

Aspect-Based Sentiment Analysis

Deep Context-and Relation-Aware Learning for Aspect-based Sentiment Analysis

- Oh, S.*, Lee, D.*, Whang, T., Park, I., Seo, G., Kim, E., & Kim, H. (2021). Deep Context-and Relation-Aware Learning for Aspect-based Sentiment Analysis. In Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics.

01. 개요

- Aspect-Based Sentiment Analysis (ABSA)는 문장 내 aspect 용어들과 관련된 감정을 찾아내는 task입니다. ABSA는 다음과 같이 크게 3가지로 나눌 수 있습니다; 1) Aspect Term Extraction (ATE), 2) Opinion Term Extraction (OTE), 3) Aspect-based Sentiment Classification (ASC). 예를 들어, “Food is good, but service is dreadful”의 문장에서 ATE는 “food”, “service”와 같은 aspect 용어들을 추출하고, OTE는 각 aspect에 대한 의견인 “good”, “dreadful”을 추출, ASC는 각 aspect에 대한 감정 분류를 “food (good)”, “service (dreadful)”을 진행합니다.
- 최근 Pre-trained LM에 힘입어 ABSA task를 joint training하여 좋은 성능을 보이고 있지만, ABSA에는 여전히 한계점이 존재합니다. 먼저, opinion의 말 그대로의 의미가 아닌 문장 전체를 반영하는 의미를 이해하는 것이 필요합니다. 아래 표의 E1의 경우 “better”가 “Japanese food”의 긍정적 의미로 사용되는 것처럼 보일 수도 있지만, 실제로는 다른 곳에서 더 좋은 음식을 먹었다는 의미이므로, 부정 (negative)으로 분류되어야 합니다. 또한, 한 문장 내에 여러 개의 aspect가 존재할 때 aspect-opinion pair들을 찾는 것과 각각의 감성 분석을 하는 것은 더욱 어렵습니다.

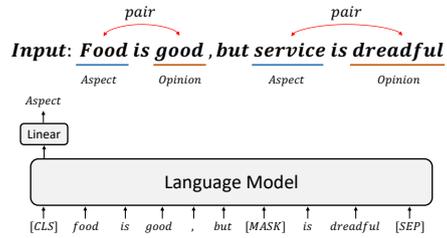
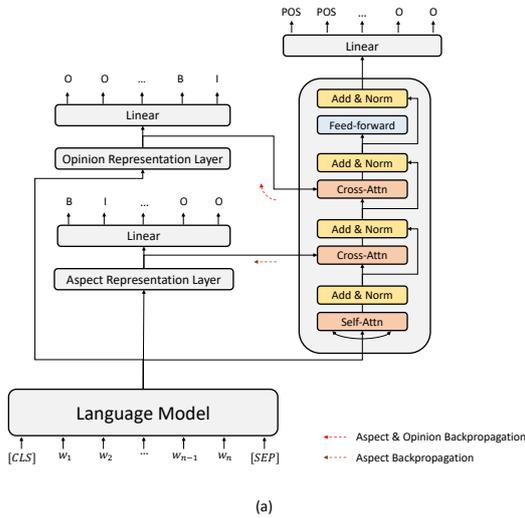
	Examples (Ground Truth)	Model	Aspect (Polarity)	Opinion
E1	I've had <i>better Japanese food</i> (neg) at a mall food court.	RACL	Japanese food (pos)	better
		DCRAN	Japanese food (neg)	better
E2	The <i>sushi</i> (neg) is cut in blocks <i>bigger</i> than my cell phone.	RACL	sushi (neu)	bigger
		DCRAN	sushi (neg)	bigger
E3	While the <i>smoothies</i> (neg) are a little <i>bigger</i> for me, the <i>fresh juices</i> (pos) are the <i>best</i> I have ever had !	RACL	smoothies (pos) fresh juices (pos)	bigger fresh best
		DCRAN	smoothies (neg) fresh juices (pos)	bigger best

02. 연구 내용

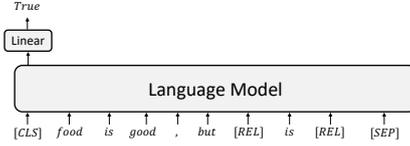
- 본 연구에서는 Deep Context Relation-Aware Network (DCRAN) 모델을 제안하였습니다. 먼저, context를 각 subtask의 관점에서 종합적으로 이해하게 도와주는 aspect와 opinion propagation decoder를 설계하였습니다. 또한, 여러 개의 aspect와 opinion을 포착하는 데에 효과적인 두 개의 self-supervised 방법을 제안함으로써 성능 향상을 이끌어 냈습니다. 실험 결과, 총 세 개의 benchmark 데이터셋 (REST, LAB)에서 state-of-the-art의 성능을 얻었으며, ablation 실험을 통해 제안하는 모델에 대해 검증하였습니다. 모델 전체 구조는 아래 그림과 같습니다.
- Self-supervised 방법은 모델이 aspect-opinion pair를 잘 포착할 수 있게 설계하였습니다. 1) Type-Specific Masked Term Discrimination (TSMTD): 문장 내 임의의 Aspect, Opinion, Outside를 [MASK] token으로 치환하여, [CLS] token이 representation을 통해 [MASK] token이 어떤 것인지 분류하는 task입니다. 2) Pairwise Relation Discrimination (PRD): 문장 내 aspect-opinion pair를 [REL] token으로 치환하고 치환된 [REL] 토큰들이 적절한 pair인지 [CLS] token representation으로 분류하는 task입니다.

03. 실험 및 결과

- 실험은 아래 표와 같이 ABSA에서 주요 benchmark 데이터셋으로 활용되는 REST-14, REST-15, LAB-14에 대해서 DCRAN의 성능을 기존 baseline과 비교 평가를 진행하였습니다.



Type – Specific Masked Term Discrimination(TSMTD)



Pairwise Relation Discrimination(PRD)

- Pre-trained LM의 경우 BERT와 ELECTRA 두 가지 언어 모델에 대해서 실험을 진행하였고, 기존 baseline 모델들과 공정한 비교를 위해 각 언어모델의 base와 large 모델에 대해서 평가를 진행하였습니다.
- 실험 결과, DCRAN-BERT_{large}와 DCRAN-ELECTRA_{large}가 모든 benchmark 데이터셋에서 주목할 만한 성능 향상을 보여주었습니다. 각 subtask인 ATE, OTE, ASC는 baseline별로 편차가 존재하였으나, 이를 포괄하는 ABSA-F1의 경우 본 연구에서 제안한 DCRAN 모델이 뛰어난 성능 향상을 보여주었습니다.

		LAP14				REST14				REST15			
		ATE-F1	OTE-F1	ASC-F1	ABSA-F1	ATE-F1	OTE-F1	ASC-F1	ABSA-F1	ATE-F1	OTE-F1	ASC-F1	ABSA-F1
MNN (Wang et al., 2018)	GloVe	76.94	77.77	65.98	53.80	83.05	84.55	68.45	63.87	70.24	69.38	57.90	56.57
E2E-TBSA (Li et al., 2019)	GloVe	77.34	76.62	68.24	55.88	83.92	84.97	68.38	66.60	69.40	71.43	58.81	57.38
DOER (Luo et al., 2019)	GloVe	80.21	-	60.18	56.71	84.63	-	64.50	68.55	67.47	-	36.76	50.31
IMN ^{-d} (He et al., 2019)	GloVe	78.46	78.14	69.62	57.66	84.01	85.64	71.90	68.32	69.80	72.11	60.65	57.91
RACL (Chen and Qian, 2020)	GloVe	81.99	79.76	71.09	60.63	85.37	85.32	74.46	70.67	72.82	78.06	68.69	60.31
WHW (Peng et al., 2020)	GloVe	-	74.84	-	62.34	-	82.45	-	71.95	-	78.02	-	65.79
IKTN (Liang et al., 2020)	BERT _{base}	80.89	78.90	73.42	62.34	86.13	<u>86.62</u>	74.35	71.75	71.63	<u>76.79</u>	69.85	62.33
SPAN (Hu et al., 2019)	BERT _{large}	82.34	-	62.50	61.25	<u>86.71</u>	-	71.75	73.68	<u>74.63</u>	-	50.28	62.29
IMN ^{-d} (He et al., 2019)	BERT _{large}	77.55	81.00	75.56	61.73	84.06	85.10	75.67	70.72	69.90	73.29	70.10	60.22
Dual-MRC (Mao et al., 2021)	BERT _{large}	<u>82.51</u>	-	<u>75.97</u>	<u>65.94</u>	86.60	-	82.04	<u>75.95</u>	75.08	-	73.59	65.08
RACL (Chen and Qian, 2020)	BERT _{large}	81.79	<u>79.72</u>	73.91	63.40	86.38	87.18	<u>81.61</u>	75.42	73.99	76.00	<u>74.91</u>	<u>66.05</u>
DCRAN (Ours)	BERT _{base}	81.76	78.84	77.02	65.18	88.21	86.36	78.67	75.77	71.61	75.86	73.30	63.19
	BERT _{large}	83.40	<u>79.72</u>	78.75	68.07	88.73	86.07	80.64	77.28	74.45	78.45	76.30	67.92
	ELECTRA _{base}	85.69	80.19	79.36	70.22	89.64	<u>87.30</u>	84.12	80.00	77.41	78.80	78.55	71.67
	ELECTRA _{large}	85.61	79.77	80.78	71.47	89.67	87.59	84.22	80.32	79.68	79.90	77.99	73.67

- 실제 제안하는 self-supervised 방법이 aspect-opinion pair를 잘 포착하는지에 대해 ablation을 진행하였습니다. 실험 결과, 본 연구에서 제안한 방법이 있을 때 sentence-level accuracy 성능이 크게 향상되는 것을 확인하였습니다.

		REST14		REST15	
		ABSA-F1	Sent-level Acc.	ABSA-F1	Sent-level Acc.
Single-Aspect	DCRAN.ELECTRA _{base}	78.62	74.48	66.23	67.69
	w/o TSMTD & PRD	78.42	73.79	64.21	66.67
	w/o TSMTD & PRD & AP & OP	77.45	73.10	62.50	64.29
Multiple-Aspect	DCRAN.ELECTRA _{base}	81.19	64.24	68.20	52.34
	w/o TSMTD & PRD	80.22	61.70	65.16	48.60
	w/o TSMTD & PRD & AP & OP	79.88	61.39	64.84	46.73