Structural Bias For Aspect Sentiment Triplet Extraction

Chen Zhang, Lei Ren, Fang Ma, Jingang Wang, Wei Wu, Dawei Song

北京理工大學
BEIJING INSTITUTE OF TECHNOLOGY

美团 美团

Outline

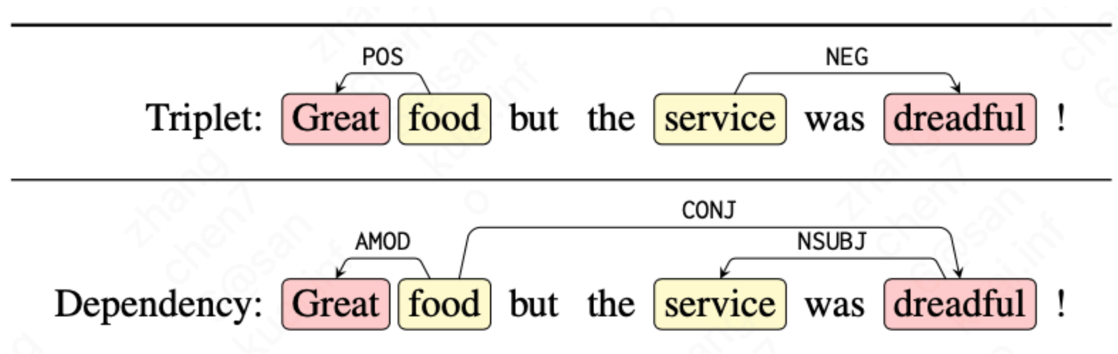
- Motivation
- Method
- Data
- Results
- Conclusion

Motivation

Aspect Sentiment Triplet Extraction

- Given a sentence (comment), an ASTE model is required to output triplets of (aspect, opinion, sentiment).

- It is commonly recognized that the triplets are highly correlated with syntactic dependency structures.



Motivation

ASTE with Syntactic Dependency Structures

- Syntactics dependency structures can be incorporated into non-LM-based ASTE models and performance improvements can be yielded.
- However, we are curious about whether this sort of structural bias is necessary for LMs, especially regarding LMs are implicitly structure learners. We study this by explicitly injecting structural bias into LMs.
- By the way, we highlight two design inefficiencies:
  - Parameter-inefficiency and latency-inefficiency.

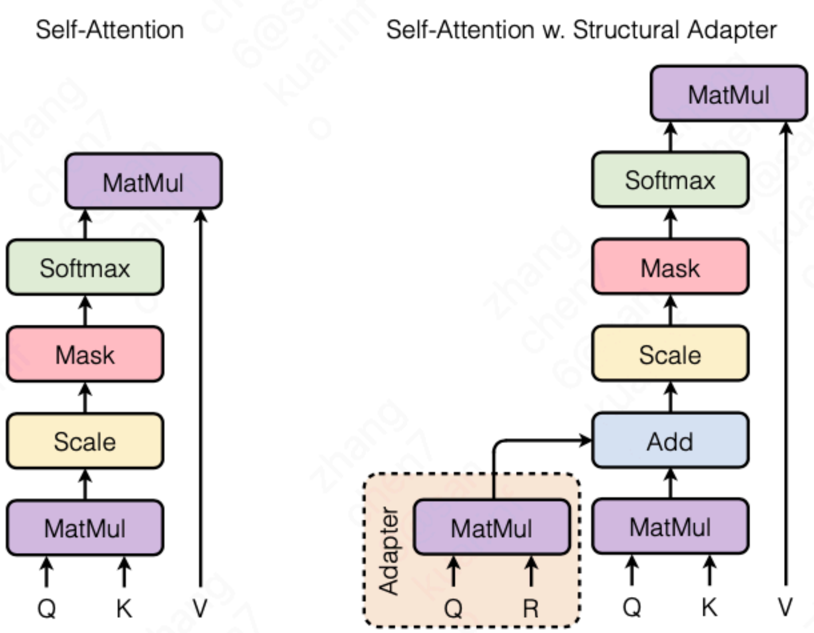
Motivation

ASTE with Structural Adapter

- Parameter-inefficiency
  - Adapter-like fashion.
- Latency-inefficiency
  - Relative position structure as an alternative inspired previous studies in aspect-based sentiment analysis.

Method
Structural Adapter

- Additive adapter.

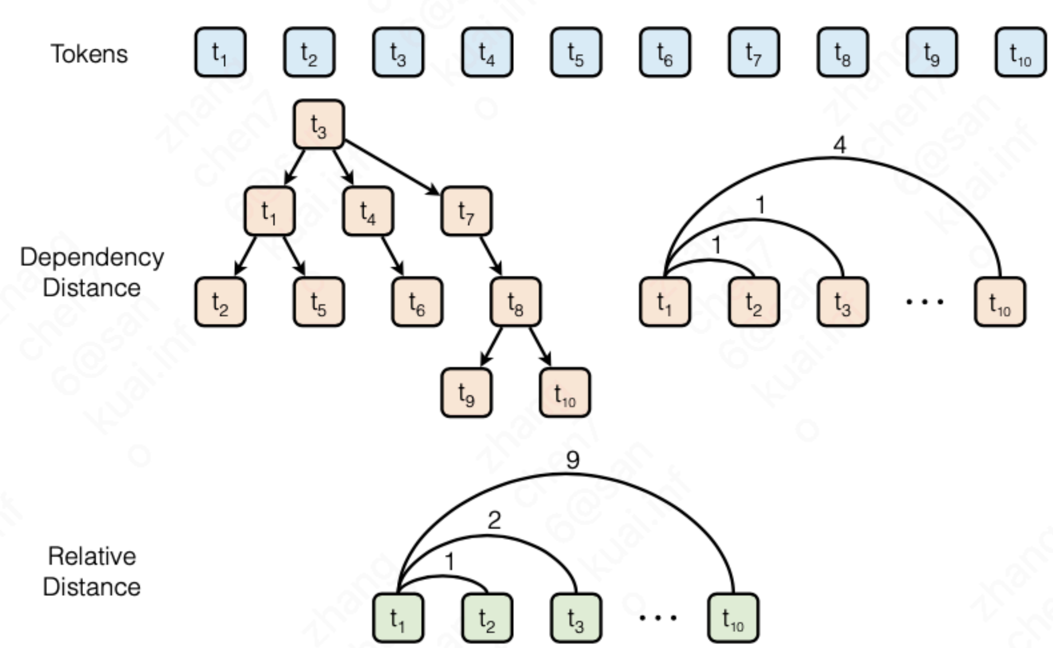


$$e_{ij} = \frac{\mathbf{x}_i \mathbf{W}_Q (\mathbf{x}_j \mathbf{W}_K + \mathbb{r}_{ij})^\top}{\sqrt{d}}$$
$$= \underbrace{\frac{\mathbf{x}_i \mathbf{W}_Q (\mathbf{x}_j \mathbf{W}_K)^\top}{\sqrt{d}}}_{\text{raw attention map}} + \underbrace{\frac{\mathbf{x}_i \mathbf{W}_Q \mathbb{r}_{ij}^\top}{\sqrt{d}}}_{\text{structured attention map}}$$

Method

Structure Derivation

- Syntactic dependency v.s. relative position.



Method

Efficiency

- Parameter and latency efficiencies are improved.

Model	#Params+
MuG BERT	0.00 M
w/ StructLyr-Dep	14.17 M
w/ StructLyr-Rel	14.17 M
w/ StructApt-Dep	0.01 M
w/ StructApt-Rel	0.01 M

While dependency distance derivation costs around 4 micro-seconds per token (250 tokens/ms in other words), relative distance derivation only spends 3e-3 micro-seconds per token (333,000 tokens/ms in other words). That is, the relative distance derivation enjoys a 1,000× speed-up compared with the dependency distance derivation. Hence, the *latency efficiency* of relative distance derivation is numerically verified.

Data

Lasted: A Large-scale ASTE Dataset from dianping.com

- Previous datasets v.s. Lasted.

Dataset		#S	#T	#T/S	#Tk/S
SemEval R14	train	1266	2336	1.85	17.31
	dev	310	577	1.86	15.81
	test	492	994	2.02	16.34
Lasted	train	19485	38050	1.95	34.94
	dev	2783	5334	1.92	34.88
	test	5567	10820	1.94	35.04

Results

Benchmarking Evaluation

- Results on SemEval

Model	L14			R14			R15			R16		
	P	R	F_1	P	R	F_1	P	R	F_1	P	R	F_1
KWHW BiLSTM*	37.38	50.38	42.87	43.24	63.66	51.46	48.07	57.51	52.32	46.96	64.24	54.21
JET^o BiLSTM*	53.03	33.89	41.35	61.50	55.13	58.14	64.37	44.33	52.50	70.94	57.00	**63.21**
MTL BiLSTM‡	51.00	40.07	44.81	63.87	54.76	58.90	57.50	42.56	48.73	59.03	54.84	56.73
GTS BiLSTM‡	60.32	38.98	**47.25**	71.08	56.38	**62.85**	66.60	46.91	**55.02**	68.75	56.02	61.71
JET^o BERT*	55.39	47.33	51.04	70.56	55.94	62.40	64.45	51.96	57.53	70.42	58.37	63.83
GTS BERT‡	57.09	50.33	53.48	69.49	67.75	68.59	61.59	58.21	59.81	65.75	68.32	66.99
w/ StructApt-Rel	57.89	51.57	54.47	68.94	68.26	68.60	62.17	58.63	60.28	66.17	69.79	67.91
Span BERT‡	62.57	56.02	59.08	71.77	70.42	71.06	62.06	63.26	62.63	68.57	71.12	69.79
w/ StructApt-Rel	64.72	56.80	**60.47**	72.53	71.75	**72.13**	62.80	63.79	**63.17**	68.94	70.74	**69.80**
MuG BERT	58.30	52.21	55.06	68.40	67.64	68.00	60.65	54.12	57.10	66.26	67.39	66.74
w/ StructApt-Dep	59.39	52.95	55.95	67.69	68.90	68.27	60.74	55.77	58.11	64.73	68.33	66.45
w/ StructApt-Rel	59.54	52.56	55.75	68.92	68.12	68.50	59.83	56.78	58.17	65.31	68.83	67.01
UGF BART†	61.41	56.19	58.69	65.52	64.99	65.25	59.14	59.38	59.26	66.60	68.68	67.62
GAS T5†	–	–	**60.78**	–	–	**72.16**	–	–	62.10	–	–	70.10
MuG RoBERTa	64.18	57.03	60.33	70.47	71.88	71.16	63.78	61.88	62.79	68.61	72.20	70.34
w/ StructApt-Dep	64.18	56.41	60.03	71.62	71.92	71.72	63.96	61.67	62.70	68.85	71.81	70.28
w/ StructApt-Rel	64.12	57.16	60.53	73.26	71.93	71.17	62.86	63.82	**63.12**	69.15	74.12	**70.44**

Results
Large-scale Evaluation

- Results on Lasted

Model	Lasted P	Lasted R	Lasted F$_1$
GTS BERT-base	43.81	46.11	44.92
w/ StructApt-Rel	45.38	46.22	45.79
MuG BERT-base	47.20	45.28	46.22
w/ StructApt-Rel	49.64	45.02	47.22
MuG RoBERTa-base	48.10	44.98	46.49
w/ StructApt-Rel	50.40	44.77	47.42
MuG RoBERTa-large	49.49	46.85	48.13
w/ StructApt-Rel	48.33	47.91	48.13

Conclusion

- Structural bias is still a necessity even with LMs.
- Is a structure-biased pretrained language model beneficial?
- arXiv https://arxiv.org/abs/2209.00820
- Code & Data https://github.com/GeneZC/StructBias
- Contact czhang@bit.edu.cn
- Many thanks.