

CNN 보조 손실을 이용한 차원 기반 감성 분석*

전민진

한양대학교 파이낸스경영학과
(3alswls@gmail.com)

황지원

한양대학교 경영학부
(jiwonhwang11131@gmail.com)

김종우

한양대학교 경영대학 경영학부
(kjiw@hanyang.ac.kr)

텍스트를 바탕으로 한 차원 기반 감성 분석(Asspect-Based Sentiment Analysis)은 다양한 산업에서 유용성을 주목을 받고 있다. 기존의 차원 기반 감성 분석에서는 타겟(Target) 혹은 차원(Asspect)만을 고려하여 감성을 분석하는 연구가 대다수였다. 그러나 동일한 타겟 혹은 차원이더라도 감성이 나뉘는 경우, 또는 타겟이 없지만 감성은 존재하는 경우 분석 결과가 정확하지 않다는 한계가 존재한다. 이러한 문제를 해결하기 위한 방법으로 차원과 타겟을 모두 고려한 감성 분석(Target-Aspect-Sentiment Detection, 이하 TASD) 모델이 제안되었다. 그럼에도 불구하고, TASD 기존 모델의 경우 구(Phrase) 간의 관계인 지역적인 문맥을 잘 포착하지 못하고 초기 학습 속도가 느리다는 문제가 있었다. 본 연구는 TASD 분야 내 기존 모델의 한계를 보완하여 분석 성능을 높이고자 하였다.

이러한 연구 목적을 달성하기 위해 기존 모델에 합성곱(Convolution Neural Network) 계층을 더하여 차원-감성 분류 시 보조 손실(Auxiliary loss)을 추가로 사용하였다. 즉, 학습 시에는 합성곱 계층을 통해 지역적인 문맥을 좀 더 잘 포착하도록 하였으며, 학습 후에는 기존 방식대로 차원-감성 분석을 하도록 모델을 설계하였다. 본 모델의 성능을 평가하기 위해 공개 데이터 집합인 SemEval-2015, SemEval-2016을 사용하였으며, 기존 모델 대비 F1 점수가 최대 55% 증가했다. 특히 기존 모델보다 배치(Batch), 에폭(Epoch)이 적을 때 효과적으로 학습한다는 것을 확인할 수 있었다. 본 연구에서 제시된 모델로 더욱 더 세밀한 차원 기반 감성 분석이 가능하다는 점에서, 기업에서 상품 개발 및 마케팅 전략 수립 등에 다양하게 활용할 수 있으며 소비자의 효율적인 구매 의사결정을 도와줄 수 있을 것으로 보인다.

주제어 : 온라인 리뷰 분석, 차원 기반 감성 분석, TASD, CNN, BERT

논문접수일 : 2021년 8월 10일 논문수정일 : 2021년 11월 12일 게재확정일 : 2021년 11월 22일
원고유형 : 학술대회용 Fast Track 교신저자 : 김종우

1. 서론

감성 분석은 단어와 문맥을 분석하여 텍스트에 있는 감정을 파악하는 기술이다. 그 중 차원 기반 감성 분석은 특정 대상에 대하여 차원별로 감성을 분석하는 것으로, 타겟, 차원, 감성을 모두 분석하거나 차원, 감성만 분석하는(Asspect Sentiment Detection, 이하 ASD) 등, 목적에 따라

다양한 형태로 연구되고 있다(Wan et al. 2020). 여기서 타겟은 차원-감성 분석 시, 분석의 대상이 되는 단어나 구를 가리킨다. 일반적인 감성 분석이 텍스트가 가지고 있는 감성이 무엇인지를 분석한다면, 차원 기반 감성 분석은 텍스트에 나타나는 대상의 차원, 그에 따른 감성, 타겟이 무엇인지를 분석하는 연구이다.

차원 기반 감성 분석에서는 리뷰를 바탕으로

* 이 논문은 '대학혁신지원사업 비즈니스 랩 기반의 프로젝트 학기제'의 지원을 받아서 수행된 연구임.

어떠한 상품에 대해 차원별로 감성을 분석하는 것이 가능하다. 예를 들어, “스테이크가 맛있긴 하지만, 가격에 비해선 별로이다”라는 리뷰 문장이 있다면 일반적인 감성 분석에선 스테이크에 대해 긍정으로 분류하지만, 차원 기반 감성 분석에서는 ‘음식’이라는 차원에 대해서는 긍정적인 감성으로, ‘가격’이라는 차원에 대해서는 부정적인 감성으로 분류한다. 즉, 일반적인 감성 분석에 비해 보다 상세하게 텍스트의 감성을 분류할 수 있는 것이다. 이처럼 차원 기반 감성 분석은 리뷰를 바탕으로 특정 상품에 대해 차원별 점수를 산출할 수 있어 기업에서 다양하게 활용할 수 있다. 예를 들어, 프랜차이즈 사업자의 경우 사업장의 여러 가지 서비스 수준을 일정 수준 이상으로 유지하는 것이 중요한데, 해당 모델을 활용하여 다차원의 서비스 수준을 측정하여 주기적으로 모니터링을 할 수 있다. 또한, 타겟 추출까지 가능하기 때문에 서비스에 문제가 생겼을 경우 구체적인 원인을 찾아 개선하기에도 용이하다.

차원 기반 감성 분석을 하기 위해서는 우선 텍스트 안에 포함된 차원이나 타겟을 추출하고, 이에 대한 감성을 분석해야 한다. 지금까지 대부분의 연구에서는 차원 혹은 타겟만을 고려하여 감성을 분석해왔다(Wang et al. 2016; Xue and Li, 2018; Schmitt et al. 2018; Sun, Huang, and Qiu, 2019). 하지만, 감성을 분석할 때 차원과 타겟을 함께 고려하지 않을 경우 감성을 정확히 예측할 수 없는 경우가 발생한다. 즉 이에는 3가지 경우가 존재하는데, 각각 ‘같은 차원에서 상반된 감성이 존재하는 경우’, ‘같은 타겟에서 상반된 감성이 존재하는 경우’, ‘타겟이 없음에도 감성이 존재하는 경우’ 등이 발생할 수 있다. 예를 들어, 첫 번째 예시로는 “피자, 샐러드는 좋았지만, 스

테이크는 실망스러웠다.”와 같은 문장으로, 해당 문장 내에서는 ‘음식’이라는 같은 차원임에도 상반된 감성이 공존하게 된다. 두 번째 예시로는, “새우는 맛있었지만, 가격이 상당했다”와 같은 문장을 생각해볼 수 있다. 이 경우 같은 타겟인 ‘새우’에 대해 상반된 감성이 공존하게 된다. 마지막으로 “음식이 너무 늦게 도착해서 식어 있었다.”라는 문장은 타겟이 없지만(NULL) ‘서비스’라는 차원에서 부정적인 감성을 띄게 된다. 이와 같이 차원과 타겟을 모두 고려하지 않으면 이중 의존성(dual dependence)의 문제가 발생하게 된다(Wan et al. 2020).

이러한 문제점은 타겟과 차원을 동시에 고려해 감성을 분석하는 TASD를 통해 해결할 수 있다. 현재 TASD분야에서 성능이 높은 모델 중 하나는 Target-Aspect-Sentiment Joint Detection for Aspect-Based Sentiment Analysis(이하 TAS-BERT) 모델이다(Wan et al. 2020). TAS-BERT 모델은 BERT(Bidirectional Encoder Representation from Transformers)(Devlin et al. 2019) 기반 모델로, 리뷰의 각 단어 토큰(Token)을 사용해 타겟을 추출하고 BERT의 첫 번째 토큰을 사용하여 감성과 차원을 추출한다. TAS-BERT는 감성 분석 시에 타겟과 차원을 모두 고려하기 때문에 감성 분석을 위해 타겟과 차원에 대한 의존성(Dependence)을 포착할 수 있다(Wan et al. 2020). 그 결과, 감성 분석 분야에서 대표적으로 사용되는 데이터인 SemEval 레스토랑 데이터셋에서 타겟과 차원과 감성을 동시에 감지하는 부문인 SemEval-2015 Task 12(Pontiki et al. 2015), SemEval-2016 Task 5(Pontiki et al. 2016)에서 높은 성과를 보였다. 하지만 BERT 기반 감성 분석 모델의 특성상 지역적 문맥(Local context)을 포착해야 하는 경우에는 한계를 가진다.

이러한 한계점을 보완하고자 본 연구에서는 TAS-BERT 모델을 기반으로 합성곱을 사용해 지역적 문맥을 더욱 잘 포착하도록 하였다. 차원별 감성 분석에 사용되는 기존의 손실 함수에 합성곱 계층에서 얻은 보조 손실을 더하여 학습을 하는 방식을 채택하여 모델 학습을 진행하였다. 그 결과, 기존 모델에 비해 TASD 분야에서 작은 배치, 적은 에폭에서 정확도를 효과적으로 높일 수 있었다. 이를 통해, 한정된 자원으로도 단시간 내에 비교적 높은 정확도를 얻을 수 있다는 시사점을 제시한다.

본 논문의 구성은 다음과 같다. 1장을 통해 연구 배경과 필요성을 설명하고, 2장에선 본 연구와 관련된 연구들을 요약한다. 3장에서 본 연구에서 제안하고자 하는 모델에 대해 진행 단계별로 설명한다. 4장에서는 본 모델과 타 모델의 정확도를 비교한 실험 결과에 대해 서술하고, 5장에서 본 연구의 의의와 효과, 한계, 향후 연구 방향을 제시하며 본 논문을 마무리한다.

2. 관련 연구

2.1 BERT

BERT는 2018년 구글에서 발표한 범용 언어 모델이다(Devlin et al. 2018). BERT의 핵심은 대량의 언어 데이터를 사전 학습(Pre-Training)시키는 것으로, 이를 통해 자연어 처리(Natural Language Processing, 이하 NLP) 분야의 정확도를 크게 높일 수 있었다. BERT는 깊은 신경망(Deep Neural Network)의 최하단에서 시작하며, 왼쪽과 오른쪽 문맥을 모두 사용하는 양방향성을 가진다.

실무적인 관점에서 살펴보면, 본 모델은 지도 학습 모델로 텍스트 데이터를 연구 모델에 맞게 전처리하는 과정이 중요하다. 훈련 데이터셋이 없는 경우, 웹 크롤링을 통해 데이터를 수집하여 각 데이터마다 타깃, 차원, 감성을 라벨링하고 모델에 맞게 가공하는 과정이 필요하며, 이는 많은 비용을 수반한다. 따라서 데이터 라벨링 작업을 최소화하면서 본 모델을 활용할 수 있는 방법에 대한 연구가 필요할 것으로 보인다.

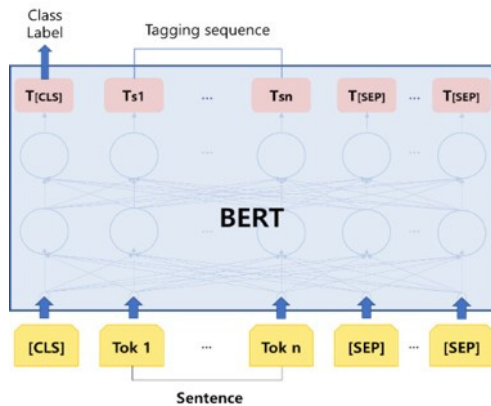
BERT는 트랜스포머(Transformer, Ashish Vaswani et al. 2017)를 기반으로 하며, 트랜스포머의 인코더(Encoder)-디코더(Decoder) 중 인코더만을 사용한다. 그리고 BERT는 언어의 특성을 잘 학습할 수 있도록 MLM(Masked Language Model)과 심층 양방향 트랜스포머 인코더를 사용하여 가려진(Masking) 단어를 예측하고, NSP(Next Sentence Prediction) 방식을 사용하여 문장 사이의 관계를 학습한다. MLM은 입력 문장에서 임의로 토큰을 가리고 그 토큰을 맞추는 방식이며, NSP는 두 문장이 주어졌을 때 두 번째 문장이 첫 번째 문장의 바로 다음에 오는 문장인지 여부를 예측하는 방식이다.

BERT를 이용한 모델은 크게 2단계로 나뉘서 생각해볼 수 있다. 거대 인코더가 입력 문장들을 임베딩(Embedding) 하여 언어를 모델링하는 사전 학습과, 이를 파인 튜닝(Fine-tuning)하여 특정 자연어 처리 과업을 수행하는 과정으로 나눌 수 있다. 이 중 사전 학습은 초기 비용이 많이 들지만, 각 언어별로 한 번만 실행하면 된다는 특징이 있다. 따라서 사전 학습된 BERT를 활용하면 자연어 처리 연구자들이 각자 사전 학습하는 비용을 절감할 수 있다.

파인 튜닝은 기존의 사전 학습된 BERT 모델에 최종 출력 레이어를 추가하여 모델을 조정하

는 방식으로 많은 자연어 처리 과업들을 간단하게 처리할 수 있다. 특정 과업에만 쓰이는 (Task-specific) 파라미터를 최대한 줄이고, 사전 학습된 파라미터들을 특정 과업에 맞게 조정하는 방식이다. 기본적으로 사전 학습 단계에서 학습된 높은 언어 이해도를 바탕으로 파인 튜닝, 즉 특정 과업에 대한 학습을 진행함으로써 자연어 처리 작업의 성능 향상을 기대할 수 있다.

BERT는 총 3.3억 단어를 바탕으로 MLM, NSP 방식을 적용하여 사전 학습한다. BERT는 모델 크기에 따라 BERT-base(L=12, H=768, A=12), BERT-large(L=24, H=1024, A=16), 2가지로 분류할 수 있다. 현재 BERT가 많이 사용되는 분야는 언어 모델링(Banerjee et al. 2021)이며, 그 외에 Q&A(Question and Answering)(Chaybouti et al. 2021), 텍스트 분류(Khezrian et al. 2020), 감성 분석(Azhar et al. 2021) 등의 분야에서도 활용되고 있다.



〈Figure 1〉 BERT architecture for Text Classification(Devlin et al. 2018)

본 연구에서는 uncased BERT-Base 모델을 사용하였으며, 문장의 단어 토큰과 차원, 감성 토큰

들을 입력한 후 BERT 인코딩(Encoding), 파인 튜닝 단계를 거쳐 클래스 레이블(Class Label)과 태깅 시퀀스(Tagging sequence)를 출력값으로 얻는다.

2.2 CNN과 텍스트 분류

2.2.1 CNN

• 합성곱 계층(Convolution Layer)

CNN은 필터링 기법을 인공지능경망에 적용함으로써 이미지를 더욱 효과적으로 처리하기 위해 개발되었다(LeCun et al. 1989). 기존의 신경망과는 달리 입력 원본 데이터를 바로 연산하여 출력하지 않고, 특징(Feature) 추출 단계와 분류 단계를 거쳐 출력값을 산출하는 것이 특징이다(문현철 외 2018). 이러한 CNN은 합성곱 계층, 풀링(Pooling) 계층, 완전 연결(Fully Connected, 이하 FC) 계층으로 구성되어 있다. 합성곱이란 데이터의 특징을 추출하는 과정으로, 데이터의 각 성분에 인접한 성분들을 조사한 후 특징을 도출하는 과정이다.

• 풀링 계층(Pooling Layer)

CNN에서 이미지는 합성곱 계층과 ReLU와 같은 비선형 활성화 함수를 거쳐 생성된다. 이때 국소 영역 내부에서는 픽셀들이 이동 및 회전 등에 의해 위치가 변경되더라도 출력값은 같다. 따라서, 풀링 계층을 이용할 경우 이미지를 구성하는 요소들의 이동 및 회전 등에 의해 CNN의 출력값이 영향을 받는 문제를 완화할 수 있다. 또한, 고차원일수록 파라미터들이 증가해 과적합(Overfitting)을 초래할 수 있는 상황에서 풀링 계층은 차원을 줄여 이러한 문제를 해결할 수

있다.

풀링의 대표적인 방법으로는 최대 풀링(Max Pooling)과 평균 풀링(Average Pooling)이 있다. 하지만 최대 풀링(Max Pooling)의 경우, 커널 크기(Kernel Size) 등의 하이퍼파라미터(Hyperparameter)를 수동으로 설정해야 하며, 만약 입력 크기를 변경한다면 이에 맞춰 다시 설정해야 한다. 반면에 적응형 평균 풀링(Adaptive Average Pooling)은 입력 및 출력 차원이 주어졌을 경우, 이에 대한 출력을 생성하는 과정에서 필요한 정확한 커널 크기를 스스로 계산하는 평균 풀링 연산으로 정의할 수 있다(Wyk and Bosman, 2019). 이와 같은 특성으로 적응형 평균 풀링은 연속적인 합성곱 블록의 차원 불일치로 인한 위상학적 제약(Topological Constraints)을 제거하기 위해 사용된다.

2.2.2 텍스트 분류를 위한 CNN 활용

이미지 처리를 위한 CNN의 필터가 이미지의 지역적인 정보를 추출하는 역할을 한다면, 텍스트 CNN의 필터는 텍스트의 지역적인 정보, 즉 단어 등장 순서 및 문맥 정보를 추출하는 역할을 한다(Raj, Sahu and Anand, 2017; She and Zhang, 2018). CNN을 사용하는 텍스트 분류는, 텍스트를 워드 임베딩(Word Embedding)을 사용해 벡터로 변환한 후 합성곱 계층과 FC층을 활용한 모델을 통해 이루어진다. 분류하고자 하는 텍스트에 대해, 만일 단어가 n 차원으로 임베딩되며, 한 문장이 최대 m 개의 단어를 가진다면 해당 문장을 $m \times n$ 의 행렬로 표현할 수 있다. 위와 같은 방식으로 생성된 행렬에 합성곱 계층을 적용하면 텍스트를 분류할 수 있다.

간단한 구조의 CNN을 활용하여 영화 리뷰 감

성 분석을 수행한 연구(Yoon Kim, 2014)는 문장 분류에서 뛰어난 성능을 보이며 당시에 많은 주목을 받았다. 해당 모델의 입력값은 k -차원의 단어 벡터이고, 문장은 최대 n 개의 단어로 구성된다. 이때 문장의 단어 수가 정해진 최대 단어 수보다 적다면 패딩을 추가한다. 그 후, 필터를 거쳐 만들어진 특징 맵(Feature Map)의 최대값을 추출하고 그 값은 특정 필터를 설명하는 특징이 된다. 이러한 특징들은 완전 연결 은닉 계층(Fully-connected Hidden Layer)과 소프트맥스 함수(Softmax)를 거쳐 라벨에 대한 확률 분포가 된다. 해당 모델은 CNN을 활용해 만든 간단한 모델임에도 불구하고 기존 다른 모델들과 비교해서 뛰어난 성능을 보여준다는 점에서 주목할 가치가 있다.

2.3 차원 기반 감성 분석에 대한 동향

2.3.1 감성 분석과 차원 기반 감성 분석

감성 분석이란 단어와 문맥을 분석하여 텍스트의 감정을 파악하는 기술이다(이정훈, 2018). 이는 “긍정”과 “부정”같이 상반되는 극성(Polarity)의 분류부터 여러 종류의 감성에 대한 분류까지 텍스트에 내포된 의견이나 감성 등을 범주화하는 작업이다(서상현, 김준태, 2016). 이러한 감성 분석은 지도 학습 또는 비지도 학습으로 이루어진다. 이 중 지도 학습을 활용한 감성 분석 관련 연구의 경우 지도 학습 알고리즘을 통해 감성 분류기를 학습시킨다. 비지도 학습의 경우, 긍정 또는 부정을 의미하는 단어를 기존에 구축된 사전을 통해 판별하는 방식이다. 하지만 기존의 감성 분석은 문장 자체의 감성 분석에 초점을 맞췄기 때문에 문장 내의 감성과 차원이 다양하게 보일 경우에는 한계가 있다. 따라서 본

연구에서는 차원 기반 감성 분석에 집중하고자 한다.

차원 기반 감성 분석은 텍스트에서 타깃, 차원 및 감성을 탐지하는 것을 목표로 한다(Hai Wan et al. 2020). 최근 ABSA는 특히 소셜 미디어와 온라인 여론의 성장으로 점점 더 많은 관심을 받고 있다. 예를 들어, “스테이크가 맛있다.”라는 문장에서 ‘스테이크’라는 타깃을 통해서 ‘음식’이라는 차원을 추출하고 ‘맛있다’라는 표현을 통해 긍정적인 감성으로 분류한다. ABSA에 대한 대부분의 기존 연구는 차원, 감성, 타깃의 조합으로 ASD(Aspect Sentiment Detection), TSD(Target Sentiment Detection), TAD(Target Aspect Detection), AD(Aspect Detection) 및 TD(Target Detection)와 같은 TASD의 다섯 가지 하위 작업에 중점을 두고 타깃과 차원을 각각 이용해 감성을 예측했다. 차원과 감성을 동시에 분석하는 ASD는 차원과 감성을 묶어서 배정된 차원-감성이 옳은지를 판단하는 이진 분류 문제로 축소하고 사전 학습된 언어 모델을 파인 튜닝하여 이를 해결한다(Sun, Huang, and Qiu, 2019). TSD는 타깃과 감성을 공동으로 감지하는 것을 목표로 하며, 두 개의 스택형 LSTM 네트워크로 구성된 통합 모델(Li et al. 2019) 등으로 문제를 해결한다. 타깃과 차원을 같이 분석하는 TAD에 대해서는 이루어진 작업은 거의 없으며, 차원만 분석하는 AD는 SVM을 통한 분석(Kiritchenko et al. 2014; Xenos et al. 2016)부터 어텐션(Attention) 기법을 사용한 연구(He et al. 2017; Xue et al. 2017; Ma, Peng and Cambria, 2018; Movahedi et al. 2019)까지 매우 다양하다. 타깃만을 분석하는 TD 과제는 CRF(Conditional Random Field), CNN, RNN(Recurrent Neural Network)과 같은 신경망 모델이 활용되고

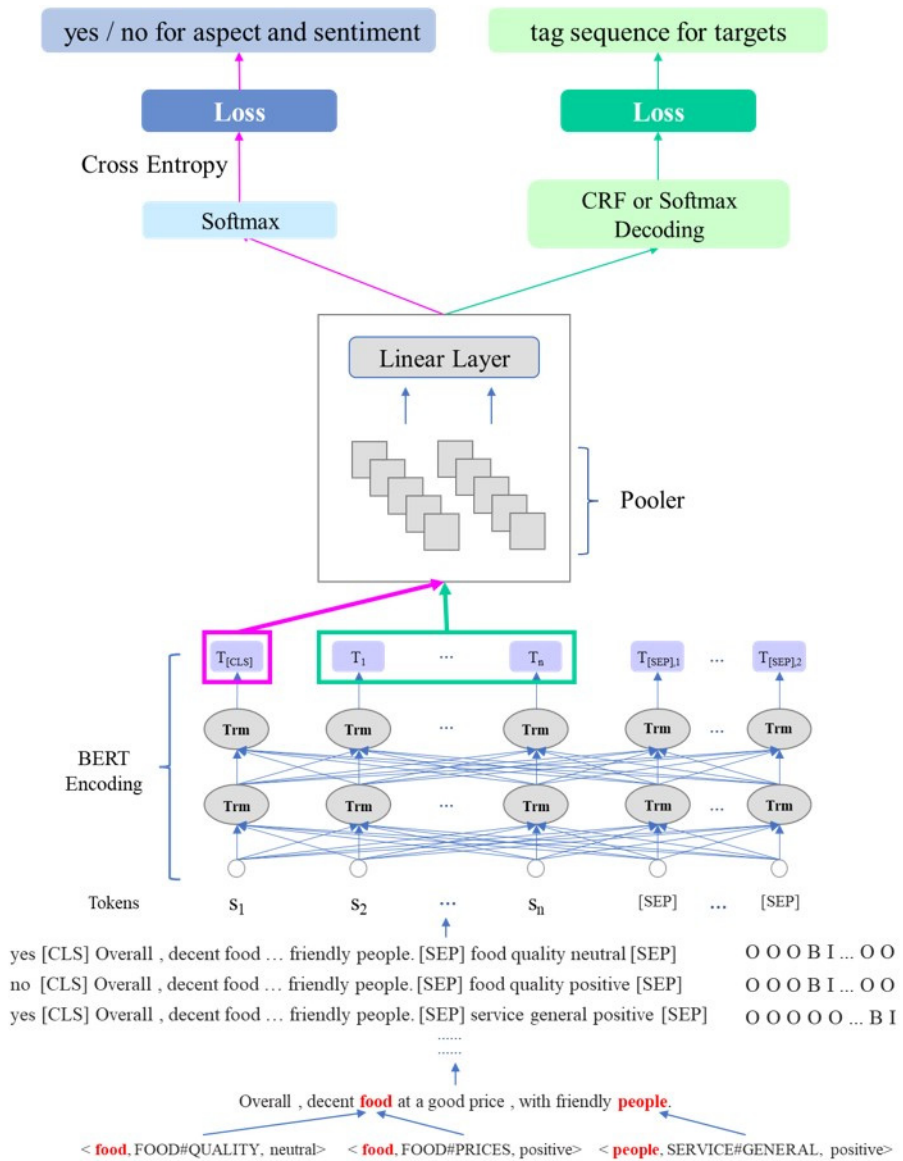
있다.

하지만 선행 연구들이 해결할 수 없는 문제가 있다. 동일한 문장 내에서 특정 차원 혹은 타깃의 감성이 혼재된 경우, 이중 의존성 문제가 발생하여 해당 차원 또는 타깃에 대한 감성 분석이 정확하지 않을 수 있다. 특히 암묵적인 타깃이 존재할 경우, TSD를 사용한 감성 예측 방법은 성능에 한계를 가진다. 이를 해결하기 위해 타깃과 차원, 감성 모두를 분석하는 TASD 연구가 존재한다. 하지만 TASD에 대한 연구는 아직 부족한 상태이다. TASD 연구 중 하나인 Brun과 Nikoulina의 연구(Brun and Nikoulina, 2018)에서 사용 가능한 구문 분석기(Parser)와 도메인별 의미 어휘에 의존하는 방법을 제안되었지만, 이 방법은 F1-score가 최대 38.1로 성능이 낮아 이를 보완할 필요가 있다.

2.3.2 차원 기반 감성 분석을 위한 타깃-차원-감성 동시 탐지

TAS-BERT(Hai Wan et al. 2020)는 TASD 연구로 이중 의존성 문제와 암묵적인 타깃을 감지하지 못하는 문제를 해결하였다. 또한, 감성에 대한 타깃-차원 쌍을 지니는 SemEval-2015/SemEval-2016 데이터에서 더 좋은 성능을 보였다. TAS-BERT에서는 동시 탐지(Joint Detection) 문제를 2단계로 축소하여 해결하였다. 첫째, 문장에 가능한 모든 차원-감성 쌍을 붙여 차원-감성 감지 문제를 이진 분류 문제로 처리하였다. 둘째, 타깃의 경우 문장에 나타난 단어 토큰을 바탕으로 시퀀스 레이블링(Sequence Labeling) 문제로 처리하였다.

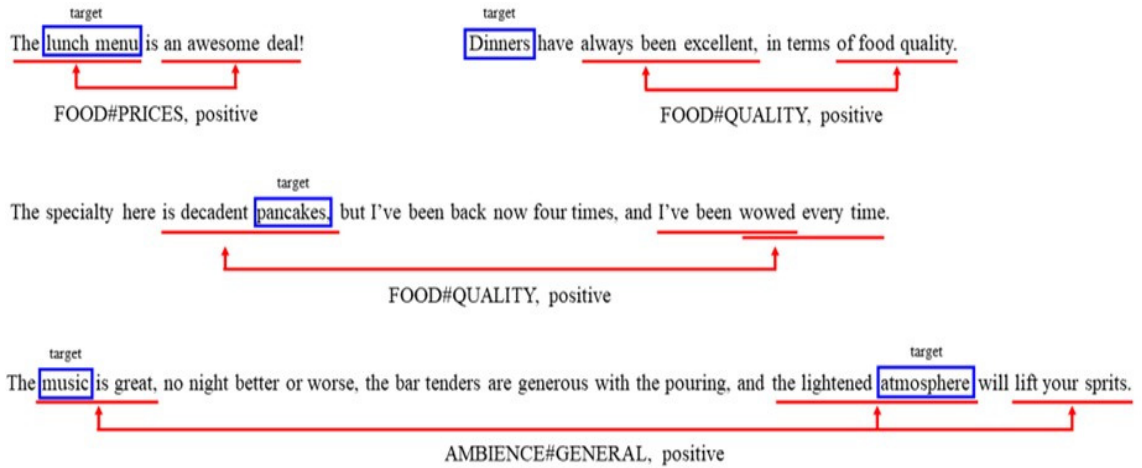
TAS-BERT 모델에서 토큰화된 문장은 BERT 인코더를 통해 임베딩되고, 임베딩한 값은 2개의 선형 FC 계층을 거친다. 그 후 차원-감성에 대한



<Figure 2> The architecture and a running example for the TAS-BERT model (Hai Wan et al, 2020)

이진 분류는 소프트맥스 디코더를 통해 결정되며, 시퀀스 레이블링 문제는 CRF 디코더 또는 소프트맥스 디코더를 통해 결정된다. 두 가지 디

코더는 BIO 태깅(Tagging), TO 태깅 두 가지와 조합된다. 특히 이 모델에서의 CRF 디코더는 오직 상태 전이 제약조건(Transition Constraints)이



(Figure 3) Example of Sentences that TAS-BERT failed to predict

있는 태깅에서 잘 작동하므로 CRF와 BIO 태깅 조합과 소프트 맥스 디코더와 TO 태깅 조합만을 고려한다. 또한, 학습 시 존재하지 않았던 단어 (Out of the vocabulary) 문제를 해결하기 위해 가장 긴 접두사 매칭(Longest Prefix Match), 단어 분할(Split Word) 등을 사용하였다.

TAS-BERT는 TASD, 그리고 ABSA 하위 연구들에서 모두 좋은 성능을 보였다. 특히 암묵적 타깃 문제에서 3배 이상의 성능을 보였고, 미지의 언어를 인식 가능한 어휘로 나누는 단어 분할 방식에서 전반적인 성능 향상이 있었다. 구체적인 예시로, SemEval Restaurant-2015 데이터셋에서는 57.51의 F1 점수를 달성하였고, Restaurant-2016 데이터셋에서는 65.89의 F1 점수를 달성하였다. 본 연구에서는 TAS-BERT에서 성능이 가장 높았던 단어 분할, BIO 태깅과 CRF를 사용한 TAS-BERT-SW-BIO-CRF 방식을 비교군으로 삼아 연구를 진행하였다.

하지만 TAS-BERT에는 지역적 문맥을 잘 포착하지 못한다는 한계가 존재한다. 아래 그림은

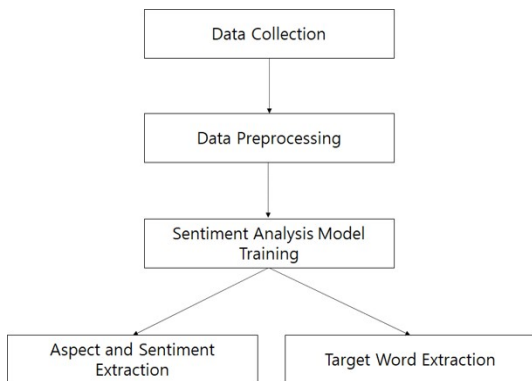
기존 모델이 잘 설명하지 못하는 문장 중 일부이다. 이를 살펴보면, 차원-감성을 예측하는 과정에서 필요한 핵심 단어들(구를 이뤄 문장에 분산되어 나타난다. 특히, 마지막 문장의 경우 분위기에 관련한 타깃과 표현들이 “music is great”, “the lightened atmosphere”, “lift your sprits” 등으로 문장 내에 구를 이뤄 분산되어 나타나 있다. 즉, 기존 모델의 경우 단어 간 상관관계를 분석하여 문맥을 포착하는 것은 뛰어나나, 구 간의 관계를 제대로 포착하지 못하는 한계가 있다.

본 연구에선 이와 같은 한계를 CNN의 필터를 사용해 해결하고자 하였다. 필터를 통해 일정한 크기의 구의 특성을 추출할 수 있고, 풀링을 통해 적절한 크기의 예측 값으로 변환해 차원-감성을 예측한다. 이를 바탕으로 얻은 손실을 보조 손실로 사용하여 학습 시엔 지역적 문맥을 더 잘 반영하도록 하였으며, 학습 후엔 기존의 방법을 사용하여 차원-감성을 예측하였다.

3. 제안 모델

3.1 연구 설정 배경

본 연구는 차원 기반 감성 분석의 발전을 위해 앞서 언급된 TAS-BERT의 성능을 높이는 것을 목표로 한다. TAS-BERT는 모델 자체로 TASD 분야에서 높은 성능을 보이지만 몇 가지 문제점을 발견하였다. TAS-BERT가 BERT, 즉 트랜스포머, 어텐션 기반이기 때문에 단어 하나하나의 관계를 포착하는 것은 뛰어나지만 단어가 묶여 구를 이뤄야 정확한 의미를 가질 때, 구 간 관계를 포착하는 것은 부족하다. 또한, 예폭 30으로 모델을 실행시킬 때는 안정적인 F1 점수를 보이지만 예폭 수가 적어질수록 F1 점수가 점점 작아지는 결과를 보인다. 이번 연구에서는 TAS-BERT의 이러한 부분을 해결하기 위해 CNN을 원 모델에 적용하여 지역적 문맥을 보다 잘 포착하고자 하였다. 기존 모델에서는 BERT에 단순히 선형 계층(Linear Layer)을 더해 차원-감성을 예측했다면, 본 모델에서는 BERT에 선형 계층과 CNN을 병렬로 결합하여 차원-감성을 예측했다.



〈Figure 4〉 The overall process of the proposed model

본 모델이 제시하는 전체적인 프로세스는 아래 그림과 같다. 우선 학습 데이터를 수집하고, 전처리를 통해 모델이 학습할 수 있는 형태로 변환한다. 이후 모델에 데이터를 넣어 학습하고, 학습된 모델을 바탕으로 차원, 감성, 타깃 단어를 추출한다. 모델 학습에 관한 자세한 내용은 3.2.3에서 설명하였다.

3.2 사전 준비 단계

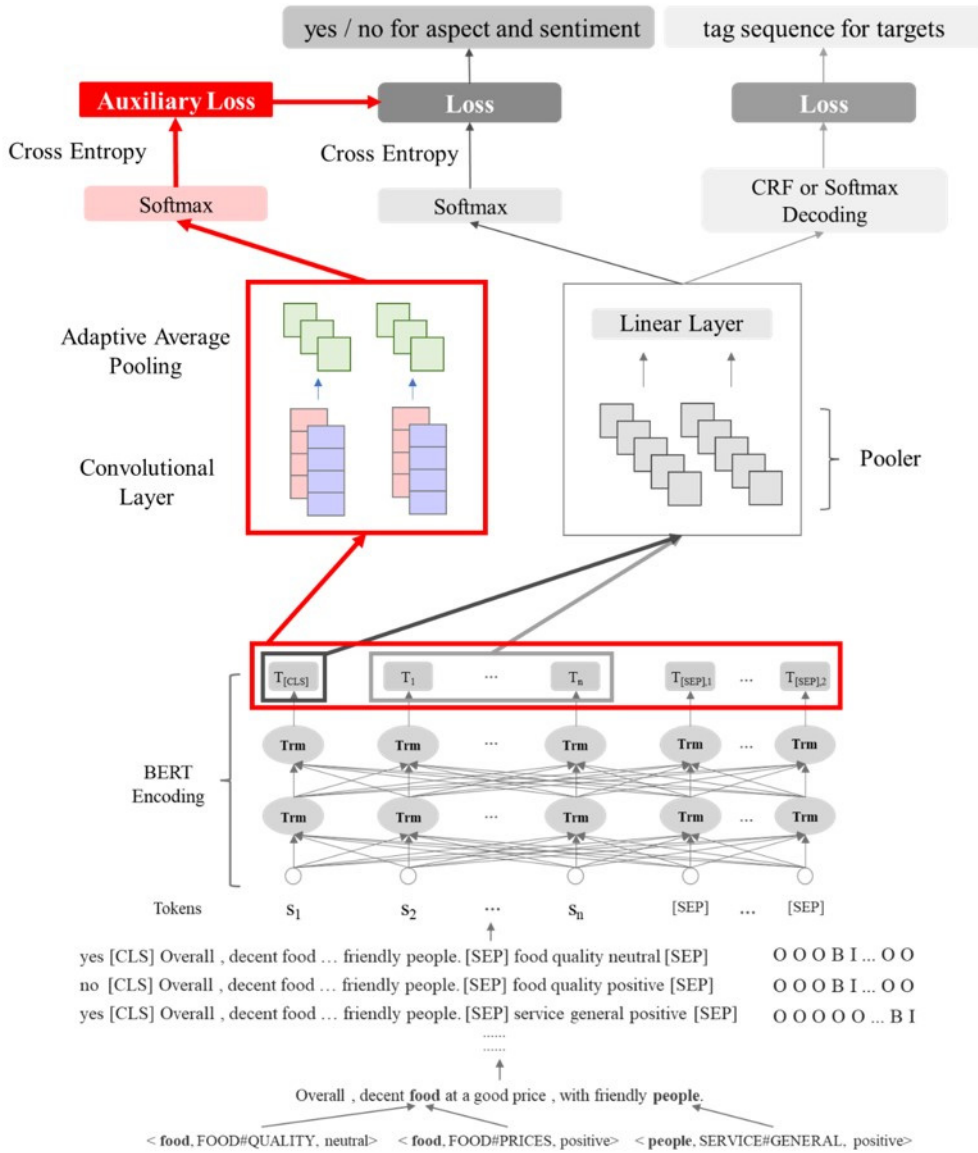
3.2.1 학습 데이터 수집

모델 학습을 위한 데이터는 주어진 자료를 사용하거나 웹 크롤링(Web Crawling)을 통해 텍스트를 수집한 후, 라벨링 과정을 거쳐 모델에 맞는 형식으로 수정한다. 본 연구에서는 SemEval 2015의 Restaurant-2015(Ponitiki et al. 2015)와 SemEval 2016의 Restaurant-2016(Pontiki et al. 2016)를 사용하였다. 자세한 데이터 라벨링 형식 및 과정은 ‘4. 실험 설계 및 결과’에서 설명하도록 한다.

3.2.2 학습 데이터 전처리

• CRF디코더와 BIO 태깅

시퀀스 레이블링의 문제를 해결하기 위해 본 연구에서는 CRF 디코더를 사용한다. CRF는 입력 연속값에 대한 출력 연속값의 조건부 확률이다. 하나의 연속값만 보고 분석하는 것이 아니라 가능성이 있는 시퀀스 후보를 몇 개 선택한 후, 가장 적합한 하나의 라벨 시퀀스를 고르는 방법이다. 그리고 또한, 2.3.2.에서 언급했듯이, CRF는 과도 제약이 있는 태깅에서만 잘 작동하기 때문에 BIO 태깅과 조합하여 사용되었다. 개체명을 인식하는 BIO 태깅의 B는 Begin의 약자로 개



〈Figure 5〉 The architecture and a running example for the proposed model 'BERT with CNN Auxiliary Loss'

체명이 시작되는 부분, I는 Inside의 약자로 개체명의 내부 부분, O는 Outside의 약자로 개체명이 아닌 부분을 뜻한다.

• 단어 분할

사전 학습된 BERT를 학습시킬 때, 학습 시 존재하지 않았던 단어(Out of the vocabulary) 문제

를 해결하기 위해 단어 분할 방법을 사용한다. 이때 단어 분할 방법은 알 수 없는 단어를 사전을 기반으로 여러 개의 알아볼 수 있는 단어로 분해하는 방법이다.

3.2.3 감성 분석 모델 학습

본 연구는 기존의 손실 함수에 보조 손실을 추가하는 방식으로 기존 모델의 성능을 향상시키고자 하였다. 본 연구에서 제안하는 모델은 <Figure. 5>과 같다. 그림에서 붉은 박스와 화살표로 표시된 부분이 기존의 TAS-BERT 모델을 확장한 부분이다. 크게 BERT 인코딩, 선형 FC 계층, 합성곱 계층 및 적응형 평균 풀링, 손실 함수, 감성/차원 추출 및 타깃 단어 추출 단계로 이루어진다.

(1) BERT 인코딩

훈련 단계에서는 (S, a, p, f, T)라는 튜플(Tuple)이 주어진다. S는 n 개의 단어로 이루어진 문장이며 각 단어는 s_1, s_2, \dots, s_n 으로 표시된다. a는 차원, p는 감성, f는 'yes/no' 라벨 그리고 T는 레이블 시퀀스(Label sequence)이다. BERT 인코더의 입력값은 토큰 시퀀스(Token sequence)로 "[CLS], $s_1, \dots, s_n, [SEP], a_1, \dots, a_m, p, [SEP]$ " 구성되며, $n+m+4$ 개의 토큰 연속값을 사용한다. 이 토큰 연속값은 BERT 인코더에 입력되어, BERT의 마지막 층에서 d 차원 벡터의 연속값을 출력한다. d 차원 벡터들의 연속값은 $T_{[CLS]}, T_{s_1}, \dots, T_{s_n}, T_{[SEP,1]}, T_{a_1}, \dots, T_{a_m}, T_p, T_{[SEP,2]}$ 으로 이루어져 있다. 임베딩 시에는 사전 학습된 BERT-Base 모델을 활용하였다.

(2) 선형 FC 계층

BERT 인코딩의 출력 시퀀스 중 제일 첫 번째 토큰인 $T_{[CLS]}$ 는 풀러(Pooler)와 FC 층을 거치면서 $T_{[CLS]}$ 토큰에 대한 은닉 상태(Hidden state)를 계산하게 된다. 이 은닉 상태를 통해 소프트맥스를 통해 차원과 감성의 'yes/no' 라벨링을 학습하게 된다.

(3) 타깃 추출

T_{s_1}, \dots, T_{s_n} 의 토큰들은 라벨 시퀀스를 예측하는데 사용된다. (2)와 다른 FC 층을 거치고 CRF 디코더를 사용하여 학습하게 된다.

(4) 합성곱 계층 및 적응형 평균 풀링

(2), (3)과는 달리 이 단계에서는 BERT 인코딩 단계에서 출력된 모든 토큰들을 사용한다. $T_{[CLS]}, T_{s_1}, \dots, T_{s_n}, T_{[SEP,1]}, T_{a_1}, \dots, T_{a_m}, T_p, T_{[SEP,2]}$ 까지의 모든 토큰들을 합성곱 계층에 입력하고, 적응형 평균 풀링을 거쳐 특징 맵을 출력하게 된다.

(5) 손실 함수

제안 모델에서 손실 함수는 크게 두 부분으로 구성된다. 첫 번째는 차원-감성 이진 분류를 통해 나온 손실이고, 두 번째는 CRF방식으로 타깃을 찾을 때 계산되는 손실이다. 우선 차원-감성 이진 분류에 대한 손실 함수를 살펴보면, 2가지 방법을 통해 차원-감성 이진 분류를 하게 된다.

첫 번째는 BERT 인코딩을 통한 나온 은닉 상태의 첫 번째 토큰을 바탕으로 선형 계층을 거쳐 차원-감성 분석을 하는 방법이다. 위의 BERT 인

코딩을 통해 나온 은닉 상태의 첫 번째 토큰, $T_{[CLS]}$ 를 가지고 선형 계층, 하이퍼볼릭탄젠트 (\tanh), 소프트맥스 계층을 통과시켜 차원-감성 분석을 위한 확률 분포 벡터 g_1 을 도출할 수 있다.

$$\begin{aligned} P_{[CLS]} &= \tanh(W_1 T_{[CLS]} + b_1) \\ g_1 &= \text{softmax}(P_{[CLS]}) \\ \text{이때, } w_1 &\in \mathbb{R}^{d \times 2}, b_1 \in \mathbb{R}^2 \end{aligned}$$

두 번째 방법은 은닉 상태의 모든 토큰을 바탕으로 CNN 계층을 거쳐 차원-감성 분석을 하는 방법이다. BERT 인코딩을 통해 나온 은닉 상태의 각각의 토큰을 합쳐 $T_{[ALL]} = \{T_{[CLS]}, T_{S_1}, \dots, T_{S_n}, T_{[SEP,1]}, T_{a_1}, \dots, T_{a_m}, T_p, T_{[SEP,2]}\}$, 즉 $(n + m + 4) \times d$ 크기의 행렬로 만든다. 이를 CNN 계층 및 적응형 평균 풀링으로 구성된 합성곱 함수에 넣으면 특징 맵 c 벡터를 얻을 수 있다. 이를 소프트맥스에 대입하면 다음과 같은 차원 감성 분석을 위한 확률 분포 벡터 g_2 를 계산할 수 있다.

$$\begin{aligned} c &= \text{Conv}(T_{[CLS]}) \\ g_2 &= \text{softmax}(c) \end{aligned}$$

$$\text{이때, } T_{[ALL]} \in \mathbb{R}^{(n+m+4) \times d}, c \in \mathbb{R}^2$$

두 가지 확률 분포 벡터를 바탕으로 손실 함수로 크로스 엔트로피(Cross entropy)를 채택해 손실을 계산하면 다음과 같다.

$$\begin{aligned} \text{loss}_g &= - \sum_{i=1}^2 I(y_n(i) = f) \log(g_1^i) - \sum_{i=1}^2 I(y_n(i) = f) \log(g_2^i) \\ g_i &= i\text{번째 } g, y_n(1) = \text{yes}, y_n(2) = \\ \text{no, } X &\text{가 true면 } I(X) = 1, \text{ 아니면 } I(X) = 0 \end{aligned}$$

타깃 단어를 위한 손실 함수를 보면 다음과 같다. 은닉 상태의 단어 토큰을 선형 함수를 거쳐 BIO 태깅을 위한 형태로 바꾼다. 이를 사용하여 CRF 디코더로 레이블 시퀀스를 예측하기 위한 확률 행렬을 계산한다. 타깃 단어 추출에 대한 손실 함수는 아래와 같은 형태로 주어진다.

$$\begin{aligned} P_{S_i} &= \tanh(W_1 T_{S_i} + b_2) \\ (w_2 &\in \mathbb{R}^{d \times 3}, b_2 \in \mathbb{R}^2) \\ P(T|P) \end{aligned}$$

$$\text{이 때, } P = [P_{S_1}, \dots, P_{S_n}] \quad P \in \mathbb{R}^{n \times 3}$$

$$\text{loss}_h = -\log(P(T|P))$$

최종적으로 차원-감성에 대한 이진 분류를 통해 얻은 손실과 타깃 단어 추출에서 얻은 손실을 더해 본 모델의 손실 함수가 계산되고, 이를 바탕으로 손실을 최소화하는 방향으로 모델 학습이 진행된다.

$$\text{loss} = \sum_{i=1}^N (\text{loss}_g^i + \text{loss}_h^i)$$

(6) 타깃, 차원, 감성 추출

위의 손실 함수를 바탕으로 학습이 진행된 후에는 P 확률 행렬, g_1 확률 벡터를 통해 태깅 시

〈Table 1〉 SemEval Dataset

id = 1014458:3		
The wine list is interesting and has many good values.		
target: wine list	aspect: DRINKS#STYLE_OPTIONS	sentiment: positive
target: wine list	aspect: DRINKS#PRICES	sentiment: positive
id = 1014458:4		
For the price, you cannot eat this well in Manhattan.		
target: NULL	aspect: FOOD#QUALITY	sentiment: positive
target: NULL	aspect: RESTAURANT#PRICES	sentiment: positive
id = 1028246:1		
Service was devine, oysters where a sensual as they come, and the price can't be beat!!!		
target: service	aspect: SERVICE#GENERAL	sentiment: positive
target: oysters	aspect: FOOD#QUALITY	sentiment: positive

퀀스, 배정된 차원-감성에 대한 이진 분류 예측값을 산출한다. 태깅 시퀀스를 통해 타깃을, 차원-감성에 대한 이진 분류 예측값을 통해 차원, 감성을 예측한다.

4. 실험 설계 및 결과

4.1 실험 내용

본 연구는 자연어 처리 분야에서 세계적으로 사용되고 있는 벤치마크 데이터셋인 SemEval 2015의 Restaurant-2015(Pontiki et al. 2015)와 SemEval 2016의 Restaurant-2016(Pontiki et al. 2016)을 사용하였다. 기존의 데이터에서는 감성을 ‘긍정’, ‘부정’, ‘중립’, ‘대립’ 4가지로 분류하였으나, 상반된 감성을 표현하는 ‘대립’의 경우

그 수가 많지 않고 학습에 방해가 될 수 있기 때문에 실험에서 제외하였다. 또한, Restaurant-2015 데이터셋의 경우 사전에 정해진 차원은 13가지이나, 데이터셋에서는 식당#일반, 음식#일반, 음료#가격 등 12가지의 차원만을 가지고 있다. 반면에 Restaurant-2016 데이터셋의 경우 미리 정해진 차원과 데이터에 나타난 차원 모두 12가지이다. 데이터셋은 <Table 1> 같은 형태로 제공된다. 더불어 사용되는 데이터의 차원 및 감성 레이블에 대한 통계 정보는 <Table 2>에 나타나 있다. 각 문장에 해당하는 차원 및 감성이 여러 개일 수 있으므로 주어진 문장에서 하나의 차원만 나타날 경우에는 하나의 데이터로, 둘 이상의 차원 및 감성이 나타날 경우 차원 및 감성의 개수대로 별도로 집계하였다.

〈Table 2〉 Statistic about the number of aspect, sentiment labels per sentence of dataset

Dataset		Sentences	Aspect		Sentiment		
			single	multi	positive	neutral	negative
Res15	Train	1117	825	292	1083	52	369
	Test	580	417	163	413	45	329
Res16	Train	1702	1244	458	1501	100	679
	test	585	450	135	513	42	195

〈Table 3〉 Data pre-processing based on aspect-sentiment pairs

Sentence_ID	Yes_No	Aspect_Sentiment	Sentence(Review)	Ner_Tags
1104293:0	1	Restaurant general negative	Judging from previous posts this used to be a good place, but not any longer	O O O O O O O O O O B O O O O O O
1104293:0	0	Restaurant general positive	Judging from previous posts this used to be a good place, but not any longer	O O O O O O O O O O B O O O O O O
1104293:0	0	Restaurant general neutral	Judging from previous posts this used to be a good place, but not any longer	O O O O O O O O O O B O O O O O O
...
1104293:0	0		Judging from previous posts this used to be a good place, but not any longer	O O O O O O O O O O B

주어진 데이터를 바탕으로 문장의 각 감성, 차원에 대해 이진 분류할 수 있도록 데이터를 가공하였다. 이때 각 문장별로 차원과 감성으로 짝을 지었을 때 나타날 수 있는 모든 경우에 ‘yes_no’라는 이진 분류 결과를 나타내는 열을 추가하여 39 또는 36개의 데이터로 전처리하였다. 예를 들어, 문장의 차원이 ‘식당#일반’, 해당 차원 감성이 ‘부정’인 경우, 차원과 감성으로 짝을 지어 나타날 수 있는 모든 경우인 39개의 데이터에 대해 Restaurant general negative(식당 일반 부정)일 때에만 ‘yes_no’에 1이라는 값을 매기고 이외 다른 데이터의 ‘yes_no’에는 0이라는 값을 매겼다.

예시에 대한 내용은 <Table 3>에 자세히 나타나 있다.

이와 같은 전처리를 통해 데이터를 구축한 후, 구글이 사전 학습한 BERT-base를 활용하여 임베딩을 한 후 과인 튜닝을 하는 방식으로 모델 학습을 진행하였다. 기본 모델은 배치 8, 최대 문장 길이(Max Sequence) 40, 학습률(Learning Rate) 2e-5, 에폭 5으로 설정하였다. 동일한 하이퍼파라미터의 TAS-BERT 모델(Figure. 2)을 비교군으로 삼아 에폭, 배치 등 하이퍼파라미터를 조정하면서 실험을 진행하였다.

4.2 성과 척도

성과 척도로는 F1 점수를 사용하였다. F1 점수는 재현율(Recall)과 정밀도(Precision)의 조화 평균이다. 재현율은 실제로 ‘Positive’인 데이터를 모델이 ‘Positive’라고 인식한 데이터의 수를 나타내는 성과 척도이다. 정밀도는 모델이 ‘Positive’라고 예측한 데이터 중 실제로 ‘Positive’인 데이터의 수를 나타내는 성과 척도이다. F1 점수는 재현율과 정밀도를 단순 평균이 아닌 조화 평균으로 계산하여 재현율과 정밀도가 극단적으로 차이 나는 경우에 높은 패널티를 부과하는 특성이 있다. 재현율, 정밀도, F1 점수에 관한 식은 다음과 같다. 실제 ‘Positive’인 데이터를 모델이 ‘Positive’로 인식했을 경우 TP(True Positive), ‘Negative’로 인식했을 경우 FN(False Negative)로 표현하였다. 또한, 실제 ‘Negative’인 데이터를 모델이 ‘Positive’로 인식했을 경우 FP(False Positive), ‘Negative’로 인식했을 경우 TN(True Negative)로 표현하였다.

$$(\text{Precision}) = \frac{TP}{TP + FP}$$

$$(\text{Recall}) = \frac{TP}{TP + FN}$$

$$(\text{F1 - score}) = 2 \times \frac{1}{\frac{1}{\text{Precision}} + \frac{1}{\text{Recall}}} =$$

$$2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

본 논문에서는 크게 2가지 이유로 F1 점수를 성과 척도로 채택했다. 본 연구에 사용된 데이터

의 경우 ‘Positive’인 데이터에 비해 ‘Negative’인 데이터의 비율이 매우 높기 때문에 성과 척도를 정확도(Accuracy)로 정할 경우 모델이 모든 데이터를 ‘Negative’라고 판단했을 때도 정확도가 높아 모델의 성능이 왜곡될 우려가 존재한다. 또한, 재현율과 정밀도는 서로 트레이드오프(Trade-off) 관계를 가지는 특성이 있어, 두 지표만 사용할 경우 정확한 성능 평가가 어렵다. 이와 같이 본 연구에서는 불균형한 데이터 특성과 지표의 특성을 반영하여 F1 점수를 모델의 성과 척도로 채택하여 실험을 진행하였다.

4.3 실험 결과

실험 결과, 본 모델은 기존에 비교군으로 삼은 TAS-BERT 모델에 비해 비교적 높은 성능을 보였다. 이때 정확한 비교를 위해 동일한 환경에서 실험을 진행하였다. 학습률, 최대 문장 길이 등 동일한 하이퍼파라미터를 적용하였을 때, 본 연구의 모델이 TAS-BERT 모델보다 같거나 높은 F1 점수를 기록하였다. 이 경우 하이퍼파라미터 중 커널 사이즈, 손실 비율, 배치, 에폭 등을 조정하며 모델의 성능을 검증하였다. 토큰 수를 기반으로 하는 최대 문장 길이의 경우, SemEval 2015 데이터 기준 전체 1,117개 문장 중 17개만이 문장 길이 40을 초과하였다. 이와 같은 데이터 특성을 고려해보았을 때, 최대 문장 길이는 40이라고 판단되어 추가로 조정하지 않았다. 에폭의 경우, 40으로 설정할 때, 에폭 20 초반에서 30 중반 사이에서 최고 성능을 보여 최대 에폭 40까지만 비교 실험을 진행하였다. 성능을 살펴보았을 때 F1 점수 기준 최대 60까지의 성능을 확인할 수 있었다. 자세한 내용은 아래에 각각의 하이퍼파라미터별로 기술하였다.

〈Table 4〉 Model performance (F1-score) comparison based on batch size

Data set	Batch size	8	32	64	128
Restaurant - 2015	TAS-BERT	29	48	50	45
	TAS-BERT with CNN auxiliary loss(1D)	41	51	52	49
	TAS-BERT with CNN auxiliary loss(2D)	45	50	53	50
Restaurant - 2016	TAS-BERT	37	60	59	57
	TAS-BERT with CNN auxiliary loss(1D)	47	59	60	59
	TAS-BERT with CNN auxiliary loss(2D)	52	60	60	60

〈Table 5〉 Model performance(F1-score) comparison based on epoch

Data set	Number of epoch	5	10	20	30	40
Restaurant - 2015	TAS-BERT	29	34	27	45	44
	TAS-BERT with CNN auxiliary loss(1D)	41	42	46	46	48
	TAS-BERT with CNN auxiliary loss(2D)	45	42	38	48	47
Restaurant - 2016	TAS-BERT	37	47	52	54	54
	TAS-BERT with CNN auxiliary loss(1D)	47	54	54	54	57
	TAS-BERT with CNN auxiliary loss(2D)	52	56	55	56	56

배치 크기, 에폭 수에 따른 F1 점수의 변화는 다음과 같다. 배치 사이즈의 경우 사이즈가 8일 때 TAS-BERT와 본 연구의 모델의 차이가 가장 두드러졌다. 이후 배치가 커지면서 TAS-BERT와의 차이는 줄어들지만 여전히 비교적 높은 F1 점수를 보여주고 있다. 에폭의 경우 역시 5일 때, TAS-BERT와의 차이가 가장 두드러지고, 에폭이 증가함에 따라 TAS-BERT 모델과의 차이는 줄어들지만, 비교적 높은 F1 점수를 꾸준히 보여주

고 있다. 또한, 합성곱 계층의 경우 1D(Dimension)와 2D, 2가지로 구현하여 기존 모델과의 비교 실험을 진행하였다. 실험을 진행한 결과, 배치 크기가 작을 때에는 2D 모델이 우수하나 배치 크기가 커지면 두 모델이 성능에 큰 차이가 없었다. 또한, 에폭이 작은 경우 2D 모델이 우수하나 에폭이 커지면 성능의 우열이 경우에 따라 바뀌었다. 하지만 어떤 차원의 합성곱 계층이든 대부분 기존 모델보다는 같거나 높은 성능을 보이고

있다.

따라서 BERT를 사용하여 차원, 감성, 타깃을 분석하는 과정에서 CNN을 사용하여 추가적인 손실을 계산하는 것이 모델을 학습하는데 효과적임을 확인할 수 있었다. 특히, 배치 8, 에폭 5에서 비교 모델과 가장 큰 차이를 냈으며, 하이퍼파라미터를 조정했을 때도 비교 모델에 비해 높은 성능을 달성한다는 사실을 확인할 수 있었다. 특히, 작은 배치 크기와 적은 에폭으로 효과적으로 성능을 낼 수 있으므로, 자원이 한정되는 상황에서 보다 유용하게 사용할 수 있다

5. 결론

본 연구는 TASD를 기반으로 차원 기반 감성 분석의 정확도를 높이고자 하였다. 차원, 타깃을 모두 고려하지 않는 차원 기반 감성 분석은 이중 의존성 문제로 정확하게 분석하지 못한다는 한계를 가지고 있다. 이를 고려하여 차원 기반 감성 분석의 정확도를 높이기 위해, TASD분야의 연구를 진행하였다. 기존 TASD분야의 연구는 자원이 한정된 상황에서 텍스트의 지역적인 문맥을 잘 포착하지 못한다는 문제가 있었다. 이를 해결하고자 본 연구에서는 기존 연구에 지역적 문맥을 효과적으로 포착하는 합성곱 계층을 병렬로 결합하여 보조적인 손실로 사용하는 방안을 제시하였다. BERT 인코딩 후 기존 선형 계층과 더불어 합성곱 계층을 병렬로 구성하였고, 각각의 계층에서 계산된 손실을 더해 차원-감성 분석을 위한 손실로 사용하였다. 학습이 종료된 후에는 기존의 선형 계층만을 사용하여 차원-감성을 분석하도록 하였다. 또한, 태깅 시퀀스를 바탕으로 타깃에 대한 이진 분류 예측값을 산출하여

최종적으로 타깃, 차원, 감성을 분석했다. 제시된 방안의 유용성을 판단하기 위해 SemEval-2015 Task 12(Pontiki et al. 2015), SemEval-2016 Task 5(Pontiki et al. 2016)를 활용해 실험을 수행하였다. 실험 결과를 바탕으로 본 연구 모델이 기존 모델보다 한정된 자원에서 성능을 높이고자 한 연구 목적에 효과적임을 확인하였다.

본 연구는 기존 모델의 문제점을 해결해 차원 기반 감성 분석의 이중 의존성 문제를 해소할 수 있는 TASD 분야에서 효과적인 성능을 냈다는 학술적 의의가 있다. CNN을 활용해 구간의 관계, 즉 지역적 문맥을 추가로 반영하여 한정된 자원에서 기존 모델보다 높은 성능을 달성하였다. 또한, 실무적인 관점에서는 한정된 자원에서 효과적인 성능을 보여 컴퓨팅 자원이 충분하지 않은 경우에도 이를 활용할 수 있음을 보여주었다. 또한, 본 모델은 전이 학습을 바탕으로 한 모델로, 데이터가 주어진다면 다양한 분야에 적용이 가능하다. 따라서, 본 모델은 한정된 분야가 아닌 다양한 분야에서 인사이트를 제공할 수 있다는 시사점을 가진다.

본 논문의 한계와 추후 연구 방향은 다음과 같다. 우선 학술적인 관점에서는 합성곱 계층 관련 한 것과 하이퍼파라미터 조정 등에 관한 것으로 나눌 수 있다. 본 연구에서는 합성곱 계층을 1D, 2D로 모두 구현하여 성능을 비교하였는데, 어떤 상황에서 1D 또는 2D가 적절한지에 대한 추가 연구가 필요하다. 또한, 연구 시 한정된 자원으로 인해 다양하게 하이퍼파라미터 조정하여 실험할 수 없었다. TAS-BERT의 논문에서 최고 성능을 기록했던 “최대 문장 길이 128”로 모델을 학습시키지 못했다는 한계점이 존재한다. 또한, 기존 손실과 보조적인 손실을 더하는 비율을 다양하게 조정하기 위한 실험이 추가적으로 필요

하다. 마지막으로, 연구 분야의 특성상 연구 속도가 빠르기 때문에 추후 BARTABSA(Yan et al. 2021) 등 최신 모델과의 비교가 필요하다.

참고문헌(References)

- Ahn, C. a
- Azhar, A. N., M. L. Khodra, “Fine-tuning Pretrained Multilingual BERT Model for Indonesian Aspect-based Sentiment Analysis”, ArXiv, Vol.1, (2021).
- Banerjee, P., K. K. Pal, F. Wang, C. Baral, “Variable Name Recovery in Decompiled Binary Code using Constrained Masked Language Modeling”, ArXiv, Vol.1, (2021).
- Brun, C., and V. Nikoulina, “Aspect based sentimentanalysis into the wild”, WASSA@EMNLP, Vol.1, (2018), 116~122.
- Chaybouti, S., A. Saghe, A. Shabou, “MIX : a Multi-task Learning Approach to Solve Open-Domain Question Answering Edit social preview”, Arxiv, Vol.2, (2021).
- Devlin, J., M. W. Chang, K. Lee, and K. Toutanova, “BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding”, arXiv, Vol.2, 1810.04805(2018).
- He, R.; W. S. Lee, H. T. Ng, and D. Dahlmeier, “An unsupervised neural attention model for aspect extraction”, ACL, Vol.1, (2017), 388~397.
- Khezrian, N., J. Habibi, I. Annamradnejad, “Tag Recommendation for Online Q&A Communities based on BERT Pre-Training Technique”, ArXiv, Vol.1, (2020).
- Kim, Y., “Convolutional Neural Networks for Sentence Classification”, EMNLP, Vol.1, (2014), 1746 ~1751.
- Kiritchenko, S., X. Zhu, C. Cherry, and S. Mohammad, “Detecting aspects and sentiment in customer reviews”, SemEval@COLING, Vol.1, (2014), 437~442.
- LeCun, Y., B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, L. D. Jackel, “Backpropagation Applied to Handwritten Zip Code Recognition”, MIT Press, Vol.1, (1989), 541~551.
- Lee, J., “Sentiment Analysis based on Korean using Recurrent Neural Network: focused on Online Movie Review”, Graduate School of Kookmin University, (2018).
- Lee, J., “Sentimental Analysis Research Trends”, KIPS, No.25, (2018), 358~361.
- Li, X.; L. Bing, P. Li, and W. Lam, “A unified model for opinion target extraction and target sentiment prediction.”, AAAI, Vol.1, (2019), 6714~6721.
- Ma, Y.; H. Peng, and E. Cambria, “Targeted aspect-based sentiment analysis via embedding commonsense knowledge into an attentive LSTM”, AAAI, (2018), 5876~5883.
- Moon, H.C, A. Yang, and J.G. Kim, “CNN-Based Hand Gesture Recognition for Wearable Applications”, Journal of Broadcast Engineering, (2018), 246~252.
- Movahedi, S., E. Ghadery, H. Faili, and A. Shakery, “Aspect Category Detection via Topic-Attention Network”, ArXiv, (2019).
- Pontiki, M., D. Galanis, H. Papageorgiou, S. Manandhar and I. Androustopoulos, “Semeval-2015 task 12: Aspect based sentiment analysis”, SemEval, Vol.1, (2015), 486~495.

- Pontiki, M., D. Galanis, H. Papageorgiou, S. Manandhar and I. Androutsopoulos, S. Manandhar, M. Al-Smadi, M. Al-Ayyoub, Y. Zhao, B. Qin, O. D. Clercq, V. Hoste, M. Apidianaki, N. V. Tannier, X.; Loukachevitch, E. V. Kotelnikov, N. Bel, S. M. J. Zafra and G. Eryigit, “Semeval-2016 task 5: Aspect based sentiment analysis”, *SemEval*, Vol.1, (2016), 19~30.
- Raj, D., S.K. Sahu, and A. Anand. “Learning Local and Global Contexts Using a Convolutional Recurrent Network Model for Relation Classification in Biomedical Text”, *CoNLL*, (2017), Vol. 1, 311~321.
- Schmitt, M., S. Steinheber, K. Schreiber and B. Roth, “Joint aspect and polarity classification for aspect-based sentiment analysis with end-to-end neural networks”, *EMNLP*, Vol.1(2018), 1109~1114.
- Seo, S. and J. Kim, “Sentimental Analysis Research Trends based on Deep Learning”, *KMMS*, Vol.3, No.25 (2016), 8~22.
- She, X., D. Zhang, “Text Classification Based on Hybrid CNN-LSTM Hybrid Model”, *ISCID*, (2018), 185~189.
- Sun, C., L. Huang, and X. Qiu, “Utilizing BERT for Aspect-Based Sentiment Analysis via Constructing Auxiliary Sentence”, *NAACL-HLT*, Vol.1, (2019), 380~385.
- Vaswani A., N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser and I. Polosukhin, “Attention Is All You Need”, *arXiv*, Vol.5, (2017).
- Wan, H., Y. Yang, J. Du, Y. Liu, K. Qi, and J. Z. Pan, “Target-Aspect-Sentiment Joint Detection for Aspect-Based Sentiment Analysis”, *AAAI*, Vol.34, No.5(2020), 9122~9129.
- Wang, Y., M. Huang, X. Zhu and L. Zhao, “Attention-based LSTM for aspect-level sentiment classification”, *EMNLP*, Vol.1, (2016), 606~615.
- Wyk, G. J. v., and A. S. Bosman, “Evolutionary Neural Architecture Search for Image Restoration”, *IJCNN*, Vol.2, (2019).
- Xenos, D., P. Theodorakakos, J. Pavlopoulos, P. Malakasiotis, and I. Androutsopoulos, “Ensembles of Classifiers and Embeddings for Aspect Based Sentiment Analysis”, *SemEval@NAACL-HLT*, Vol.1, (2016), 312~317.
- Xue, W. and T. Li, “Aspect based sentiment analysis with gated convolutional networks”, *ACL*, Vol.1, (2018), 2514~2523.
- Xue, W., W. Zhou, T. Li and Q. Wang, “A Neural Multi-task Model for Aspect Category Classification And Aspect Term Extraction on RestaurantReviews”, *IJCNLP*, Vol.2, (2017), 151~156.
- Yan, H., J. Dai, T. Ji, X. Qiu and Z. Zhang, “A Unified Generative Framework for Aspect-Based Sentiment Analysis”, *arXiv preprint arXiv:2106.04300*, (2021).

Abstract

Target-Aspect-Sentiment Joint Detection with CNN Auxiliary Loss for Aspect-Based Sentiment Analysis

Min Jin Jeon* · Ji Won Hwang** · Jong Woo Kim***

Aspect Based Sentiment Analysis (ABSA), which analyzes sentiment based on aspects that appear in the text, is drawing attention because it can be used in various business industries. ABSA is a study that analyzes sentiment by aspects for multiple aspects that a text has. It is being studied in various forms depending on the purpose, such as analyzing all targets or just aspects and sentiments. Here, the aspect refers to the property of a target, and the target refers to the text that causes the sentiment. For example, for restaurant reviews, you could set the aspect into food taste, food price, quality of service, mood of the restaurant, etc. Also, if there is a review that says, "The pasta was delicious, but the salad was not," the words "steak" and "salad," which are directly mentioned in the sentence, become the "target."

So far, in ABSA, most studies have analyzed sentiment only based on aspects or targets. However, even with the same aspects or targets, sentiment analysis may be inaccurate. Instances would be when aspects or sentiment are divided or when sentiment exists without a target. For example, sentences like, "Pizza and the salad were good, but the steak was disappointing." Although the aspect of this sentence is limited to "food," conflicting sentiments coexist. In addition, in the case of sentences such as "Shrimp was delicious, but the price was extravagant," although the target here is "shrimp," there are opposite sentiments coexisting that are dependent on the aspect. Finally, in sentences like "The food arrived too late and is cold now." there is no target (NULL), but it transmits a negative sentiment toward the aspect "service." Like this, failure to consider both aspects and targets - when sentiment or aspect is divided or when sentiment exists without a target - creates a dual dependency problem.

To address this problem, this research analyzes sentiment by considering both aspects and targets

* Department of financial management, Hanyang University

** School of Business, Hanyang University

*** Corresponding author: Jong Woo Kim

School of Business, Hanyang University

222 Wangshimni-ro, Seongdong-gu, Seoul 04763, Korea

Tel: +82-2-2220-1067, Fax: +82-2-2220-1169, E-mail: kjw@hanyang.ac.kr

(Target-Aspect-Sentiment Detection, hereby TASD). This study detected the limitations of existing research in the field of TASD: local contexts are not fully captured, and the number of epochs and batch size dramatically lowers the F1-score. The current model excels in spotting overall context and relations between each word. However, it struggles with phrases in the local context and is relatively slow when learning. Therefore, this study tries to improve the model's performance.

To achieve the objective of this research, we additionally used auxiliary loss in aspect-sentiment classification by constructing CNN(Convolutional Neural Network) layers parallel to existing models. If existing models have analyzed aspect-sentiment through BERT encoding, Pooler, and Linear layers, this research added CNN layer-adaptive average pooling to existing models, and learning was progressed by adding additional loss values for aspect-sentiment to existing loss. In other words, when learning, the auxiliary loss, computed through CNN layers, allowed the local context to be captured more fitted. After learning, the model is designed to do aspect-sentiment analysis through the existing method.

To evaluate the performance of this model, two datasets, SemEval-2015 task 12 and SemEval-2016 task 5, were used and the f1-score increased compared to the existing models. When the batch was 8 and epoch was 5, the difference was largest between the F1-score of existing models and this study with 29 and 45, respectively. Even when batch and epoch were adjusted, the F1-scores were higher than the existing models. It can be said that even when the batch and epoch numbers were small, they can be learned effectively compared to the existing models. Therefore, it can be useful in situations where resources are limited.

Through this study, aspect-based sentiments can be more accurately analyzed. Through various uses in business, such as development or establishing marketing strategies, both consumers and sellers will be able to make efficient decisions. In addition, it is believed that the model can be fully learned and utilized by small businesses, those that do not have much data, given that they use a pre-training model and recorded a relatively high F1-score even with limited resources.

Key Words : Online review analysis, ABSA, TASD, CNN, BERT

Received : August 10, 2021 Revised : November 12, 2021 Accepted : November 22, 2021

Corresponding Author : Jong Woo Kim

저 자 소개



전민진

현재 한양대학교 파이낸스경영학과, 컴퓨터소프트웨어학부 학생으로 재학중이다. 한양대학교 대학혁신지원사업 비즈니스 랩 기반 프로젝트 학기제 인턴 과정에서 Business & AI Lab에 근무하였다. 주요 관심분야는 기계학습 및 딥러닝, 자연어처리이다.



황지원

한양대학교에서 경영학부 학사를 취득하였으며 한양대학교 대학혁신지원사업 비즈니스 랩 기반 프로젝트 학기제 인턴 과정에서 Business & AI Lab에 근무하였다. 주요 관심분야는 데이터 마케팅, 기계학습 및 딥러닝 기법의 비즈니스 활용, 시장 조사 분석 등이다.



김종우

현재 한양대학교 경영대학 경영학부 교수로 재직 중이다. 서울대학교 수학과에서 학사를 마쳤으며, 한국과학기술원에서 경영과학으로 석사학위를, 산업경영학으로 박사학위를 취득하였다. 주요 연구 관심분야는 데이터마이닝 기법과 응용, 기계학습과 딥러닝, 텍스트 마이닝, 오피니언 마이닝, 상품추천기술, 지능형 정보시스템, 집단지성, 사회 네트워크 분석 등이다.