

감성 분석 기반의 시 추천 웹 서비스

유민경, 배병철

홍익대학교 게임학부

mingyeongyu8@gmail.com, byuc@hongik.ac.kr

A Poem Recommendation Web Service Based on Sentiment Analysis

Yu Min-kyeong, Byung-Chull Bae

Hongik University, School of Games

요약

본 논문은 네이버 영화 리뷰 데이터(NSMC)를 학습시킨 koBERT 언어 모델을 활용하여 사용자가 입력한 자신의 하루에 대한 감상평을 감성 분석하고, 이와 적합한 현대시를 제공하는 웹 서비스를 제안한다. 극성을 가진 NSMC 데이터를 변형해 모델을 학습시킨 후, 사용자 평가 결과 제공된 감성 분석에 대해 88%, 현대시에 대해 72%의 만족도를 보였다. 또한, 원인과 결과에 따라 같은 감성이라도 다른 감성으로 해석될 경우를 제외한다면, 잘못 분류되어 제공된 시더라도 다양하게 해석될 수 있는 시의 특징에 따라 다소의 오류가 있어도 사용자가 이를 자신의 상황에 맞게 해석할 수 있음을 확인했다.

1. 서론

감성 분석(Sentiment Analysis)에 대한 관심이 높아지며, 주어진 데이터셋으로 감성 분석의 성능을 높이는 연구가 활발히 이루어지고 있다[1][2]. 이러한 감성 분석 연구는 2000년대 이후 본격적으로 확대되어 영화평, 뉴스 기사 등의 데이터에 적용하여 선호도 분석 및 챗봇의 자연어 생성 등의 응용에서 사용되고 있다[3]. 이러한 연구에 요구되는 데이터셋은 일반적으로 1,000~10,000개 이상으로 구성되며, 의미있는 데이터셋 구축에는 긴 시간이 소요된다[4]. 따라서, 본 논문에서는 일반에게 공개되어 현재 다양한 용도로 사용되고 있는 네이버 영화 리뷰 데이터(NSMC; Naver Seniment Movie Corpus)[5]를 활용하고, 사전 학습시킨 감성 모델을 효과적으로 이용하여 사용자가 입력한 ‘자신의 하루’에 대한 감상평과 적합한 현대시를 제공하는 웹 서비스를 제안한다.

2. 웹 서비스 제공 및 평가

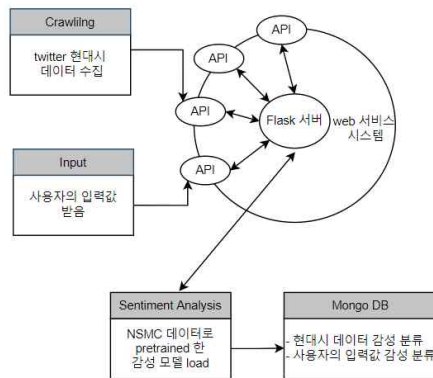
본 연구의 웹 서비스 시스템은 그림 1과 같다. 그림에서 koBERT 언어모델을 활용한 감성 분석 모델은 사용자가 입력한 자신의 하루와 크롤링한 시 데이터를 감성 분석할 때 사용된다. 저장한 감성 분석 모델의 가중치를 불러와 모델을 로드하면, 0~1 사이의 값인 predict score(예측값)가 출력된다. predict score에 따른 감성 분석 기준은 임의의 임계치에 따라 5가지(매우 긍정, 긍정, 보통, 부정, 매우 부정)로 정의한다.

제안한 시스템 아키텍처를 기반으로 사용자가 자신의 하루에 대해 간단하게 입력하면, 그에 대한 감성 분석 결과와 관련된 시를 한 편 추천할 수 있도록 하였고, 이를 위해 그림 2와 같은 간단한 웹사이트를 구축하였다.

향후 실제 서비스화를 위해 페이스북 <한국 인공지능 커뮤니티> 등을 통해 일반 사용자들로부터의 평가를 수집하였으며 (n=200), 만족, 보통, 불만족의 3가지 평가 기준에서

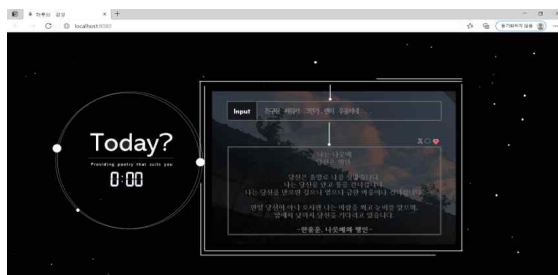
사용자가 불만을 선택했을 경우, 해당 현대시가 5가지의 감성 기준 중 새롭게 분류되어야 할 감성을 택하게 함으로써 해당 현대시의 라이브러리가 수정되게 하였다.

(그림 1) 시스템 아키텍처



(Figure 1) System Architecture

(그림 2) 감성 평가 및 시 추천을 위한 웹 서비스



(Figure 2) A Web Service for Sentiment Analysis and Poem Recommendation

사용자들의 평가를 분석한 결과, 감성에 대한 만족도는 88%, 현대시에 대한 만족도는 72%로 각각 나타났다. 예를 들면, 한용운의 <이별은 미의 창조>의 경우 ‘이별을 찬미하는 내용으로, 해석에 따라 5개의 라이브러리 중 ‘긍정’ 쪽에 해당할 수도 있지만, 감성 분석은 ‘부정’에 분류되어 있었다. 이처럼 해석에 따라 의도와 다른 값을 반환하는 경우라도, 사용자는 이를 긍정적으로 평가하여 라이브러리가 수정되지 않는 경우가 발생했다. 이는 다양한 표현 때문에 여러 의미로 해석될 수 있는 시의 특성 상, 독자가 자신의 상황에 맞춰 해석하고 이해할 수 있는 것으로 보인다.

반면, 수정된 라이브러리 중 감동에 의한 ‘슬픔’이 ‘부정’으로 분류되어 있어 불만족한 경우도 있었다. 이처럼 원인과 결과에 따라 같은 감성이라도 다른 의미를 지닐 때 분류의 모호성 경우가 다수 발생했으며, 이에 따라 사용자도 결과에 대해 불만족하는 상황이 발생하기도 했다.

3. 결론 및 향후 과제

본 논문에서는 네이버 영화 리뷰 데이터 (NSMC)를 기반으로 학습한 koBERT 모델을 이용하여 사용자 입력에 따른 감성 분석을 시행하고, 사용자 입력에 적합한 현대시를 출력하는 웹 서비스 시스템을 구현하고 간단한 사용자 평가를 시행하였다.

원인과 결과에 따라 같은 감성이라도 다른 감성으로 해석되었던 감동에 의한 ‘슬픔’과 상실에 의한 ‘슬픔’처럼 이를 다르게 해석할 수 있도록 시의 비유, 역설적 표현과 같이 다양한 표현에 대해 단순한 감성 분석외의 추가 연구가 필요할 것으로 보인다. 또한, 문맥을 고려한 분석 기준이 적용될 경우 다양한 기준으로 시 분류 또는 군집이 가능할 수 있다. 향후 연구로는 문맥 정보와 역설적 의미를 고려하여 감성 분석을 시행하고 이에 따른 시 분류 및 군집을 적용할 계획이다.

Acknowledgement

This work was supported by the National Research Foundation of Korea(NRF) grant (No. 2021R1A2C1012377) funded by the Korea government(MSIT)

참고 문헌

- [1] 서상현, 김준태, “딥러닝 기반 감성분석 연구 동향”, 한국멀티미디어학회지, 20권, 3호, 8-22, 2016
- [2] 오영택 외 2명, “Parallel Stacked Bidirectional LSTM 모델을 이용한 한국어 영화리뷰 감성 분석”, 한국정보과학회, 46권, 1호, 45-49, 2019
- [3] 이정훈, “감성분석 연구동향”, 한국정보처리학회, 358-361, 2018
- [4] Andrew Ng, “Machine Learning Yearning”, 7장, 2018
- [5] Naver Sentiment Movie Corpus (NSMC) <https://github.com/e9t/NSMC>