

ABSA 기반 영화 리뷰 감상 포인트 분석 시스템

서민석, 한재혁, 정인상, 황경호

국립한밭대학교 컴퓨터공학과

seominseok@edu.hanbat.ac.kr, hanjaehyeok@edu.hanbat.ac.kr, jeonginsang@edu.hanbat.ac.kr,
gabriel@hanbat.ac.kr

ABSA based point analysis system for movie reviews

Min-Seok Seo, Jae-Hyeok Han, In-Sang Jeong, Gyung-Ho Hwang

Hanbat National University, Dept. Computer Engineering

요약

속성 기반 감성 분석을 통해 실제 관람객이 작성한 영화 리뷰에서 감상 포인트를 분석한다. 리뷰 텍스트에서 모델을 통해 예측된 속성어에 따라 5개의 감상 포인트로 이를 분류하고, 감상 포인트별 긍·부정 퍼센티지 및 관련 리뷰 텍스트를 반응형 웹을 통해 제공한다. 모델 훈련을 위해 직접 구축한 한국어 데이터셋을 사용했으며, 모델은 SemEval 2014 Task에서 SOTA를 달성한 LCF-ATEPC를 사용하였다.

I. 서론

감성 분석(Sentiment Analysis)은 자연어 처리의 한 분야로서, 텍스트 내에 있는 의견, 평가 등의 주관적인 정보를 분석하는 기술이다. 전자상거래 시장이 성장함에 따라 많은 기업이 소비자들의 리뷰 데이터에 감성 분석 기술을 활용하여 상품 및 서비스에 대한 인사이트를 도출한다. 최근에는 텍스트에서 사용자의 의견이 들어간 속성어(aspect terms)를 추출하고, 각각의 속성(aspect)에 대한 감성을 분석하는 ABSA(Asspect-Based Sentiment Analysis, 속성 기반 감성 분석)에 대한 연구가 활발히 진행되고 있다.

본 논문에서는 네이버 영화 리뷰를 크롤링하여, 속성 기반 감성 분석을 사용해 5개의 감상 포인트(연출, 연기, 스토리, 영상미, OST)에 대한 관람객들의 평가를 분석하는 시스템을 개발하였다. 모델 훈련을 위한 훈련 및 검증용 한국어 데이터셋을 직접 구축하였으며, SemEval 2014 Task에서 SOTA(State-of-the-Art)를 달성한 LCF-ATEPC[2] 모델을 사용하였다.

II. 본론

2.1 데이터셋 구축

현재 국립국어원에서 제공하는 속성 기반 감성 분석용 데이터셋의 경우, 블로그 게시물 또는 SNS에 작성된 리뷰를 문장 단위로 줄바꿈 처리한 뒤, 문장 단위로 분할한 텍스트에 대해서 레이블링을 진행해 텍스트 형태의 다양성이 부족하다. 또한 은어, 비속어, 이모티콘 등이 빈번히 사용되는 영화 리뷰 텍스트와는 다르게 맞춤법이 정확하고, 정제된 리뷰들이 많아 실제 리뷰 텍스트와는 많은 차이가 존재한다.

이에 따라 본 논문에서는 실제 관람객이 작성한 네이버 영화 리뷰를 크롤링해, IOB(Inside-Outside-Beginning) 태그 방식으로 직접 데이터셋을 구축하였다.

2.2 Aspect Based Sentiment Analysis

기존의 감성 분석은 주로 문장 또는 문서 단위로 감성 극성을 예측하였기 때문에 <그림 1>과 같이 문장 내에 여러 개의 평가가 들어가 있

는 경우, 이를 제대로 분류하지 못한다. ABSA는 이러한 감성 분석 기술을 더욱 발전시킨 것으로, <그림 1>과 같이 문장이나 문서 내에 있는 여러 aspect에 대한 독립적인 감성을 예측할 수 있다. ABSA는 텍스트 안에 포함된 속성어를 추출하는 ATE(Asspect Term Extraction) Task와 주어진 속성어에 대한 감성을 예측하는 APC(Asspect Polarity Classification) Task로 구성되는데, 두 개의 태스크를 동시에 학습하기 위해 multi task learning 모델인 LCF-ATEPC를 사용하였다.

배우들 연기는 좋았으나, 스토리가 너무 뻔함

그림 1. 속성어가 여러 개 포함된 영화 리뷰 예시

2.3 LCF-ATEPC

본 논문에서 사용한 LCF-ATEPC 모델은 ATE, APC Task를 동시에 학습하는 multi task learning 모델이다. 모델 네트워크 구조는 <그림 2>와 같다. 이전까지의 연구들에서는 속성에 대한 감성을 예측할 때, global context(문장 전체의 맥락)만을 가지고 예측했지만, LCF-ATEPC에서는 local context(속성과 가까운 문맥 단어들과의 맥락)까지 고려함으로써 APC Task에서의 성능 향상을 이루어냈다. local context를 학습하기 위해, <그림 3>의 aspect sequence와 context sequence를 입력으로 받아, context 토큰과 aspect 토큰 쌍의 SRD(Semantic-Relative Distance)를 계산한다. SRD 값이 임계값보다 작은 context 토큰은 local context로 간주하며, global context와 local context는 각각의 BERT-Shared 레이어를 통해 임베딩 된다.

Local context의 경우, LCF(Local Context Focus) 레이어에서 CDM(Context-features Dynamic Mask), CDW(Context-features Dynamic Weighting) 레이어를 통해 문맥 단어의 특징을 보존하면서, 동시에 의미상으로 상대적인 단어의 피처를 마스킹하거나 가중치를 약화한다. 이후 MHSA(Multi-Head Self-Attention) 인코더에서 SDA(scale-dot attention) 연산을 통해 시맨틱 피처(semantic feature)를 추출한다. 최종적으로 생성된 local context와 global context 피처를 결합하는 Feature Interactive Learning 과정을 통해 APC Task를 학습하며, A

TE Task는 global context 피쳐만을 사용하여 학습한다.

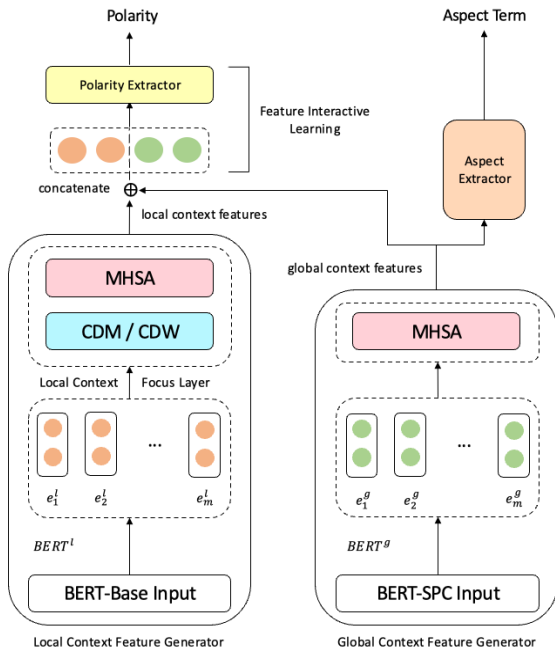


그림 2. LCF-ATEPC 네트워크 구조

$$s^t = \{w_i^t, w_{i+1}^t, \dots, w_j^t\}$$

$$s = \{w_0, w_1, \dots, w_n\} \quad (1 \leq i < j \leq n)$$

그림 3. aspect sequence(s^t), context sequence(s)

2.4 결과 확인 및 데이터 시각화

본 논문에서 개발한 시스템 구성도는 <그림 4>와 같다. 네이버 영화 사이트에서 관람객 리뷰를 크롤링한 뒤, 크롤링한 리뷰 데이터를 모델의 입력으로 전달한다. 이후 모델에서 예측한 속성어를 사전에 정의한 감상 포인트별 키워드에 따라 5가지 감상 포인트로 분류한 뒤 각 감상 포인트의 긍·부정 퍼센트를 계산한다. 이때 웹의 구성 화면은 <그림 5>, <그림 6>과 같다. 사용자가 상단의 검색바를 통해 원하는 영화를 검색하면, 해당 영화의 감상 포인트별 긍·부정 퍼센트를 바 차트 형태로 보여준다. 추가적으로 리뷰 보기 페이지 중앙에 있는 감상 포인트 버튼을 클릭하면, 해당 감상 포인트로 분류된 리뷰 텍스트를 보여준다.

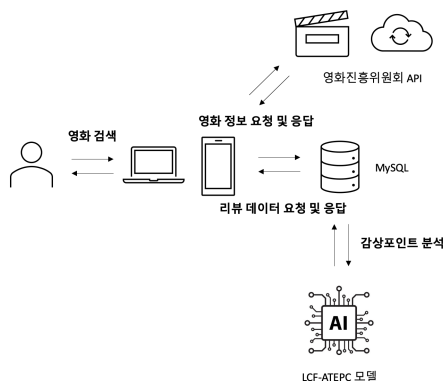


그림 4. 영화 리뷰 감상 포인트 분석 시스템 구성도

III. 결론

본 논문에서는 속성 기반 감성 분석을 통해 영화 리뷰의 감상 포인트를 분석하는 시스템을 개발하였다. 반응형 웹 기반으로 개발하여 다양한 환경에서도 쉽게 접근할 수 있으며, 감상 포인트별 실제 리뷰 텍스트를 제공하기 때문에 사용자가 영화에 대한 상세한 정보까지 알 수 있다. 현재는 미리 정의한 키워드에 따라 감상 포인트를 분류하였지만, 추후 군집화를 통해 분류를 진행할 예정이며, 모델 성능 및 속도 개선, 웹 UI 개선 등을 진행할 예정이다.

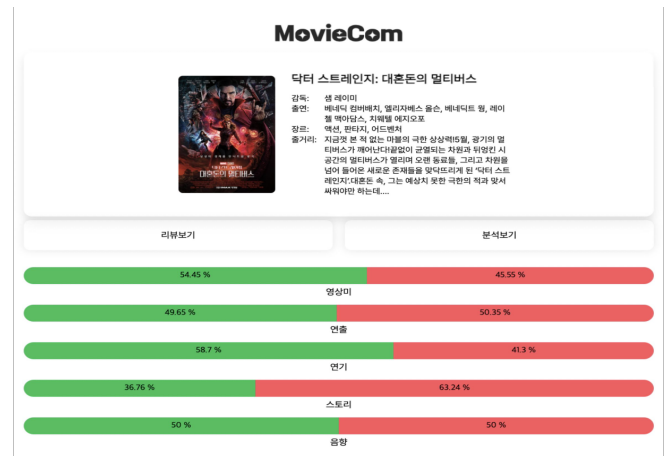


그림 5. 감상 포인트별 긍·부정 퍼센트

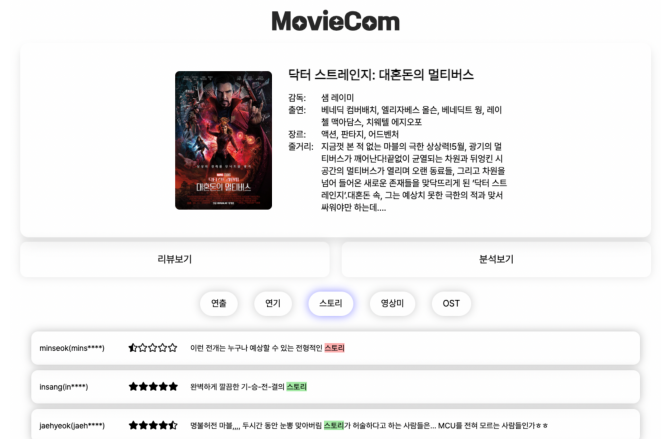


그림 6. 감상 포인트별 상세 리뷰

참 고 문 헌

- [1] Yang, Heng, and Ke Li. "PyABSA: Open Framework for Aspect-based Sentiment Analysis." arXiv preprint arXiv:2208.01368, 2022.
- [2] Yang, Heng, et al. "A multi-task learning model for chinese-oriented aspect polarity classification and aspect term extraction." *Neurocomputing*, vol.419, pp.344-356, Jan. 2021.
- [3] Zeng, Biqing, et al. "Lcf: A local context focus mechanism for aspect-based sentiment classification." *Applied Sciences*, vol. 9, no.16, 2019.