

Utilizing BERT for Aspect-Based Sentiment Analysis via Constructing Auxiliary Sentence

NAACL 2019

Chi Sun, Luyao Huang, Xipeng Qiu

- Sentiment analysis는 자연어 처리의 중요한 task중 하나.
- sentiment 분석의 기본 가정은 텍스트가 극성(polarity)을 갖는다는 것.
- 문장에는 여러 aspect가 포함될 수 있다.

Ex) “이 책은 하드커버 버전인데 가격이 좀 비싸다”

‘외관’의 극성은 긍정이고, ‘가격’에 대한 극성은 부정이다

- aspect 기반 sentiment 분석(ABSA)은 특정 aspect에 대한 sentiment를 식별하는 것을 목표로 한다.

- ABSA는 sentence-level task지만 해당 task는 여러 target을 가진 문장을 처리할 수 없다.
- "The design of the space is good in Boqueria but the service is horrid, on the other hand, the staff in Gremio are very friendly and the food is always delicious."
- 주어진 대상과 관련된 특정 aspect에 대한 fine-grained option극성을 식별하는 것을 목표로 하는 targeted aspect 기반 정서 분석(TABSA) 작업을 소개한다.
- 작업은 두 단계로 나눌 수 있다.
 - (1) 첫 번째 단계는 각 target과 관련된 aspect를 결정하는 것이고,
 - (2) 두 번째 단계는 target에 대한 aspect의 극성을 풀어내는 것이다.

- BERT의 입력 표현은 단일 문장과 문장 쌍을 모두 나타낼 수 있으므로 (T)ABSA를 문장 쌍 분류 작업으로 변환하고 pre-trained BERT를 fine-tuning한다.

Model	Aspect			Sentiment	
	Acc.	F_1	AUC	Acc.	AUC
LR (Saeidi et al., 2016)	-	39.3	92.4	87.5	90.5
LSTM-Final (Saeidi et al., 2016)	-	68.9	89.8	82.0	85.4
LSTM-Loc (Saeidi et al., 2016)	-	69.3	89.7	81.9	83.9
LSTM+TA+SA (Ma et al., 2018)	66.4	76.7	-	86.8	-
SenticLSTM (Ma et al., 2018)	67.4	78.2	-	89.3	-
Dmu-Entnet (Liu et al., 2018)	73.5	78.5	94.4	91.0	94.8
BERT-single	73.7	81.0	96.4	85.5	84.2
BERT-pair-QA-M	79.4	86.4	97.0	93.6	96.4
BERT-pair-NLI-M	78.3	87.0	97.5	92.1	96.5
BERT-pair-QA-B	79.2	87.9	97.1	93.3	97.0
BERT-pair-NLI-B	79.8	87.5	96.6	92.8	96.9

Table 3: Performance on SentiHood dataset. We boldface the score with the best performance across all models. We use the results reported in Saeidi et al. (2016), Ma et al. (2018) and Liu et al. (2018). “-” means not reported.

2. 연구 방법

- **TABSA**에서 문장은 일반적으로 일련의 단어로 구성된다.
단어 중 일부는 Saeidi 외 연구진(2016)에 따라 사전 식별된 target이다.
- 문장 S , target entity 집합 T 와 aspect 집합 A 가 주어지면 target-aspect pair의 전체 집합 $\{(t, a) : t \in T, a \in A\}$ 에 대한 sentiment 극성 ($y \in \{\text{positive, negative, none}\}$)을 예측하는 3-class 분류 문제로 설정한다.
- **ABSA**에서는 target이 없으므로 target-aspect 쌍 $\{t, a\}$ 은 aspect a 가 된다.

보조 문장 구성

- Sentences for QA-M
 - target-aspect 쌍으로부터 생성하고자 하는 문장은 질문이다.
 - ex) "what do you think of the safety of location - 1 ?"
- Sentences for NLI-M
 - NLI 작업의 경우 문장을 생성할 때 설정하는 조건이 덜 엄격하고 형식이 훨씬 간단하다.
만들어진 문장은 단순한 pseudo 문장이다.
 - ex) "location - 1 - safety".

Example:

LOCATION2 is central London so extremely expensive, LOCATION1 is often considered the coolest area of London.

Target	Aspect	Sentiment
LOC1	general	Positive
LOC1	price	None
LOC1	safety	None
LOC1	transit-location	None
LOC2	general	None
LOC2	price	Negative
LOC2	safety	None
LOC2	transit-location	Positive

Table 1: An example of SentiHood dataset.

Methods	Output	Auxiliary Sentence
QA-M	S.P.	Question w/o S.P.
NLI-M	S.P.	Pseudo-sentence w/o S.P.
QA-B	{yes,no}	Question w/ S.P.
NLI-B	{yes,no}	Pseudo-sentence w/ S.P.

Table 2: The construction methods. Due to limited space, we use the following abbreviations: *S.P.* for *sentiment polarity*, *w/o* for *without*, and *w/* for *with*.

- Sentences for QA-B

- sentiment label 정보를 추가하고 TABSA를 이진 분류 (label \in {yes, no})로 임시 변환하여 확률 분포를 얻는다.

- 이때 각 target-aspect 쌍은

“the polarity of the aspect safety of location – 1 is positive”,

“the polarity of the aspect safety of location – 1 is negative”,

“the polarity of the aspect safety of location – 1 is none”과 같은

세 가지 시퀀스를 생성한다.

- Yes의 확률 값을 일치 점수로 사용한다.

- 세 개의 시퀀스 (positive, negative, none)를 생성하는 target-aspect 쌍의 경우

예측 범주에 대해 일치하는 점수가 가장 높은 시퀀스의 클래스를 사용한다.

Example:

LOCATION2 is central London so extremely expensive, LOCATION1 is often considered the coolest area of London.

Target	Aspect	Sentiment
LOC1	general	Positive
LOC1	price	None
LOC1	safety	None
LOC1	transit-location	None
LOC2	general	None
LOC2	price	Negative
LOC2	safety	None
LOC2	transit-location	Positive

Table 1: An example of SentiHood dataset.

- Sentences for NLI-B

- QA-B와의 차이점은 보조 문장이 질문에서 **pseudo 문장**으로 바뀐다는 것.

- 보조 문장은

"location - 1 - safety - **positive**",

"location - 1 - safety - **negative**",

"location - 1 - safety - **none**" 이다.

보조 문장을 구성한 후에는 TABSA 작업을

단일 문장 분류 작업에서 문장 쌍 분류 작업으로

변환할 수 있다.

Example:

LOCATION2 is central London so extremely expensive, **LOCATION1** is often considered the coolest area of London.

Target	Aspect	Sentiment
LOC1	general	Positive
LOC1	price	None
LOC1	safety	None
LOC1	transit-location	None
LOC2	general	None
LOC2	price	Negative
LOC2	safety	None
LOC2	transit-location	Positive

Table 1: An example of SentiHood dataset.

Model	Aspect			Sentiment	
	<i>Acc.</i>	F_1	AUC	<i>Acc.</i>	AUC
LR (Saeidi et al., 2016)	-	39.3	92.4	87.5	90.5
LSTM-Final (Saeidi et al., 2016)	-	68.9	89.8	82.0	85.4
LSTM-Loc (Saeidi et al., 2016)	-	69.3	89.7	81.9	83.9
LSTM+TA+SA (Ma et al., 2018)	66.4	76.7	-	86.8	-
SenticLSTM (Ma et al., 2018)	67.4	78.2	-	89.3	-
Dmu-Entnet (Liu et al., 2018)	73.5	78.5	94.4	91.0	94.8
BERT-single	73.7	81.0	96.4	85.5	84.2
BERT-pair-QA-M	79.4	86.4	97.0	93.6	96.4
BERT-pair-NLI-M	78.3	87.0	97.5	92.1	96.5
BERT-pair-QA-B	79.2	87.9	97.1	93.3	97.0
BERT-pair-NLI-B	79.8	87.5	96.6	92.8	96.9

Table 3: Performance on SentiHood dataset. We boldface the score with the best performance across all models. We use the results reported in Saeidi et al. (2016), Ma et al. (2018) and Liu et al. (2018). “-” means not reported.

Fine-tuning pre-trained BERT

- Fine-tuning 절차
- 입력 시퀀스의 fixed-dimensional pooled representation을 얻기 위해 CLS 토큰의 최종 은닉 상태를 입력으로 사용한다.
- 여기서는 벡터를 $C \in R^H$ 로 표시한다.
- 그 다음 parameter matrix가 $W \in R^{K \times H}$ 인 분류 계층을 추가합니다.
K는 category의 수입니다.
- 마지막으로, 각 category P의 확률은 $\text{softmax}(CW^T)$ 에 의해 계산됩니다.

- **BERT-single for (T)ABSA**

- 단일 문장 분류 작업을 위한 BERT.
- target 범주의 수가 n_t 이고 aspect 범주가 n_a 라고 가정합니다.
- TABSA를 $n_t \cdot n_a$ 개의 target-aspect-related sentiment 분류 문제의 조합으로 간주하고, 먼저 각 sentiment 분류 문제를 분류한 다음 얻은 결과를 요약한다.
- ABSA의 경우 pre-trained BERT 모델을 fine-tuning하여 모든 aspect에 대한 n_a 분류기를 학습시킨 다음 결과를 요약한다.

- **BERT-pair for (T)ABSA**

- 문장 쌍 분류 작업에 대한 BERT.
- 앞에서 구성한 보조 문장을 기반으로, (T)ABSA를 해결하기 위해 문장 쌍 분류 접근법을 사용한다.
- 보조문장을 구성하는 네 가지 방법에 따라 각 모델의 이름을

BERT-pair-QA-M, BERT-pair-NLI-M, BERT-pair-QA-B 및 BERT-pair-NLI-B로 지정한다.

Datasets

- 5,215개의 문장으로 구성된 SentiHood_(Saeidi et al., 2016) dataset로 평가.
이 중 3,862개는 단일 target을 포함하고 나머지는 다중 target을 포함한다.
각 문장은 sentiment 극성 y 를 가진 target-aspect 쌍 $\{t, a\}$ 의 목록을 포함한다.
- 문장과 문장의 target t 가 주어지면,
 - (1) target t 에 대한 aspect a 의 언급을 감지하고,
 - (2) 탐지된 target-aspect 쌍에 대한 sentiment 극성 y 를 결정한다.
- 또한 aspect 기반 정서 분석을 위해 SemEval-2014 Task 4_(Pontiki et al., 2014) dataset로 평가한다.
- SentiHood와 유일한 차이점은 target-aspect 쌍 $\{t,a\}$ 이(가) 단지 aspect a 가 된다는 것이다.

```

94     {
95         "opinions": [],
96         "id": 530,
97         "text": ":// When we stayed in London the first time, we stayed in LOCATION2 near LOCATION1"
98     },
99     {
100         "opinions": [],
101         "id": 226,
102         "text": "@b: LOCATION1 is in Hertfordshire not London, and it is even worse than Bexley"
103     },
104     {
105         "opinions": [
106             {
107                 "sentiment": "Negative",
108                 "aspect": "safety",
109                 "target_entity": "LOCATION1"
110             }
111         ],
112         "id": 292,
113         "text": "A Brazilian man was shot there 3 years ago in LOCATION1 station by a policeman i think he was a suspected terrorist of the 7/7 bombings on London transport"
114     },
115     {
116         "opinions": [
117             {
118                 "sentiment": "Negative",
119                 "aspect": "safety",
120                 "target_entity": "LOCATION1"
121             },
122             {
123                 "sentiment": "Positive",
124                 "aspect": "live",
125                 "target_entity": "LOCATION1"
126             },
127             {
128                 "sentiment": "Positive",
129                 "aspect": "general",
130                 "target_entity": "LOCATION1"
131             }
132         ],
133         "id": 78,
134         "text": "A friend of mine lived in LOCATION1 and she liked it, though other people have told me it's a bit rough"
135     },

```

평가: TABSA

- SentiHood dataset을 평가하는 동안, 가장 자주 볼 수 있는 네 가지 속성(general, price, transit-location, safety)만 고려한다.
- BERT-pair는 aspect 감지 및 정서 분석에서 다른 모델을 상당한 차이로 능가하여 DmuEntnet에 비해 9.4 Macro-average-F1 및 2.6 Accuracy 향상을 얻었다.
- 전반적으로, 4개의 BERT-pair 모델의 성능은 거의 비슷하다.
- BERT-pair-NLI 모델은 aspect 감지에서 상대적으로 성능이 우수한 반면, BERT-pair-QA 모델은 sentiment 분류에서 성능이 우수하다.
- 또한 BERT-pair-QA-B 및 BERT-pair-NLI-B 모델은 다른 모델보다 sentiment 분류에서 더 나은 AUC 값을 달성했다.

Model	Aspect			Sentiment	
	<i>Acc.</i>	F_1	AUC	<i>Acc.</i>	AUC
LR (Saeidi et al., 2016)	-	39.3	92.4	87.5	90.5
LSTM-Final (Saeidi et al., 2016)	-	68.9	89.8	82.0	85.4
LSTM-Loc (Saeidi et al., 2016)	-	69.3	89.7	81.9	83.9
LSTM+TA+SA (Ma et al., 2018)	66.4	76.7	-	86.8	-
SenticLSTM (Ma et al., 2018)	67.4	78.2	-	89.3	-
Dmu-Entnet (Liu et al., 2018)	73.5	78.5	94.4	91.0	94.8
BERT-single	73.7	81.0	96.4	85.5	84.2
BERT-pair-QA-M	79.4	86.4	97.0	93.6	96.4
BERT-pair-NLI-M	78.3	87.0	97.5	92.1	96.5
BERT-pair-QA-B	79.2	87.9	97.1	93.3	97.0
BERT-pair-NLI-B	79.8	87.5	96.6	92.8	96.9

Table 3: Performance on SentiHood dataset. We boldface the score with the best performance across all models. We use the results reported in Saeidi et al. (2016), Ma et al. (2018) and Liu et al. (2018). “-” means not reported.

평가: ABSA

Models	P	R	F1
XRCE	83.23	81.37	82.29
NRC-Canada	91.04	86.24	88.58
BERT-single	92.78	89.07	90.89
BERT-pair-QA-M	92.87	90.24	91.54
BERT-pair-NLI-M	93.15	90.24	91.67
BERT-pair-QA-B	93.04	89.95	91.47
BERT-pair-NLI-B	93.57	90.83	92.18

Table 4: Test set results for Semeval-2014 task 4 Subtask 3: Aspect Category Detection. We use the results reported in XRCE (Brun et al., 2014) and NRC-Canada (Kiritchenko et al., 2014).

Models	4-way	3-way	Binary
XRCE	78.1	-	-
NRC-Canada	82.9	-	-
LSTM	-	82.0	88.3
ATAE-LSTM	-	84.0	89.9
BERT-single	83.7	86.9	93.3
BERT-pair-QA-M	85.2	89.3	95.4
BERT-pair-NLI-M	85.1	88.7	94.4
BERT-pair-QA-B	85.9	89.9	95.6
BERT-pair-NLI-B	84.6	88.7	95.1

Table 5: Test set accuracy (%) for Semeval-2014 task 4 Subtask 4: Aspect Category Polarity. We use the results reported in XRCE (Brun et al., 2014), NRC-Canada (Kiritchenko et al., 2014) and ATAE-LSTM (Wang et al., 2016). “-” means not reported.

Discussion

- 왜 BERT-pair 모델의 실험 결과가 더 좋은가?

-> target 및 aspect 정보를 보조 문장으로 변환하는데, 이는 말뭉치를 확장하는 것과 같다.

- 원본 데이터 집합의 문장 s_i 는 문장 쌍 분류 작업에서 $(s_i, t_1, a_1), \dots, (s_i, t_1, a_{na}), \dots, (s_i, t_{nt}, a_{na})$ 로 확장된다.
- TABSA는 SA보다 target 및 aspect 정보가 더 복잡하다.
- TABSA에서 pre-trained BERT를 directly fine-tuning하는 것은 성능 향상을 달성하지 못한다.

그러나 target과 aspect을 분리하여 보조 문장을 구성하고 TABSA를 문장 쌍 분류 작업으로 변환하면 QA 및 NLI와 유사하므로 pre-trained BERT 모델의 장점을 충분히 활용할 수 있다.

- 논문에서 이 접근 방식은 TABSA에 제한되지 않으며, 이 구성 방법은 다른 유사한 작업에 사용될 수 있다고 함.
- ABSA의 경우 동일한 접근 방식을 사용하여 aspect만 있는 보조 문장을 구성할 수 있다.
- BERT-pair 모델에서 BERT-pair-QA-B와 BERT-pair-NLI-B는 레이블 정보의 모델링 때문에 정서 분류에서 더 나은 AUC 값을 달성한다.

Conclusion

- 본 논문에서는 (T)ABSA를 단일 문장 분류 작업에서 문장 쌍 분류 작업으로 변환하기 위한 보조 문장을 구성하였다.
- 문장 쌍 분류 작업에서 pre-trained BERT 모델을 fine tuning하는 것으로 SOTA를 얻었다.
- BERT fine-tuning을 기반으로 단일 문장 분류와 문장 쌍 분류의 실험 결과를 비교하고, 문장 쌍 분류의 장점을 분석하고, 변환 방법의 유효성을 검증했다.
- 이 연구에서는 target에 대한 정보가 직접 주어지는데 이를 응용하기 위해서는 모델이 스스로 target을 찾을 수 있도록 해야 할 것이다.
- 문장-보조문장 순으로 입력이 들어가는데 이 순서가 바뀌면 어떻게 될까?
- 더 많은 aspect를 예측해야 한다면 성능이 어떻게 변할까?

{

```
"opinions": [  
  {  
    "sentiment": "Positive",  
    "aspect": "shopping",  
    "target_entity": "LOCATION1"  
  }  
],  
"id": 302,  
"text": " LOCATION1 is just a normal area that happens to have an alternative market"  
},
```

```
In [9]: 1 train_dev['train'][229]
```

```
Out[9]: {'comments': 'ㅋㅋㅋ 생각이라는 걸 하고 산다면...ㅋㅋㅋㅋ',  
         'contain_gender_bias': False,  
         'bias': 'none',  
         'hate': 'offensive',  
         'news_title': '"현아♥이던, 거침없는 키스...당당한 사랑꾼 껌딱지 커플"'}
```

ratsgo's nlpbook

질의/응답 (Question Answering)

지문(context)과 질문(question)이 주어졌을 때 질문에 적절한 답을 지문에서 찾습니다.

context 지문을 입력하세요

question 질문을 입력하세요

답 찾기

context

일론 머스크는 2022년 기준 세계에서 가장 부유한 인물이며 인류 역사상 최초로 개인의 자산이 3,000억 달러를 돌파한 사람이다. 항공우주, 전기차, 인공지능, 인터넷 위성군, 초고속열차 등 다양한 첨단과학분야 기업을 경영하며 전세계 기술 시장을 선도하고 있다. 또한 그의 영향력은 특정 첨단기술 업계에만 국한되는 것이 아니라 단순한 트윗 한 줄 만으로 주식, 암호화폐와 같은 투자분야의 상품들을 수십조 원 씩 상승 또는 하락시키기도 하며 상당한 양의 기부금을 사용하여 미국의 정치에 관여하고[34] 심지어 전쟁의 판도[35] 에도 영향을 끼치는 웅장함을 지니고 있다. 그러나 엄청난게 강한 파워에 걸맞지 않은 진중하지 못한 태도와 종잡을 수 없는 행보로 인해, 신뢰할 수 없는 소시오패스라는 악평과 천재적인 능력을 가진 기업가라는 칭송을 동시에 받는 인물이다

question

누가 세계에서 가장 돈이 많지?

answer

[CLS] 누가 세계에서 가장 돈이 많지? [SEP] 일론 머스크

이 웹 데모가 어떻게 동작하는지 자세히 확인하고 싶으신 분은 [ratsgo's nlpbook](#)을 참고하세요.

Copyright © 2020 Gichang LEE. Distributed by an [CC BY-NC-SA 3.0 license](#).