	1	625	1,251
	2	626	
긍정	4	21,228	21,228

2.2 KoBERT를 활용한 감성분석

KoBERT는 한국어에 특화된 BERT(Bidirectional Encoder Representations from Transformers) 모델이다[3]. BERT 모델의 경우 한국어를 포함한 다국어를 위한 모델이지만, 상대적으로 한국어 비중이 낮아 한국어 언어처리에 성능이 원하는 만큼 좋지 않은 편이다. KoBERT는 한국어 데이터의 정확도를 높이기 위해 한국어 위키피디아 등을 학습시킨 모델이다. 본 논문에서는 KoBERT를 활용하여 반려동물 사료제품의 리뷰글에 대한 감성분석을 수행한다.

기존에 구축된 KoBERT 언어모델에 반려동물 사료제품에 대한 리뷰글을 추가로 학습하기 위한 미세

조정(fine tuning) 방식을 적용하였다. 즉, KoBERT 언어모델을 일반적인 한국어 문장과 단어의 언어모델에서 반려동물 사료제품에 대한 문장과 단어로 특화된 언어모델로 확대되도록 전이 학습(transfer learning)을 수행한 것이다. 본 논문에서는 전이학습된 KoBERT 언어모델을 사용하여 새로운 반려동물 사료제품에 대한 감성분석을 수행한다.

3. 실험

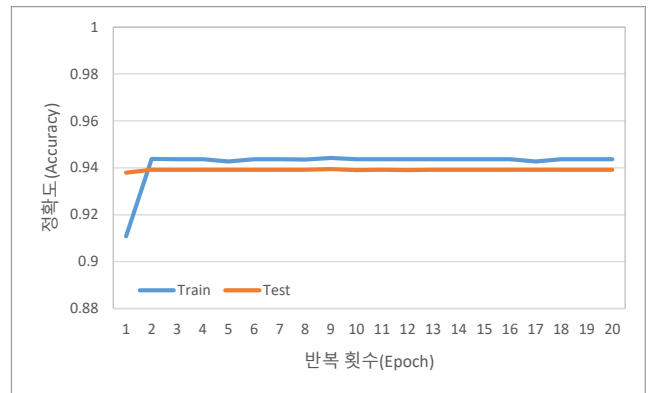
3.1 실험 환경

먼저 반려동물에 특화된 KoBERT 언어모델의 생성을 위해 앞서 (표 1)에서 제시된 데이터가 사용되었다. 총 22,479개의 리뷰글 데이터에 대해서 80%는 학습 데이터로 사용되었고 20%는 테스트 데이터로 사용하였다. 학습을 위한 배치 크기(batch size)는 32로 설정하였으며, 학습률(learning rate)은 5×10^{-5} 로 설정하였다 총 20번의 학습 반복(epoch)을 통해 학습을 진행하였다. PyTorch 1.12.1+cu113에 기반하여 학습 알고리즘을 구현하였으며, 학습 수행은 구글 코랩 환경에서 GPU 장치를 이용하였다. 한편, 긍정과 부정 리뷰글 데이터 각각은 서로 길이가 다르므로 32개, 64개, 128개 문자를 기준으로 리뷰글 데이터를 정규화하여 학습과 테스트를 진행하였다.

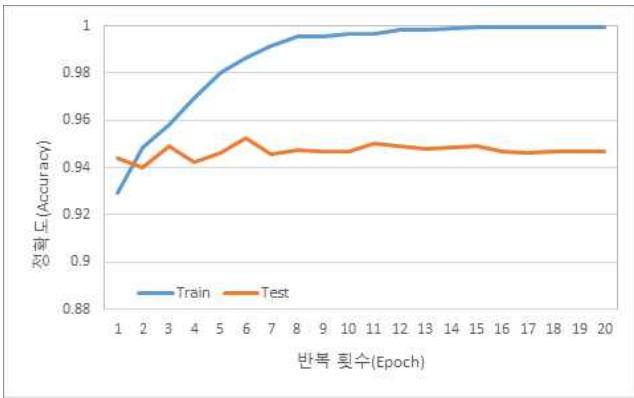
3.2 실험 결과

본 논문의 반려동물 사료제품의 감성분석 모델의 성능은 학습 반복 횟수에 정확도(accuracy)로 측정하였다. 여기서 정확도는 실제값과 예측값을 비교하여 올바르게 예측한 비율을 의미한다.

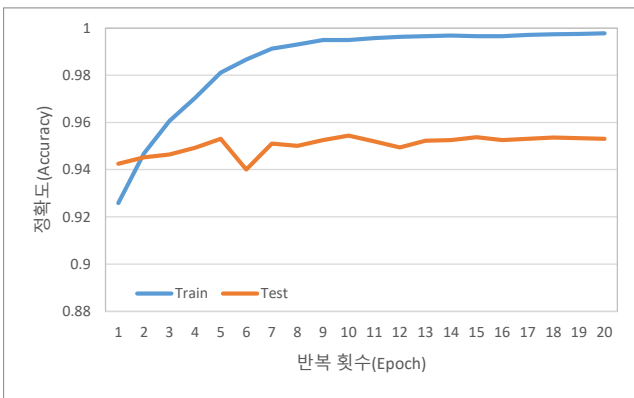
학습 데이터와 테스트 데이터를 나누어 정확도를 측정하였으며, 긍정과 부정 리뷰글의 길이의 3가지 종류로 나누어 정확도를 산출하였다. (그림 2)는 리뷰글의 길이에 따른 정확도를 나타낸다.



(a) 리뷰글의 길이가 32인 경우



(b) 리뷰글의 길이가 64인 경우



(c) 리뷰글의 길이가 128인 경우

(그림 2) 학습 반복 횟수에 따른 정확도 성능

(그림 2)의 결과는 리뷰글의 길이가 64와 128인 경우 반복 횟수가 증가함에 따라 학습 데이터에 대해서는 1.0 값으로 수렴하고 테스트 데이터에 대해서는 0.94에서 0.95 사이의 값을 가짐을 보여준다. 그러나, 리뷰글의 길이가 32인 경우는 학습 데이터에 대해서 1.0 값으로 수렴하지 못하고 0.95 값에 머무르고 있으며, 테스트 데이터에 대해서는 0.94 이하 값을 갖는다. 따라서, 리뷰글의 길이가 길어질수록 높은 정확도를 가짐을 알 수 있다. 이는 리뷰글의 길이가 길수록 학습 과정 중 리뷰글의 모든 부분이 학습에 참여할 가능성이 높아지기 때문인 것에 기인한다.

한편, 학습과 테스트 과정에 참여하지 않은 실제 반려동물 사료제품의 긍정과 부정 데이터를 검증 데이터로 활용하여 검증 실험을 실시하였다. 앞선 실험에서 리뷰글의 길이가 128일 때가 가장 좋은 정확도를 보였기 때문에, 이 경우에 생성된 모델에 검증 데이터를 적용하였다. <표 2>는 대표적인 2가지 반려동물 사료제품에 대한 리뷰글 개수와 정확도 성능

을 보여준다. <표 2>의 결과는 사료제품 모두에 대해서 0.93 이상의 정확도를 산출하였다. 이러한 결과는 새로운 사료제품 출시에 따른 리뷰글에 대해서도 우수한 정확도를 산출할 수 있음을 보여준다.

<표 2> 검증 데이터를 통한 성능 평가

제품명	리뷰글 개수	정확도
A 제품[5]	1,281	0.93
B 제품[6]	589	0.94

4. 결론

본 논문에서는 자연어처리 기술을 활용하여 맞춤형 반려동물 사료제품 추천 모델을 설계하고 구현하였다. 반려동물 사료제품의 구매 사이트에 게시된 리뷰글을 크롤링하며, KoBERT 모델을 활용하여 반려동물 사료제품 리뷰글에 대한 감성분석을 수행하였다. 성능평가 결과, 리뷰글의 길이가 길수록 비교적 좋은 정확도 성능을 보여주었으며, 새로운 사료제품에 대한 검증 실험에서도 0.93 이상의 정확도를 보여주었다.

한편, 본 논문에서 구현된 반려동물 사료제품 추천 모델의 성능은 정확도에 근거하여 측정하였다. 하지만, 최근에는 자연어처리 분야에서 이진 분류를 위한 측정 도구로 MCC(Mathew's Correlation Coefficient) 방법이 널리 활용되고 있다. 향후 MCC 방법을 활용하여 다양한 환경에서 성능 평가를 수행하고자 한다.

Acknowledgment

이 논문은 경상북도의 재원으로 2022년 경북형소프트웨어인력양성사업의 지원을 받아 수행한 연구임.

참고문헌

- [1] 황원경, 손광표, 2021 한국반려동물보고서, KB금융지주 연구보고서, 2021년 3월.
- [2] 김유영, 송민, “영화 리뷰 감성분석을 위한 텍스트 마이닝 기반 감성 분류기 구축”, 지능정보연구, 제22권, 제3호, pp. 71-89, 2016.
- [3] 전희원, 정승환, 김형준, 구글 BERT의 정석, 한빛미디어, 2021.
- [4] KoBERT, <https://github.com/SKTBrain/KoBERT>
- [5] 오리젠 펄피, <https://mypetlife.co.kr/10893/>
- [6] Now. 그레이프리 소몰브리드 어댑트, <https://www.eaglevet.co.kr/>