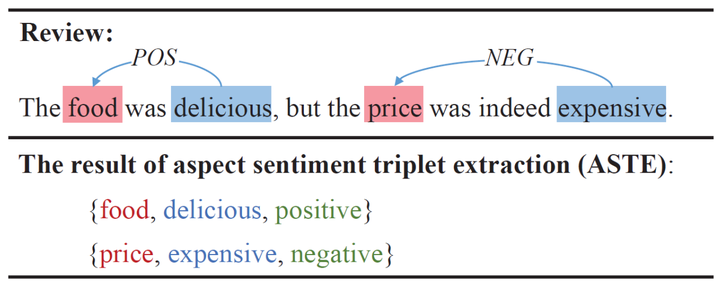
Bidirectional Machine Reading Comprehension for Aspect Sentiment Triplet Extraction

Shaowei Chen, Yu Wang, Jie Liu*, Yuelin Wang
College of Artificial Intelligence, Nankai University, Tianjin, China
{shaoweichen, yuwang17, 1711496}@mail.nankai.edu.cn, jliu@nankai.edu.cn

Aspect sentiment triplet extraction task (ASTE)

Aim to identify triplets from review sentences.

- <aspect, opinion expression, sentiment>

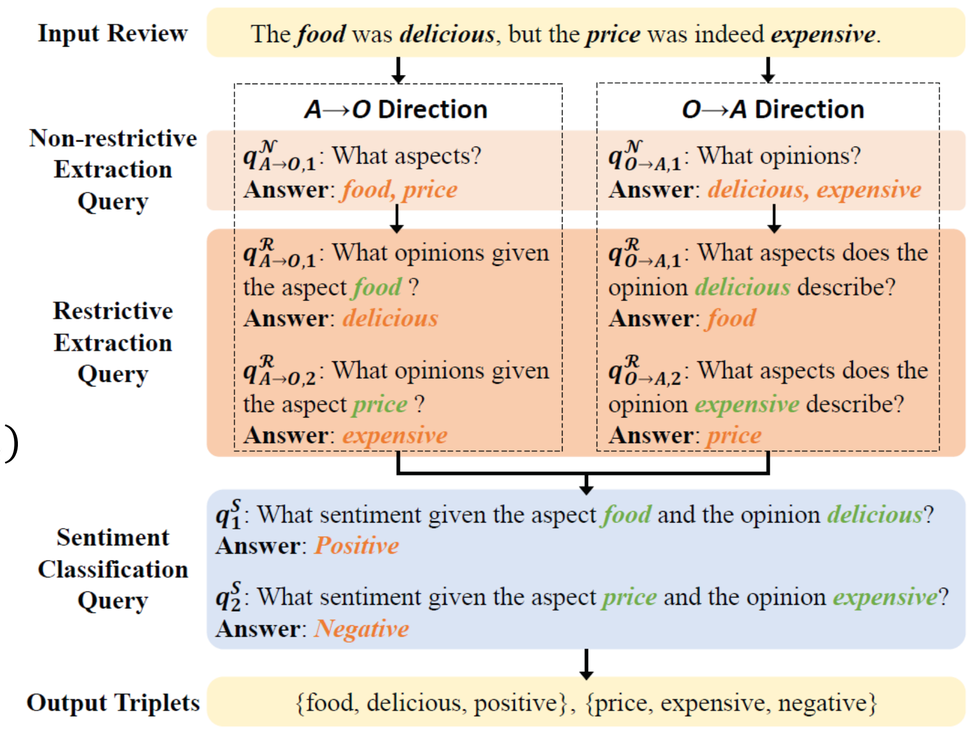


Three challenges:

- How to adequately learn the **association** between aspect term extraction (ATE) and opinion term extraction (OTE) and make them **mutually beneficial.**
- How to flexibly and exactly detect the **complicated relations** between aspects and opinion expressions.
- How to properly **introduce the detected relations** into sentiment classification task.

Model

- Transform ASTE task into a **multi-turn machine reading comprehension** (MTMRC) task.
- Propose a **bidirectional MRC** (BMRC) framework.
- **Three types of queries**:
  - Non-restrictive extraction query $Q^{\mathcal{N}} = \{q_i^{\mathcal{N}}\}_{i=1}^{|Q^{\mathcal{N}}|}$, Restrictive extraction query $Q^{\mathcal{R}} = \{q_i^{\mathcal{R}}\}_{i=1}^{|Q^{\mathcal{R}}|}$, Sentiment classification query $Q^{\mathcal{S}} = \{q_i^{\mathcal{S}}\}_{i=1}^{|Q^{\mathcal{S}}|}$
- **Encoding Layer**: BERT
- **Answer Prediction**:
  - For extraction query:
    $p\left(y_i^{start}|x_i, q\right) = \text{softmax}(\boldsymbol{h}_{|q|+2+i}W_s)$, $p\left(y_i^{end}|x_i, q\right) = \text{softmax}(\boldsymbol{h}_{|q|+2+i}W_e)$
  - For sentiment classification query:
    $$p\left(y^{\mathcal{S}}|X, q\right) = \text{softmax}(\boldsymbol{h}_1 W_c)$$
- **Joint Learning**: minimize the cross-entropy loss.
  $$\mathcal{L}(\theta) = \mathcal{L}_{\mathcal{N}} + \mathcal{L}_{\mathcal{R}} + \mathcal{L}_{\mathcal{S}}$$
- **Inference**: fuse the answers to different queries and obtain triplets.



Experiments

Experimental results (%). 'A-S', 'O', 'P' and 'T' denote aspect term and sentiment co-extraction, opinion term extraction, aspect-opinion pair extraction, and aspect sentiment triplet extraction.

Evaluation	Models	14-Lap				14-Res				15-Res				16-Res			
		A-S	O	P	T	A-S	O	P	T	A-S	O	P	T	A-S	O	P	T
Precision	TSF	63.15	78.22	50.00	40.40	76.60	84.72	47.76	44.18	67.65	78.07	49.22	40.97	71.18	81.09	52.35	46.76
	RINANRTE+	41.20	78.20	34.40	23.10	48.97	81.06	42.32	31.07	46.20	77.40	37.10	29.40	49.40	75.00	35.70	27.10
	Li-unified-R+	66.28	76.62	52.29	42.25	73.15	81.20	44.37	41.44	64.95	79.18	52.75	43.34	66.33	79.84	46.11	38.19
	RACL+R	59.75	77.58	54.22	41.99	75.57	82.28	73.58	62.64	68.35	76.25	67.89	55.45	68.53	82.52	72.77	60.78
	Ours	**72.73**	**84.67**	**74.11**	**65.12**	**77.74**	**87.22**	**76.91**	**71.32**	**72.41**	**82.99**	**71.59**	**63.71**	**73.69**	**85.31**	**76.08**	**67.74**
Recall	TSF	61.55	71.84	58.37	47.24	67.84	80.39	68.10	62.99	64.02	78.07	65.70	54.68	72.30	86.67	70.50	62.97
	RINANRTE+	33.20	62.70	26.20	17.60	47.36	72.05	51.08	37.63	37.40	57.00	33.90	26.90	36.70	42.40	27.00	20.50
	Li-unified-R+	60.71	74.90	52.94	42.78	74.44	83.18	73.67	68.79	64.95	75.88	61.75	50.73	74.55	86.88	64.55	53.47
	RACL+R	**68.90**	**81.22**	**66.94**	51.84	**82.23**	**90.49**	67.87	57.77	**70.72**	**83.96**	63.74	52.53	**78.52**	**91.40**	71.83	60.00
	Ours	62.59	67.18	61.92	**54.41**	75.10	82.90	**75.59**	**70.09**	62.63	73.23	**65.89**	**58.63**	72.69	83.01	**76.99**	**68.56**
F_1-score	TSF	62.34	74.84	53.85	43.50	71.95	82.45	56.10	51.89	65.79	78.02	56.23	46.79	71.73	83.73	60.04	53.62
	RINANRTE+	36.70	69.60	29.70	20.00	48.15	76.29	46.29	34.03	41.30	65.70	35.40	28.00	42.10	54.10	30.70	23.30
	Li-unified-R+	63.38	75.70	52.56	42.47	73.79	82.13	55.34	51.68	64.95	77.44	56.85	46.69	70.20	83.16	53.75	44.51
	RACL+R	64.00	**79.36**	59.90	46.39	**78.76**	**86.19**	70.61	60.11	**69.51**	**79.91**	65.46	53.95	**73.19**	**86.73**	72.29	60.39
	Ours	**67.27**	74.90	**67.45**	**59.27**	76.39	84.99	**76.23**	**70.69**	67.16	77.79	**68.60**	**61.05**	73.18	84.13	**76.52**	**68.13**

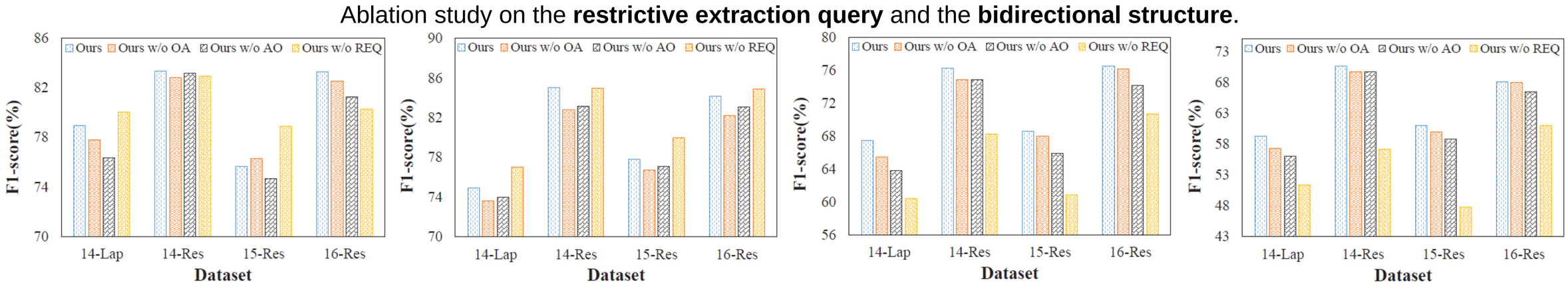
Ablation study on **aspect sentiment triplet extraction** (F1-score, %).

Models	14-Lap	14-Res	15-Res	16-Res
TSF	43.50	51.89	46.79	53.62
Ours w/o BERT	48.15	63.32	53.77	63.16
Ours w/o REQ	51.40	57.20	47.79	61.03
Ours	**59.27**	**70.69**	**61.05**	**68.13**

Ablation study on **sentiment classification** (F1-score, %).

Datasets	A		A-S	
	Ours	Our w/o REQ	Ours	Ours w/o REQ
14-Lap	78.94	**80.06**	**67.27**	61.61
14-Res	**83.31**	82.73	**76.39**	66.26
15-Res	75.67	**79.00**	**67.16**	56.82
16-Res	**83.28**	80.60	**73.18**	68.82

Ablation study on the **restrictive extraction query** and the **bidirectional structure**.



(a) Aspect Term Extraction (b) Opinion Term Extraction (c) Aspect-Opinion Pair Extraction (d) Triplet Extraction