

2020 국어 정보 처리 시스템 경진대회 – 지정분야 2차 보고서

BERT 기반 Aspect-based Sentiment Analysis를 이용한 한국어 영화 리뷰 분석 시스템

2020-09-28

Team 리프 – 최찬호, 고대훈

자연어 처리 분야는 딥러닝 기술과 결합되어 비약적인 발전을 이루었습니다. 최근에는 BERT (Bidirectional Encoder Representations from Transformers) 와 같이 사전 훈련을 이용하는 언어 모델이 제안되어, 다양한 자연어 처리 분야에서 State-of-the-Art 성능을 달성하고 있습니다.

감성 분석 분야에서도 BERT 기반의 언어 모델은 강력한 성능을 보여줍니다. SKTBrain에서 개발한 KO-BERT는 Naver Sentiment Movie Corpus 데이터에 대하여 90.1%의 정확도로 긍정/부정을 구분합니다. 자연어 전처리에 있어서도 World-Piece Model, BPE Algorithm 등을 결합한 범용 전 처리 기술이 개발되어 비교적 쉽게, 고성능의 자연어 처리 모델을 구현할 수 있게 되었습니다. 저희는 이러한 최신 자연어 처리 기술들을 활용하여, 데이터로부터 새로운 가치를 창출하는 감성 분석 시스템을 개발하고자 합니다.

영화 리뷰와 같은 데이터는 긍정-부정 감성 인식 분야에 적극적으로 이용되었습니다. 영화 리뷰 말뭉치는 해당 영화에 대한 주관적 감성과 높은 연관이 있으므로, 리뷰 작성자의 평점 정보를 이용하여 긍정-부정 학습 구조를 설계할 수 있었습니다. 하지만 저희는 영화 리뷰와 같은 말뭉치 데이터에 다른 관점으로 접근합니다.

영화 리뷰는 단순히 "영화"라는 단일 대상에 대한 감성 정보만을 갖는 것이 아닙니다. 영화라는 대상을 중심으로, 다양한 속성(Aspect)에 대한 사용자들의 의견 정보를 가지고 있습니다. 저희는 영화 리뷰에서 속성(Aspect) 단위의 사용자 의견 정보들을 추출하여, 활용하는 시스템을 개발하였습니다.

속성 단위 감성 분석(ABSA: Aspect-based Sentiment Analysis)은 말뭉치에 포함된 다양한 개체들을 대상으로, 각 개체들에 대한 감성 정보를 분석하는 자연어 처리 기술입니다. 최근에는 BERT와 같은 사전 학습 자연어 모델과 결합한 연구들이 활발하게 진행되고 있습니다. 하지만 아직 속성 단위 감성 분석에 대한 딥러닝 구조를 학습할 수 있는 한국어 데이터셋은 충분하지 않은 상태입니다.

해당 시스템에서는 속성 단위 감성 분석을 수행하는 딥러닝 모델을 학습시키기 위하여, 속성(Aspect) 단위의 긍정-부정 정보가 포함되어 있는 10,000개의 화장품 리뷰 데이터를 활용합니다. 이는 일반적으로 목표 딥러닝 모델을 학습시키기에는 충분한 데이터가 아닙니다. 이러한 한계를 극복하기 위하여 BERT의 강력한 자연어 처리 능력과 새로운 모델 구조, Data Augmentation 기법을 활용하였습니다.

결과적으로, 영화 리뷰 데이터에 대해서도 분석 능력을 가지는 Aspect-based Sentiment Analysis 딥러닝 모델을 학습하였으며, 이를 기반으로 영화 리뷰 분석 시스템을 개발하였습니다.

1-1. 시스템 개요

해당 시스템은 딥러닝 기반의 자연어 처리 모델을 이용하여 영화 리뷰 말뭉치에 대해서 속성 단위 감성 분석(Asspect-based Sentiment Analysis)을 수행합니다. 단, 말뭉치로부터 속성 추출 (AE: Aspect Extraction) 작업은 수행하지 않고, 감성 분류 (ASC: Aspect Sentiment Classification) 작업만을 수행합니다. 곧, 미리 지정된 속성에 대해서 감성 분석을 수행합니다. (Targeted Aspect-based Sentiment Analysis Model)

자연어 처리 모델은 BERT (Bidirectional Encoder Representations from Transformer)를 기반으로 작동합니다. BERT 는 대규모의 라벨 없는 데이터들로 사전 학습(pre-training)을 수행한 후, 미세 조정(fine-tuning)하는 방법으로 다양한 자연어 처리 문제를 해결합니다. 해당 시스템은 사전 학습된 BERT 모델을 사용하기 위하여 SKT-Brain 에서 개발한 오픈 소스 프로그램 KO-BERT (<https://github.com/SKTBrain/KoBERT>)를 이용하였습니다. KO-BERT 기반의 감성 분석 모델은 Naver Sentiment Movie Corpus 데이터에 대한 긍정/부정을 90.1% 정확도로 구분합니다.

개발된 프로토타입 프로그램은 웹 크롤러를 이용하여 다음 영화 (<https://movie.daum.net>)로부터 영화 리뷰 말뭉치들을 수집합니다. 또한, 수집된 영화 리뷰 말뭉치를 바탕으로 속성 단위 감성 분석을 수행합니다. 결과적으로 개별 영화에 대한 사용자 의견 정보를 추출하고 이를 요약하여 활용합니다.

1-2. 감성 분석 문제 정의

해당 시스템은 영화 리뷰 말뭉치에서 사용자 의견 정보를 추출하기 위하여 지정된 속성들에 대하여, 속성 단위 감성 분석을 수행하고자 합니다. 이는 영화 리뷰 데이터에 대한 Target Dependent Sentiment Analysis 또는 ABSA (Aspect-based Sentiment Analysis) 모델을 학습하여 달성할 수 있습니다.

하지만, 속성 단위 감성 분석(ABSA: Aspect-based Sentiment Analysis) 모델을 학습시킬 수 있는 한국어 말뭉치는 충분하지 않습니다. 저희는 자연어 처리 모델 학습을 위하여, 한국 경영 학회의 온라인 상품 평 데이터 (<http://www.drbr.or.kr/>)와 Naver Sentiment Movie Corpus 데이터 (<https://github.com/e9t/nsmc>)만을 이용하였습니다.

한국 경영 학회에서 제공하는 온라인 상품 평 데이터는 속성 단위 감성 분석을 위하여 구성되었으나 화장품에 관련된 것으로, 영화 리뷰와는 다른 속성들에 대한 말뭉치 데이터입니다. 또한 Naver Sentiment Movie Corpus 데이터는 속성 단위로 구성되어 있지 않고, 문장 전체에 대한 긍정/부정 정보만을 가지고 있습니다. 영화 리뷰에 대해서 속성 단위 감성 분석 모델 학습을 수행할 수 있는 데이터는 확보하지 못하였습니다.

곧, 목표로 하는 속성 단위 감성 분석 모델은 학습 데이터와 평가 데이터가 다른 영역(domain)을 가지고 있습니다. 목표 모델은 위에 언급된 데이터들로 학습되지만, 영화 리뷰 말뭉치로부터 속성 단위 감성 분석 능력을 갖추어야 합니다.

1-3. 평가 데이터

영화 리뷰 말뭉치로부터 속성 단위 감성 분석을 수행하는 모델을 학습시키고자 합니다. 이 때, 성능 척도로 사용할 수 있는 평가 데이터를 직접 구성하였습니다. 평가 데이터는 Git-hub 에 공유되어 있습니다. (https://github.com/chanho096/Sentiment_Analysis)

구성된 데이터는 Naver Sentiment Movie Corpus 데이터의 Test-Set 을 기반으로 작성되었습니다. 지정된 속성 9 가지 중 2 가지 이상이 평가된 리뷰 1458 개를 추출하여, 각각의 속성에 대한 긍정/부정 정보를 입력하였습니다.

총 1458 개의 평가 데이터는 감성 분석 모델 학습으로는 사용되지 않았으며, 학습된 모델의 분석 능력을 평가하는 Validation-Set 또는 Test-Set 으로 사용되었습니다.

1458 개의 평가 데이터는 216 개의 대립 사례 데이터와 1242 개의 일치 사례 데이터로 구분됩니다. 일치 사례 데이터는 말뭉치에 언급된 모든 속성들이 일치하는 긍정/부정 값을 가지는 경우를 의미합니다. 대립 사례는 일치 사례가 아닌 경우로, 말뭉치에 언급된 속성들 중 하나 이상이 다른 긍정/부정 정보를 가지고 있습니다.

일치 사례 - 영성한 액션에, 시나리오는 왜그러지...

대립 사례 - 스토리가 좀 아쉬운건 있지만 감성적으로 본다면 아름답고 슬픈영화

1-4. 기존 연구

BERT 모델은 대규모의 라벨 없는 데이터들로 사전 학습 (pre-training)을 수행 후, 목적에 맞게 미세 조정(fine-tuning)하는 방법으로 여러가지 자연어 처리 분야에서 State-of-the-Art 성능을 달성했습니다. 이러한 BERT 를 활용한 ASC (Aspect Sentiment Classification) 작업 또한 연구되었으며, 높은 성능을 달성하였습니다. ([4] Attentional Encoder Network for Targeted Sentiment Classification)

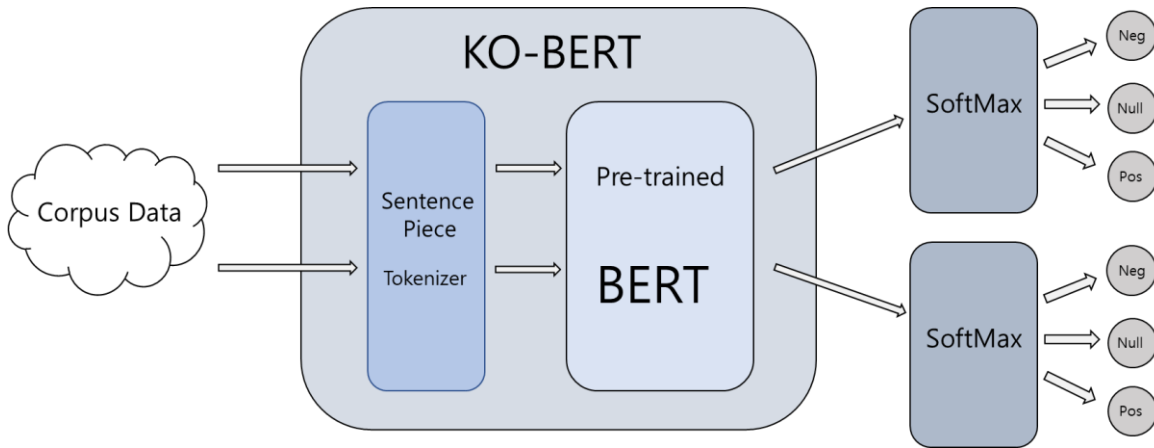
기존에는 BERT 의 SPC (Sentence-Pair Classification) 구조를 이용하여 ABSA 모델을 구성하였습니다. BERT 의 첫번째 문장으로는 분석하고자 하는 문장이 주어지며, 두번째 문장으로는 분석하고자 하는 속성이 주어집니다. ABSA 모델은 BERT 의 SPC 결과를 활용하여 문장으로부터 주어진 속성에 대한 감성을 분석합니다.

말뭉치 데이터 - 두 개 샀는데 둘 다 색상 예쁘네요

BERT-SPC 입력 - [CLS] 두 개 샀는데 둘 다 색상 예쁘네요 [SEP] 색상 [SEP]

BERT-SPC 구조의 ABSA 모델은 화장품에 관련된 온라인 상품 평 데이터에 대하여 95.1% 정확도로 속성 단위의 감성 분석을 수행하였습니다. 평가 데이터에서는 대립 사례에서 **64.79%**, 일치 사례에서 **84.18%**의 정확도를 달성하였습니다.

● ABSA Model with BERT – Masked Pair Training



2-1. 모델 제안

기존 BERT-SPC 기반 ABSA 모델의 학습 결과, 대립 사례에서 상대적으로 낮은 정확도를 가집니다. 대립 사례의 적중률이 낮고 일치 사례의 적중률이 높으므로, ABSA 모델이 속성 단위보다는 문장 전체에 대한 감성 분석을 학습한 것입니다. 저희는 대립 사례에 대해서 명시적으로 학습시켜주는 방법으로 이러한 문제점을 보완할 수 있다고 생각했습니다.

하지만, 기존까지의 ABSA 모델은 한 번에 하나의 속성에 대해서만 감성 분석을 수행하므로 여러가지 대립되는 속성들이 있는 사례를 명시적으로 학습할 수 없습니다. 저희는 대립 사례에 대한 명시적인 학습을 진행하기 위해서 다음과 같은 특징들을 가지는 새로운 모델 구조를 고안하였습니다.

● Pair Output

해당 시스템의 ABSA 모델은 한 번에 최대 두개의 속성에 대하여 감성 분석을 수행합니다. 곧, Output Layer 로 두 개의 SoftMax Layer 를 사용하여 각각 하나씩 서로 다른 속성에 대한 감성 분석을 수행합니다. 해당 시스템에서는 긍정/부정의 극성(polarity) 분류만을 목표로 하고 있지만, 특정 SoftMax Layer 가 감성 분석할 대상이 없을 경우를 고려하여 중립 출력을 추가하였습니다. (입력으로 한 개의 속성이 주어지는 경우, 하나의 SoftMax Layer 는 중립 출력을 가집니다.) 곧, 각각의 SoftMax Layer 는 긍정, 중립, 부정 중 하나의 출력을 가집니다.

● Masked Aspect

해당 모델이 Pair Output 구조를 가지므로, 입력으로 주어지는 각 속성들이 어떤 SoftMax Layer 에 의해서 분석되는지 순서를 정해줄 필요가 있습니다. 단순히 BERT-SPC 구조에 두 개의 속성을 입력으로 주는 방법으로는 개별 SoftMax Layer 가 분석할 대상을 적절히 지정해주지 못한다고 판단하였습니다. 따라서 저희는 각 SoftMax Layer 에 서로 다른 Mask 단어를 부여하여, 개별 SoftMax Layer 가 자신의 Mask 단어에 대한 감성 분석을 수행하는 구조를 고안하였습니다.

곧, ABSA 모델에 입력으로 주어지는 말뭉치의 Targeted Aspect 는 지정된 Mask 단어로 대체됩니다. 각 속성이 Mask 단어로 대체되므로, 개별 속성 단어에 종속된 감성 분석은 수행할 수 없습니다. 하지만, 문장 구조와 해당 단어의 위치 정보를 이용한 속성 단위의 감성 분석을 수행할 수 있습니다.

또한, Masked Aspect 구조는 모든 속성들을 동일한 구조로 취급하므로, 적은 수의 학습 데이터를 효율적으로 활용할 수 있습니다.

품질대란 **색상** 이쁘네 오
그라 해서 바르면 **지속력** 좀 떨어져요
배송 6일만에 받았어요 너무 오래 걸려요

⇒

품질대란 **측면** 이쁘네 오
그라 해서 바르면 **측면** 좀 떨어져요
측면 6일만에 받았어요 너무 오래 걸려요

지속력 대박 **가격** 싸게 잘 샀어요
배송은 빨리오는데, 절대 저런 **색상** 아니에요 ㅠ

⇒

측면 대박 **객체** 싸게 잘 샀어요
객체는 빨리오는데, 절대 저런 **측면** 아니에요 ㅠ

2-2. 모델 구조

모델의 입력 말뭉치 데이터는 분석하고자 하는 속성이 Mask 단어로 대체되어 있습니다. 입력 말뭉치는 Sentence-Piece Tokenizer 로 전 처리하여 BERT 모델 입력으로 사용됩니다. BERT 를 이용하여 추출된 특징(feature) 벡터들은 FFNN(Feed Forward Neural Network) 구조의 pooler 를 통과한 후, 각 SoftMax Layer 의 입력으로 사용됩니다. 최종적으로 각 SoftMax Layer 는 자신의 Mask 단어에 대한 분류(Classification) 작업을 수행합니다. 개별 SoftMax Layer 는 서로 독립적인 BR(Binary Relevance) 구조를 가지고 있습니다.

2-3. 데이터 증강 (Data Augmentation)

제안된 모델의 Pair Output, Masked Aspect 구조를 활용하여 입력 말뭉치에 다양한 데이터 증강 기법을 적용하였습니다. 특히, 기존 데이터들을 결합하여 새로운 대립 사례 데이터들을 생성하는 방법으로 대립 사례에 대한 명시적인 학습을 수행할 수 있었습니다.

배송 빨라서 좋아요
객체, 측면 빨라서 좋아요

⇒

지속력이 그다지입니다
배송 빨라서 좋아요
⇒ **객체**이 그다지입니다 **측면** 빨라서 좋아요

2-4. 감성 사전 학습 (Sentiment pre-training)

KO-BERT 모델을 바로 사용하여 ABSA 모델을 학습하지 않고, Naver Sentiment Movie Corpus 데이터를 이용하여 긍정/부정 극성 분류를 먼저 학습시킨 후 ABSA 모델을 학습시켰습니다. 이와 같이 BERT 모델을 감성 사전 학습 (Sentiment pre-training) 시켜서 ABSA 모델의 성능을 향상시킬 수 있었습니다.

최종적으로는 Naver Sentiment Movie Corpus 데이터를 이용한 SA(Sentiment Analysis) 모델과 해당 ASC 모델을 하나의 BERT 구조를 이용하여 동시에 학습시켜서 가장 좋은 성능을 얻을 수 있었습니다. (교차 학습)

2-5. 모델 성능

	대립 사례	일치 사례
BERT-SPC	64.79 %	84.18 %
BERT-SPC + 감성 사전 학습	54.09 %	94.80 %
Our Model	69.51 %	79.03 %
Our Model + 감성 사전 학습	71.46 %	87.26 %
Our Model + 교차 학습	73.40 %	92.02 %

평가 데이터 기준, 대립 사례와 일치 사례에 대한 정확도입니다.

새로운 모델 구조로 대립 사례를 성공적으로 학습시켰음을 확인할 수 있습니다. 결과적으로, BERT-SPC 기반의 모델보다 높은 성능의 속성 단위 감성 분석 모델을 학습시켰습니다.

추가적으로, 교차 학습된 모델은 Naver Sentiment Movie Corpus 데이터의 Test-Set 에서 87.05%의 정확도를 달성하였습니다.

3-1. 영화 리뷰 분석 시스템

영화 리뷰 분석 시스템은 DAUM 영화 (<https://movie.daum.net>)의 영화 URL 을 입력으로 받습니다. 주어진 URL 을 이용하여 해당 영화의 리뷰 데이터를 수집하고 (웹 크롤러 이용), 학습된 ABSA 모델로 속성 단위의 감성 분석을 수행합니다. 결과적으로, 개별 리뷰들이 영화의 각 속성(Aspect) 마다 어떤 평가를 내렸는지를 분석하고 해당 정보들을 종합합니다.

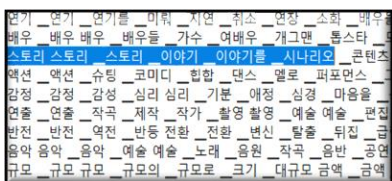
3-2. 영화 리뷰 속성 (Aspect)

영화 리뷰 데이터로부터 추출하고자 하는 속성들을 사전에 지정하였습니다. 지정된 속성들이 언급된 리뷰들을 추출하여, 감성 분석을 실행합니다.

➔ 연기, 배우, 스토리, 액션, 감정, 연출, 반전, 음악, 규모

이 때, 지정된 속성들의 유의어에 대해서도 감성 분석을 수행할 수 있도록 유의어 대체 사전을 구성하였습니다.

유의어들은 Sentence-Piece Vocabulary 에 존재하는 모든 단어들에 대해서 각 속성 단어와의 Cosine Similarity 를 비교하여 검출하였습니다. 이 때, 각각의 단어들로 임베딩 벡터를 생성하기 위하여 BERT 의 Embedding Layer 를 사용하였습니다. 이와 같은 방법으로 검출된 유의어 중, 영화 리뷰를 기준으로 각 속성들과 유사한 의미를 가진 단어들을 유의어 대체 사전에 추가하였습니다.



[Cosine Similarity 분석 결과]

3-3. 속성 기준 영화 요약

ABSA 모델의 속성 단위 감성 분석 결과를 이용하여, 영화에 가장 많이 언급된 요소들에 대한 감성 정보들을 요약합니다. 영화가 어떤 측면에서 긍정적인 평가를 받으며, 부정적인 평가를 받는지 정보를 얻을 수 있습니다. 이러한 속성 기준의 영화 요약 정보는 사용자에게 관심 영화를 추천하는데 사용될 수 있습니다.

```
### 영화 제목: [ 캐리비안의 해적 : 세상의 끝에서 ]
### 총 리뷰 개수: 1466
### 감성 분석 리뷰 개수: 201

### Top 3 Aspect: 영화 리뷰에 가장 많이 발견된 측면에 대하여 감성 분석
### 1. 스토리: 부정적 (63.64%)
### 2. 연기: 긍정적 (92.98%)
### 3. 액션: 긍정적 (60.61%)
```

3-4. 영화 추천 시스템

속성 기준의 영화 요약 정보를 활용하여, 실제 영화 추천 시스템을 구성하였습니다. 사용자가 관심 가지고 있는 속성을 기준으로, 영화 추천 서비스를 제공합니다. 영화 추천 시스템을 구성하기 위하여 1,138 개 영화에 대한 DAUM 영화 리뷰 1,061,210 개에 대하여 감성 분석을 수행한 후, 해당 정보들을 종합하였습니다.

영화 추천 시스템은 1,138 개 영화 각각에 대하여 점수를 계산 후, 높은 점수를 가지는 영화들을 추천합니다. 점수는 영화에 대한 긍정 평가, 영화에서의 관심 속성에 대한 긍정 평가, 영화에서의 관심 속성 비중 정보를 종합하여 계산합니다.

영화에 대한 긍정 평가 비율은 각 영화에 대한 리뷰 말뭉치에 긍정/부정 감성 분석을 수행하여 추정하였습니다. 영화에서의 관심 속성에 대한 긍정 평가 비율은 각 속성들이 언급된 리뷰 말뭉치에 속성 단위 감성 분석을 수행하여 추정하였습니다. 영화에서의 관심 속성 비중 정보는 각 속성들이 영화 리뷰에서 언급된 횟수를 이용하여 계산합니다. (산출된 각 비율들을 가중 조화 평균하여 추천 점수를 계산합니다.)

```
### Movie Recommendation: 관심 있는 속성을 기준으로 영화를 추천
### - 영화 리뷰 데이터 감성 분석 결과를 바탕으로 영화를 추천합니다.
### Aspect: [연기, 배우, 스토리, 액션, 감정, 연출, 반전, 음악, 규모]
### 검색 키워드 입력: 반전

### 전체 영화 개수: 1138개
### "반전" 관련 영화 개수: 273개

### 영화 추천 결과 [영화 리뷰 긍정 비율 / 관심 속성 긍정 비율]
[1]: 유주얼 서스펙트 [77.08% / 69.28%]
[2]: 헬로우 고스트 [82.54% / 77.72%]
[3]: 천사와 악마 [74.09% / 66.94%]
[4]: 데스노트 라스트 네임 [76.32% / 76.92%]
[5]: 브레이킹 던 part2 [69.72% / 62.25%]
```

3-5. 관심 속성 리뷰 검색

사용자가 관심 가지는 속성에 대하여 긍정 또는 부정 리뷰 데이터들을 검색하여 출력합니다. 속성 단위 감성 정보를 활용한 검색 서비스를 이용하여, 사용자는 관심 측면의 긍정 또는 부정 요소에 대하여 자세한 정보를 습득할 수 있습니다.

```

### Target Review: 중간 이상 숙성된 대장 미생 분해 효소 함유
### Aspect: [2기, 비유, 스토리, 역전, 감정, 전환, 반전, 음악, 규모]
### 전체 코드 입력:

### 스토리 관련 긍정적 리뷰 5개 분석 결과
### Review 1. (긍정적) "아민족도 다나미루어라 ~~~~~~ 정말 맛있고음 ~~~ ~~~~보면서 식욕 ~~~ 크로코넬지우어먹었다면알고나서는 ~~~~스토리도나름재미있었다"
### Review 2. (긍정적) "진짜 끝까지 웃음이 터져야 조니밥의 청담진진 캠퍼와 정말 상상초월의 스토리 캐릭터들의특색의 상상력의 끝은 어디인가?"
### Review 3. (긍정적) "1 2번이랑도 보았다 정말 재미있어요! 내년도 특목한 스토리 선전고과 스토리 정말 할만하네요! 정말재미!"
### Review 4. (긍정적) "스케일 대역구구 조니밥 캠퍼진진과 나다름 스토리 스토리도 나다름 스토리 재미있어! ~~~~보면서 ~~~~나미루어!"
### Review 5. (긍정적) "조니밥 동생들한테 놀아줄때는 끝까지 좀 지루한데 했지만 액제를 포함하는 부분부터 캐릭터들의 의욕이나 박진감있고 시원시원한 전개가 이어"

### 스토리 관련 부정적 리뷰 5개 분석 결과
### Review 1. (부정적) "캐릭터들이는 12345 모든 줄에 들어도 될 장면들을 너무 길게 늘려다 장면 전체 하절도 보며 조니 밥의 매력도 다 못"
### Review 2. (부정적) "웃기 무릎 잡을듯 9 2분도 넘게 지체되다나 새로운 캐릭터들은 말없이 제로 살아있는 역은 하나도 없고 말도 없는데 말도 없는데 재미있지 않"
### Review 3. (부정적) "이 설정이 이전 3부에서 완전히 다른 3부에서 보듯이 나다 말도 안하는 이 캐릭터들 인물상과도 하하하 하하하 완전 완전"
### Review 4. (부정적) "솔직히 너무 지루해요 12분이 넘게도 끝까지하고 전개가조니밥 나다 재미고 생생하다! ~~~~보면서 ~~~~즐거워! 중간중간 ~~~~했지만 ~~~~조니밥"
### Review 5. (부정적) "전반부 무대형도 2시간이 2시간이 복음 ~~~~하나는 내용이 3편에서나 ~~~~보면서 ~~~~아이가 ~~~~즐거워! ~~~~하나는 ~~~~지루한 ~~~~조니밥"

```

3-6. 사용자 지정 속성 분석

영화에 대하여 지정된 속성 이외에도, 사용자가 관심가지는 속성에 대한 정보들을 제공합니다. 영화에서 해당 속성이 어떻게 평가되었는지, 긍정적 또는 평가한 리뷰들은 어떻게 구성되어 있는지 관련 정보들을 요약하여 제공합니다.

```
### Custom Aspect: 서평자가 직접 지시한 언어 대안 감성 분석 결과 출력
### 감성 분석 수제 방법: "조니얼"
### "조니얼" 감성 분석 실행...

### "조니얼" 관련 감성 분석 결과
### 연관 리프 개수: 50
### 감성 지표: 긍정적 (78.69%)

### "조니얼" 관련 리프 목록의 5개 분석 결과 출력
### Review 1. [공정적] "조니얼ㅋㅋㅋㅋㅋㅋㅋ너무웃음있을뿐발수거없이그고열기네"
### Review 2. [공정적] "조니얼 연기 정말잘하고 재미도 있고 좋았습디다 끝부분이 좀여짜라던것 해곤 4인어
### Review 3. [부정적] "1번에 나온건 조니얼은 귀엽게 나왔던 반면 2분은 갸웃했는데 3분은 또 어드레 나옴지
### Review 4. [긍정적] "특종장황들은 참 잘맞아도 말하는데 너무 내용이 일관성 있지가 없어서 별로였어요조니얼"
### Review 5. [중립적] "최초로써 조니얼의 능력스런 연기도 끝나구마지막 액션씬에서 갈수록은 진짜 웃겼음"
```

3-7. 분석 수행 시간

영화 분석 시스템이 CPU 를 이용하여 1 개의 리뷰에 대하여 감성 분석을 수행하는데 평균 4ms 의 시간이 소요됩니다. GPU 를 이용할 경우 1630 개의 리뷰를 분석하는데 1.7s 의 시간이 소요되었습니다. 곧, 감성 분석 모델은 실시간 서비스에도 활용 가능한 분석 수행 시간을 갖고 있습니다.

영화 분석 시스템의 병목 구간은 웹 크롤러를 이용한 리뷰 수집 시간입니다. 현재 1630 개의 리뷰 데이터를 수집하는데 약 53s 의 시간이 소요됩니다. 실제 시스템 활용 시, 리뷰 데이터를 서버로부터 제공받는다면 병목 구간을 해소할 수 있습니다.

4-1. 소스 코드

프로토타입 프로그램과 모델 구현, 데이터 증강, 학습 알고리즘 등 시스템 재구현을 위한 소스 코드는 GitHub 에 공유되어 있습니다.
(https://github.com/chanho096/Sentiment_Analysis)

또한, 자체 제작한 평가 데이터 또한 해당 GitHub 에서 제공합니다.

5. 결론

감성 분석 기술을 이용하여 사용자 의견 정보를 추출하고 활용하는 시스템을 제안하였습니다. 해당 시스템은 속성 단위 감성 분석(ABSA)을 이용하여 영화 리뷰와 같은 대규모 말뭉치 데이터로부터 사용자 의견을 분석하고, 이를 활용하여 새로운 정보를 구성하였습니다. 이러한 데이터 재구성 작업은 영화 리뷰에만 적용되는 것이 아니라, 다양한 말뭉치 데이터에서 목표로 하는 요소에 대한 정보를 수집하고 활용하는데 사용될 수 있습니다.

목표로 하는 시스템과 개발 환경을 기준으로, 기존 연구보다 더 높은 성능의 ABSA 모델을 학습하였습니다. 기존 모델이 가지고 있는 문제점을 해결하기 위하여 새로운 모델 구조를 고안하였으며, 성능을 향상시킬 수 있는 다양한 방법을 제시하였습니다. 특히, 제약된 학습 데이터 조건 하에서도 속성 단위 감성 분석 모델을 학습시킬 수 있음을 보였습니다.

해당 감성 분석 기술을 활용한 사용자 의견 정보 추출 및 활용 시스템은 다양한 분야에서 응용되어 새로운 가치를 만들어 낼 수 있습니다. 그리고 이를 구현하기 위한 효율적인 모델 구조 또한 제안하였습니다. 더 많은 데이터를 활용하여 관련 모델에 대한 연구가 진행된다면 다양한 한국어 의견 정보들을 추출하고 활용할 수 있을 것입니다. 국립국어원에서 제공하는 말뭉치 데이터에도 한국어 관련 속성 단위 감성 정보 말뭉치가 추가된다면, 적극적으로 활용되어 한국어 정보화를 확대시킬 것입니다.

6. 참조 논문

[1] BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding, 2020, Jacob Devlin

[2] Fine-grained Sentiment Classification using BERT, 2019, Manish Munikar

[3] Adaptive Recursive Neural Network for Target-dependent Twitter Sentiment Classification, 2014, L Dong

[4] Attentional Encoder Network for Targeted Sentiment Classification, 2019, Y Song

[5] Utilizing BERT for Aspect-Based Sentiment Analysis via Constructing Auxiliary Sentence, 2019, C Sun