

# 생성 모델 기반 속성기반 감성 분류

류상연

가천대학교 소프트웨어학과

[teryas@gachon.ac.kr](mailto:teryas@gachon.ac.kr)

## Generative-model based Aspect-Based sentiment Analysis

Sangyeon YU

Gachon University

### 요 약

인터넷과 소셜미디어 사용량의 급증으로, 제품 리뷰, 온라인 피드백, 소셜 미디어 게시물 등을 통해 고객의 감정을 파악하는 것이 중요해졌다. 이에 따라 인공지능이 활용되어 고객이 제품이나 서비스의 어떤 부분에 만족하거나 불만을 가지는지를 분석하는 연구가 활발히 진행되고 있다. 그러나, 이러한 속성 기반 감성 분석이 해외에서는 활발하게 이루어지는 반면, 국내에서는 상대적으로 부족한 상황이다. 이 연구에서는 Aspect-Based sentiment Analysis의 두 개의 주요 작업인 Aspect Category Detection(ACD)과 Aspect Sentiment Classification(ASC)에 대해 생성 모델 중 하나인 T5 모델을 사용하는 방법론을 제시한다. 이 방법론은 기존 판별 모델을 사용하는 것에 비해 시간과 성능 측면에서 크게 향상되었음을 보여준다.

주제어: 자연어 생성, 속성 기반 감성 분류, 감성 분석

### 1. 서 론

감성분석(Sentiment Analysis)은 텍스트 데이터에서 긍정적이거나 부정적인 감성 또는 감정을 탐지하고 분류하는 과정을 말한다. 자연어 처리의 감성 분석은 다음과 같은 주요 카테고리로 분류할 수 있다. 텍스트가 긍정적인지 부정적인지 중립인지를 판단하는 극성(Polarity) 분석, 텍스트에서 여러 가지 감정을 탐지하고 강도를 측정하는 감성 스코어링(Emotion scoring), 텍스트에서 특정 속성이나 측면에 대한 감성을 분석하는 속성 기반 감성 분석(Asspect-Based Sentiment Analysis)이 있다.[1]

인터넷 및 소셜미디어의 사용량이 증가함에 따라 제품 리뷰, 온라인 피드백, 소셜 미디어 게시물 등을 통해 고객의 감정 상태를 이해하는 것이 일반화되고 있다. 기존의 감성 분석은 텍스트 또는 문장이 전체적으로 긍정적인지, 부정적인지를 판별하는 것을 중점으로 삼았다. 그러나 마케팅 전문가들은 감성 분석을 활용하여, 단순히 긍정적이거나 부정적인 평가 이상의 정보, 즉 고객이 어떤 부분에서 제품이나 서비스에 만족하거나 불만을 가지는지에 대한 상세한 정보를 요구하고 있다. 이렇게 여러 특성을 추출하여 특성별 감성 분석을 실시하는 것이 속성 기반 감성 분석(ABSA, Aspect-Based Sentiment Analysis)이다. 이러한 속성별 감성 분석이 최근에 주목받는 이유는 고객의 선호도를 특성별로 파악하고, 이를 통해 더욱 세밀하고 정확한 통찰력을 얻어 분석 결과의 유용성을

극대화할 수 있기 때문이다.

$S_i$  : 양도 많고 전용 패드도 너무너무 맘에듬

Subtask	Input	Output
Aspect Category Detection (ACD)	$S_i$	본품#일반,패키지 /구성품#일반
Aspect Sentiment Analysis (ASC)	$S_i$	positive

표 1. ABSA subtask ACD와 ASC중,  $S_i$ 는  $i$ 번째 입력된 문장일 때 입력값과 출력값,

이러한 속성 기반 감성 분석은 여러한 하위 태스크로 나뉘는데, 그 중 본 논문에서 다룰 태스크는 Aspect Category Detection (ACD)와 Aspect Sentiment Classification (ASC)이다.

ACD는 주어진 텍스트에서 어떤 속성이나 면이 언급되었는지를 찾는 태스크이다. 예를 들어 표 1에 나온 양도 많고 전용 패드도 너무너무 맘에듬 이라는 리뷰에서 ACD는 본품과 구성품 이라는 속성을 탐지하고 이러한 태스크는 텍스트에서 어떤 주제가 중점적으로 다루어졌는지를 파악하는데 유용하며, 이를 통해 특정 대상의 특정 속성에 대한 공감도를 알 수 있다.

분류	내용	예시	비고
입력	문장	좋아하는 뷰티유튜버가 만든 쿠션제품이라고 해서 어떨지 넘나 궁금했는데, 직접 써보니 대박 촉촉하고 컬러감 대박	문자열
출력	튜플 (ACD, ASC)	(제품 전체#인지도, 긍정) (본품#품질, 긍정) (본품#일반, 긍정)	튜플

표 2. 속성 기반 감성 분석 과제의 모델 입력과 출력의 예

ASC는 반면에 텍스트에서 언급된 특정 측면에 대한 감성을 판별하는 태스크이다. 이는 해당 측면에 대한 일반적인 감정이 긍정적인지, 부정적인지, 또는 중립적인지를 결정한다. 위 표1의 예시에서 본품, 패키지 구성품에 대한 감성은 긍정적으로 판별될 수 있다.

본논문은 이러한 ABSA의 하위 태스크인 ACD와 ASC를 생성모델 중 T5[2]를 기반으로 감성 분류를 제안한다.

## 2. 관련 연구

### 2.1 언어 모델

다양한 언어 모델들 중 대표적인 생성 모델로는 Google이 개발한 자연어 처리(NLP)모델 T5(Text-to-Text Transfer Transformer)는 인코더-디코더 구조의 Transformer[3] 모델을 기반으로 한다. 기존의 Transformer 모델들과는 달리 T5는 모든 자연어 처리 Task를 생성 문제로 재구성한다. 이러한 접근 방식을 통해, T5는 자연어 이해(NLU)와 자연어 생성(NLG)작업 모두를 동일한 프레임 워크 내에서 처리할 수 있다. T5는 사전 학습 단계에서 비지도 학습 방법을 사용한다. 이때 입력 텍스트에서 일부 단어나 문구를 마스킹하고, 이 마스킹된 부분을 예측하는 방식으로 학습이 이루어진다. 이는 BERT에서 사용한 방법과 유사하나, 텍스트-투-텍스트 형식을 사용하기 때문에, 마스킹된 부분을 별도로 표시하지 않고, 모델이 자연스럽게 마스킹된 부분을 예측하도록 한다.

### 2.2 속성 기반 감성 분석

속성 기반 감성 분석(ABSA)는 2010년대 초에 등장하였고, 텍스트와 감성 분석 분야에서 중요한 연구 주제로 인식되었다. SemEval(Semantic Evaluation) 2014[4]가 이 분야의 주요 데이터셋으로, 랩톱과 레스토랑 리뷰에서 중요한 속성을 식별하고, 해당 속성에 연관된 감정 극성을 부여하는 작업을 포함하고 있다.

전통적인 감성 분석은 문서나 텍스트 조각 전체의 감성을 판단하는 반면, ABSA는 텍스트 내 특정 속성이나 요소에 초점을 맞춘 감성 분석을 수행한다. 예를 들어, '드라이버 업데이트는 잘 됐지만, BIOS 업데이트가 시스템을 열리고 컴퓨터가 꺼졌다.'라는 문장에서 '드라이버'에 대한 감성은

긍정적이고, 'BIOS'에 대한 감성은 부정적임을 알 수 있다. 하지만 전통적인 감성 분석에서는 '잘 됐지만'라는 긍정적인 단어 때문에 전체 문장이 긍정적인 감성으로 판단될 수 있다.

따라서 ABSA의 장점은 문장 내에서 언급된 특정 속성에 대한 감성을 더 세밀하게 파악할 수 있다는 점이다. 이는 사용자의 특정 특징이나 요소에 대한 반응을 보다 정확하게 이해하는 데 도움이 된다.

### 2.3 인코더 기반 감성 분류

전통적인 속성 기반 감성 분류 모델은 주로 BERT[5], Roberta[6], Electra[7] 인코더만을 사용하면서 인코더-디코더 구조를 활용하지 않았다.

그러나, 속성 간의 감성이 상호연관성을 가질 수 있다는 연구 결과에 따라, 이러한 감성들이 주변 속성과 유사성을 보이는 '속성 감성 일관성'이라는 새로운 개념이 제안되었다. 이를 바탕으로, 연구자들은 속성 감성 일관성의 모델링에 초점을 맞추어 이전 연구에서 놓쳤던 간극을 해결하는 로컬 감성 통합 패러다임(Local Sentiment Aggregation paradigm, LSA)을 도입하였다.[8]

또한, 지역 컨텍스트 포커스(Local Context Focus, LCF) 메커니즘을 기반으로, 속성 용어 추출과 그 극성을 동시에 추론하는 새로운 접근법이 제시되었다. 이러한 접근법은 기존 연구에서 주로 속성 용어의 극성에만 초점을 맞추던 방식과는 달리, 속성 용어의 추출에 중점을 두는 새로운 시각을 제공하였다.[9]

### 2.4 Instruction learning 기반 감성 분류

InstructABSA[10]는 전통적인 인코더 기반 감성 분류 방법과 대비하여 Instruction learning을 활용하는 방법을 제안했다. 이는 Tk-Instruction 모델[11]을 기반으로 하였는데, 학습 과정에서는 SUPER-NATURALINSTRUCTIONS라는 자연어 처리(NLP) 태스크 분류 벤치마크를 활용하여 T5 모델에 Instruction learning 기법을 적용한 것이다. 이로 인해 모델이 다양한 컨텍스트에서 지시문을 이해하고 따르는 능력을 향상되었다. 이를 바탕으로 InstructABSA는 ABSA에 적합한 Instruction Prompt를 작성하여 Tk-Instruction 모델에 적용하는 새로운 접근 방식을 제안하였다.

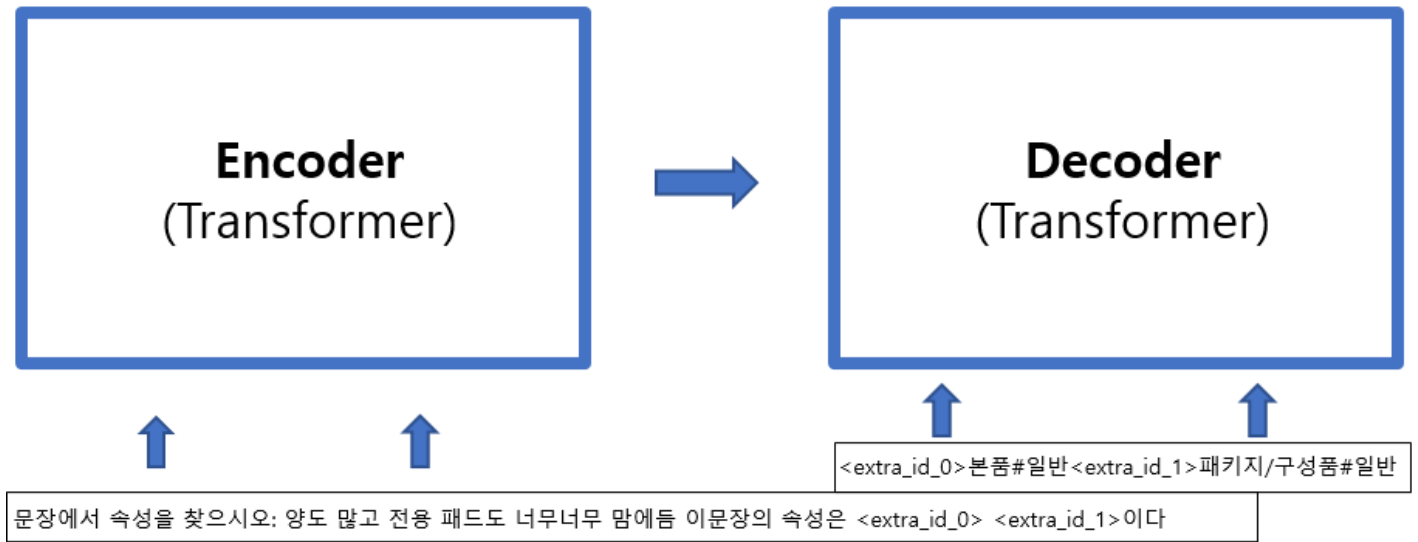


그림 1. Aspect Category Detection의 모델구조

### 3. 데이터셋 및 모델링

#### 3.1 데이터셋

본논문에서 사용한 데이터셋은 2022 국립국어원 인공 지능 언어 능력 평가 대회 말뭉치\_속성 기반 감성 분석 데이터셋이다.

	훈련	검증	시험
문장 수	3001	2794	2127

표 3. 실험 데이터셋

데이터셋의 형태는 표 2과 같으며, 여기서 속성 범주 탐지 (Aspect Category Detection, ACD)는 개체와 속성의 조합을 '#'를 사용하여 표시한다. 이때, 개체와 속성의 쌍은 총 28개로 구성되어 있다. 속성 감성 분류 (Aspect Sentiment Classification, ASC)는 이전 단계에서 추출된 개체#속성 범주에 대한 화자의 감성을 긍정, 부정, 중립 중 하나로 분류하는 과정이다.

#### 3.2 사전 학습 모델

베이스 라인으로는 pko-t5를 사용하였으며 pko-t5는 sentencepiece 대신 OOV 가 없는 BBPE를 사용했으며 한국어 데이터 (나무위키, 위키피디아, 모두의 말뭉치 등..)를 T5의 span corruption task를 사용해서 unsupervised learning만 적용하여 학습을 진행하였다

#### 3.3 모델링

ABSA 각각의 TASK에 ACD와 ASC의 알맞게 두 가지의 형태로 모델 구조를 제안한다.

ACD의 경우 T5의 Encoder 부분에 Prompt의 형태를 주어 입력값을 주었으며 표3의 예시처럼 문장에서 속성을 찾으시오 문장 그리고 이 문장의 속성은 <extra\_id\_N>의 토큰 형태로 입력값을 주었고, 그에 대응하는 정답 라벨을 Decoder에 '<pad><extra\_id\_N>+정답N'의 형태로 입력하였다. 이때 Decoder의 최종 logits을 출력하고 입력된 labels와 Cross-Entropy loss function을 이용하고 최종 loss를 계산한다.

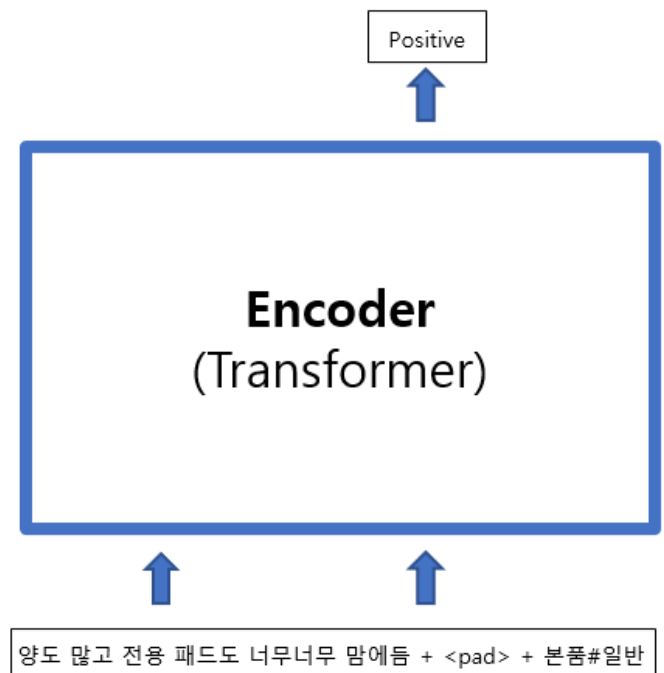


그림 2. Aspect sentimental Classification의 모델구조

ASC의 경우 ACD에서 찾아낸 특성쌍을 해당쌍이 어떤 극성을 가지는지 찾기 위해 문장과 ACD의 특성쌍을 같이 입력하여 ACD 모델과 같이 입력된 labels와 Cross-Entropy loss function을 이용하고 최종 loss를 계산한다

#### 4. 실험 및 결과

##### 4.1 실험

ACD 데이터에 대한 미세조정은 학습률 5e-6, 배치 사이즈 4으로 1 ~ 2 Epoch 진행하였다 ASC 데이터에 대한 미세조정은 학습률 5e-5, 배치 사이즈 16으로 1 ~ 2 Epoch 진행하였다 .GPU 자원은 TitanXp 1기를 사용하였다.

##### 4.2 결과

표3과 표4는 모델 평가 결과이다. 두 가지 형태의 Task 모두 실험에서 제안 모델은 비교 모델 보다 높은 성능을 보였다. ACD에서 모델이 잘 학습했는지 확인하기 위해 평가 지표를 정확도, F1-score, 한 에폭당 걸린 평균 시간을 비교하였다.

Model	Accuracy	F1-score	Time
Klue-Roberta	0.64	0.68	2100s
XLNet-Roberta	0.63	0.56	1937s
Electra	0.70	0.69	1427s
Ours(T5)	0.72	0.72	134s

표 3. Aspect Category Detection 모델 성능 평가

제안 모델은 비교 모델과 비교하여 정확도에선 최대 8%의 성능 차이를 보였고, F1-Score 에서는 최대 4%의 성능 차이를 보였다. 또한 한 에폭당 시간을 비교해 보면 약 10 배 정도의 시간이 단축됨을 확인할 수 있었다.

Model	Accuracy	F1-score(macro)	Time
Klue-Roberta	0.9722	0.5751	85s
XLNet-Roberta	0.973	0.5269	81s
Electra	0.9741	0.5842	74s
Ours	0.9747	0.5594	91s

표 4. Aspect Sentiment Classification 모델 성능 평가

제안 모델은 비교 모델과 비교하여 정확도의 경우 최대 0.2%의 성능 향상이 있었지만, F1-score과 한 에폭당 시간을

비교하면 비교 모델이 비해 상향되지 못함을 확인할 수 있었다.

#### 4.3 실험 결과 분석

ACD 의 경우 비교 모델은 28 개의 속성쌍에 대하여 각각 비교를 한 것에 반면에 제안 모델은 생성된 정답을 비교하는 방식으로 진행하였다. 이 때문에 한 에폭당 걸리는 시간이 최대 약 13 배 정도 차이 나는 유의미한 결과가 도출되었다. 또한 정확도나 F1-score 자체에서도 각각 최대 8%,4%의 성능 향상을 이뤘고 이는 생성 모델이 속성쌍을 비교하여 카테고리를 생성하는 능력이 Encoder 기반 모델보다 높음을 확인할 수 있었다. 그러나, ACD 모델은 특정 속성에 대해 정확한 답변을 도출하지 못하는 경향이 있었다. 이는 그림 3 에서 확인할 수 있듯이, 훈련 데이터 셋의 분포가 균일하지 않고, 특정 데이터가 더 많이 존재하는 경향 때문임을 확인할 수 있었다.

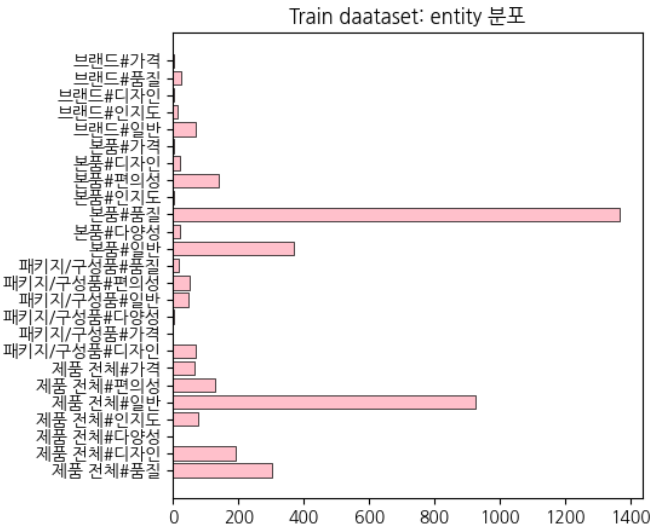


그림 3. 실험 데이터셋 속성 쌍 분포

ASC 의 경우 ACD 의 방식과 다르게 정답이 나와있는 속성 쌍에 대한 극성을 분류하였기 때문에 파라미터 수가 많은 T5 모델이 한 에폭당 걸리는 시간이 조금 더 걸렸음을 확인하였고 성능 자체도 F1-score 는 비교 모델 보다 낮음을 확인할 수 있었다.

ASC 의 경우 F1-score(macro) 점수가 낮게 나옴을 확인할 수 있었는데 이는 사용한 말뭉치\_속성 기반 감성 분석 데이터 셋이 표 5 와 같이 데이터의 불균형으로 인하여 발생한 것으로 추측된다

감성	총 합계
Positive	6028
Neutral	149
Negative	86

표 5. 실험 데이터셋 감성 분포

5. 결론

본 논문은 속성 기반 감성 분석(ABSA, Aspect-Based Sentiment Analysis)을 기존 연구의 Encoder 기반 방식이 아닌 Encoder-Decoder 를 활용한 생성 모델 기반의 방식을 제안하였다. 이러한 방식을 통해 ABSA 의 하위 태스크인 (ACD, Aspect Category Detection)에서 성능 평가 결과 정확도, F1-Score 에서 성능 향상을 확인할 수 있었고 한 에폭당 경과 시간을 단축함을 확인할 수 있었다. 그러나 (ASC, Aspect Sentiment Classification)에서의 성능 평가 결과 비교 모델과의 성능향상을 이루지 못함을 확인할 수 있었다.

이를 해결하기 위해 다른 형태의 모델 구조를 시도해 볼 계획이다. 관련 연구인 Insturction learning을 기반으로 하여 한국어 만의 Task-insturction을 이용하여 ACD, ASC의 Task에 적용해 볼 계획이고, 텍스트 만의 속성 기반 감성 분석뿐만 아니라 이미지를 같이 활용하는 멀티 모달[12] 속성 기반 감성 분석에도 적용하여 모델을 제안할 계획이다.

## 참고 문헌

- [1] 이희재, 임희석.(2022).속성 기반 감성 분석을 활용한 BERT 감성 분류 모델의 성능 개선.한국정보과학회 학술발표논문집,( ),332-334.
- [2] RAFFEL, Colin, et al. Exploring the limits of transfer learning with a unified text-to-text transformer. *The Journal of Machine Learning Research*, 2020, 21.1: 5485-5551.
- [3] VASWANI, Ashish, et al. Attention is all you need. *Advances in neural information processing systems*, 2017, 30.
- [4] KIRANGE, D. K.; DESHMUKH, Ratnadeep R.; KIRANGE, M. D. K. Aspect based sentiment analysis semeval-2014 task 4. *Asian Journal of Computer Science and Information Technology (AJCSIT) Vol*, 2014, 4.
- [5] DEVLIN, Jacob, et al. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*, 2018.
- [6] LIU, Yinhan, et al. Roberta: A robustly optimized bert pretraining approach. *arXiv preprint arXiv:1907.11692*, 2019.
- [7] CLARK, Kevin, et al. Electra: Pre-training text encoders as discriminators rather than generators. *arXiv preprint arXiv:2003.10555*, 2020.
- [8] YANG, Heng; LI, Ke. Improving Implicit Sentiment Learning via Local Sentiment Aggregation. *arXiv e-prints*, 2021, arXiv: 2110.08604.
- [9] YANG, Heng, et al. A multi-task learning model for chinese-oriented aspect polarity classification and aspect term extraction. *Neurocomputing*, 2021, 419: 344-356.
- [10] SCARIA, Kevin, et al. InstructABSA: Instruction Learning for Aspect Based Sentiment Analysis. *arXiv preprint arXiv:2302.08624*, 2023.
- [11] WANG, Yizhong, et al. Super-naturalinstructions: Generalization via declarative instructions on 1600+ nlp tasks. In: *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*. 2022. p. 5085-5109.
- [12] LING, Yan, et al. Vision-language pre-training for multimodal aspect-based sentiment analysis. *arXiv preprint arXiv:2204.07955*, 2022.