

## 과제 기술서: 속성 기반 감성 분석<sup>1)</sup>

### 1. 과제 개요

감성 분석이란 화자의 의견이나 긍·부정의 태도가 나타나는 문장의 감성 극성(긍정 또는 부정)을 분석하여 정량화하는 것이다. 주로 문장 단위로 긍정, 부정의 유무 혹은 정도를 평가해 왔으며, 제품 및 여행의 후기 분석, 추천 시스템에 널리 활용되고 있다.

속성 기반 감성 분석은 언어 자료에 나타난 개체와 속성 정보를 고려한 감성 분석 방법으로 일반적인 감성 분석에 비해 더 세밀한 정보를 얻을 수 있다는 장점이 있다. 예를 들어, 음식점 도메인의 후기 ‘콩국수가 싸서 좋다’라는 문장에서 일반적인 감성 분석의 결과는 ‘긍정’이 되지만, 속성 기반 감성 분석의 결과는 ‘개체 카테고리(개체): {음식(콩국수)}, 속성: {가격}, 감성: {긍정}’이 되므로 더 많고 구체적인 정보를 얻을 수 있다.

입력 문장			
편안해서 요즘 인기 많던데.. 프리미엄시그니처는 디자인도 예뻐하네..			
분석 결과(tuple)			
편안해서 요즘 인기 많던데.. 프리미엄시그니처는 디자인도 예뻐하네..			
ACD	제품 전체#인지도	ACD	제품 전체# 디자인
ASC	positive	ASC	positive

그림 1 속성 기반 감성 분석 입출력의 예

### 2. 과제 정의

이 과제는 국립국어원 ‘2021년 말뭉치 감성 분석 및 연구’ 사업에서 구축한 속성 기반 감성 분석 말뭉치 중 “제품” 도메인을 대상으로 제작되었다. 이 과제의 세부 과업은 속성 기반 감성 분석에서 속성 범주 탐지(Asspect Category Detection, ACD)와 속성 감성 분류(Asspect Sentiment Classification, ASC)이다.

항목	내용	예시
ACD	속성 범주 탐지	제품 전체#인지도
ASC	속성 감성 분류	긍정

표 1 세부 과업

속성 범주 탐지(Asspect Category Detection, ACD)란 주어진 문장에서 분석의 대상이 되는 범주들을 추출하는 것으로, ‘#’를 구분자로 하는 ‘개체’와 ‘속성’의 조합으로 표기한다. 예를 들어 ‘제품 전체#디자인’과 같이 개체 ‘제품 전체’와 속성 ‘디자인’을 조합한 것이 범주이다. 만약 한 문장의 범주가 ‘제품 전체#디자인’으로 주석되어 있다면 이는 필자가 제품 전체의 디자인에 대해 긍정 또는 부정의 감성을 느꼈다는 뜻이다. 전체 데이터 세트는 화자의 주관성이 반영된 문장으로만 구성된다. 아래 표 2와 같이, 제품 분야에는 여러 개체와 속성들이 존재하며, 속성 범주 탐지 과업에서는 문장에서 속성 범주를 추출한다(본 품#디자인, 브랜드#인지도 등).

1) ‘2022년 국립국어원 인공지능의 언어 능력 평가 대회’ 과제와 같음



구분	항목	설명
개체	제품 전체	제품 전체
	본품	특정 장치 및 부품 등 본품 하위 구성 요소
	패키지/구성품	제품 패키지 및 구성품(예: 브러쉬, 펌프)
	브랜드	브랜드 이미지
속성	일반	개체 자체
	가격	제품의 경제적, 마케팅 요소 평가
	디자인	형태, 색상 등 제품 외형 디자인 평가
	품질	제품의 기능, 성능, 성분, 소재, 효과 등 평가
	편의성	사용 난이도 및 편의, 사이즈 언급 등 휴대성 평가
	다양성	제품 구성품의 다양성 및 유형, 색상에 따른 제품 다양성
	인지도	유명도, 인기, 판매량 언급 등

표 2 제품 분야의 개체와 속성

속성 감성 분류(Aспект Sentiment Classification, ASC)란 속성 범주 탐지 과업을 통해 추론한 속성 범주에 대한 화자의 감성을 긍정, 부정, 중립 중 하나로 분류하는 과업이다. 즉 하나의 문장에 나타나는 속성 범주 각각에 대한 감성을 분류한다. 예를 들어 ‘아이폰은 이쁘데 비싸요.’와 같은 문장의 경우, ‘제품 전체#디자인’은 긍정으로 분류할 수 있으나, ‘제품 전체#가격’은 부정으로 분류할 수 있다.

다시 말해, 속성 기반 감성 분석 과제는 위에서 언급된 속성 범주 탐지와 속성 감성 분류 과업을 모두 수행하여, 입력 문장에 대해 속성 범주를 추출하고, 각각에 대한 감성을 분류하는 것이다. 각 참가 팀의 제출 결과는 F1 점수로 평가되어 순위표(리더보드)에 게시된다.

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
평가		F1 점수	

표 3 속성 기반 감성 분석 과제의 입력과 출력의 예

평가는 정답 데이터 세트와 예측 데이터 세트의 주석(annotation)을 문장 단위로 비교하여 F1 점수로 측정한다.

$\text{Precision} = \text{TP} / (\text{TP} + \text{FP}), \text{Recall} = \text{TP} / (\text{TP} + \text{FN}), \text{F1 Score} = 2 * (\text{Precision} * \text{Recall}) / (\text{Precision} + \text{Recall})$ <ul style="list-style-type: none"> <li>- True positive: 예측한 속성 범주와 감성 쌍이 정답에 있는 경우</li> <li>- False positive: 예측한 속성 범주와 감성 쌍이 정답에 없는 경우(속성 범주만 맞춘 경우 포함)</li> <li>- False negative: 정답 데이터 세트에 있는 속성 범주와 감성 쌍을 예측하지 못한 경우(속성 범주만 맞춘 경우 포함)</li> </ul>
---

### 3. 데이터 형식

이 과제의 데이터 세트는 제이슨-엘(JSON-L, jsonlines)로 제공되며, 각 문장에 대한 속성 기반 감성 분석 주석이 포함되어 있다. 특히 학습 데이터에는 문장에 개체가 명시적으로 표현되는 경우 이에 대한 정보를 포함하므로(예: ‘패키지#구성품’이 문장에서 ‘전용 패드’로 표현되는 경우) 참고용으로 활용할 수 있다. 제출할 때는 이를 제외하고 개체#속성 범주와 감성으로 구성된 튜플 목록만을 포함한다.

표 4는 데이터 규모를 보여준다. 이 과제의 데이터 세트는 다양한 속성 범주를 포함하므로 정당한 평가를 위해 검증 데이터와 시험 데이터의 양을 충분히 배정하였다. 특히 현실 세계에서의 속성 범주나 감성별 후기의 비중이 같지 않으므로 이와 같은 불균형성을 데이터 세트에 반영하기 위하여 감성 극성별 데이터의 규모를 인위적으로 조절하지 않았다.

구분	훈련	검증	평가
문장 수	2,999	2,792	2,126

표 4 데이터 규모

표 5는 데이터 형식의 예시이다. 주어진 훈련 데이터와 시험 데이터는 동일한 제이슨-엘(JSON-L) 형식으로 제공되며, 평가 데이터의 경우에는 각 문장에 대한 주석(annotation) 항목이 없다. 참가 팀은 해당 목록에 대해 모델의 출력 결과를 추가하여 제출한다.

항목	내용
훈련용 데이터의 예	<pre>... {   "id": "1",   "sentence_form": "양도 많고 전용 패드도 너무너무 맘에듬",   "annotation": [     ["본품#일반", ["양", 0, 1], "positive"],     ["패키지/구성품#일반", ["전용 패드", 6, 11], "positive"]   ] } ...</pre> <p>- 범주와 감성 이외에 추가로 문장 내 개체의 명시적 표현 정보도 포함함. 만약 명시적으로 표시되지 않은 경우에는 [null, 0, 0]의 형태로 제공</p>
평가용 데이터의 예 (제출 전)	<pre>... {   "id": "1",   "sentence_form": "양도 많고 전용 패드도 너무너무 맘에듬",   "annotation": [] } ...</pre> <p>- 학습용 데이터와 동일한 형태 - 주석의 대상이 되는 “annotation”을 비워두고 제공</p>
	<pre>... {   "id": "1",</pre>

제출용 데이터의 예	<pre>"sentence_form": "양도 많고 전용 패드도 너무너무 맘에듬", "annotation": [   ["본품#일반", "positive"],   ["패키지/구성품#일반", "positive"] ] ...</pre>
	<ul style="list-style-type: none"> <li>- 평가용 데이터의 “annotation”에 비어있는 리스트를 채워서 제출</li> <li>- 훈련용 데이터와 달리 범주를 표현하는 문장 내 어휘 정보는 제외하고 범주와 감성으로 구성된 튜플들의 리스트</li> </ul>

표 5 데이터 형식의 예

#### 4. 기준 모델

이 과제의 기준 모델(baseline model)은 깃허브(github)를 통해 공개되어 있다<sup>2)</sup>.

([https://github.com/teddysum/korean\\_ABSA\\_baseline](https://github.com/teddysum/korean_ABSA_baseline))

#### 5. 참가 규정 및 제출 방법

##### (1) 팀 구성

- 국립국어원 ‘언어정보나눔터(모두의 말뭉치, 인공지능(AI)말평)’ 회원은 누구나 참가할 수 있다.

##### (2) 답안 제출

- 참가 팀은 해당 과제를 해결한 결과를 제공하는 양식(샘플) 파일과 동일한 형식으로 작성하여 제출한다.
- 한 참가 팀이 복수의 결과물을 제출할 경우 제출한 모델 및 결과 중 가장 높은 성적을 기준으로 순위표(리더보드)에 게시한다.

##### (3) 모델 사용 및 제출

- 라이선스에 문제가 없는 모델 사용 가능(라이선스의 검토 책임은 참가 팀에게 있음)
- 외부 API를 추론에 직접적으로 이용 불가(예: 챗GPT 결과 제출)
- 외부 데이터 추가 사용 가능
  - 외부에 공개된 모든 데이터를 학습 데이터로 사용 가능
  - 상시 과제로 제공된 데이터는 초거대 언어모델(LLM)을 이용하여 증강 후 학습 데이터로 사용 가능

##### (4) 기타: 시험 데이터의 정답은 공개 계획 없음

#### 6. 문의: 과제의 ‘문의’ 게시판 이용

- 2) ① 해당 모델은 xlm-roberta-base 모델을 사용하여 학습되었고, ACD와 ASC 모델을 파이프라인으로 연결한 모델이다.  
 ② ACD 모델은 xlm-roberta-base 모델의 입력 형태를 <s>sentence\_form</s></s>category</s>와 같 이 하고 각 category별로 문장에서 추출될지 되지 않을지에 대한 이진 분류를 수행한다.  
 - 모델 입력 예시: <s>둘째날은 미친듯이 밟아봤더니 기어가 헛돌면서 틱틱 소리가 나서 경악.</s></s>제품 전체#일반</s>  
 - 모델 출력 예시: 0  
 ③ ASC 모델은 ACD에서 추출된 category에 대해 xlm-roberta-base 모델의 입력 형태를 <s>sentence\_form</s></s>category</s>와 같이 하고 긍정(0), 부정(1), 중립(2)로 분류한다.  
 - 모델 입력 예시: <s>둘째날은 미친듯이 밟아봤더니 기어가 헛돌면서 틱틱 소리가 나서 경악.</s></s>본품#품질</s>  
 - 모델 출력 예시: 1