Artificial

Enhanced Multi-Channel Graph Convolutional Network for Aspect Sentiment Triplet Extraction

Hao Chen,* Zepeng Zhai,* Fangxiang Feng, Ruifan Li,† Xiaojie Wang
School of Artificial Intelligence, Beijing University of Posts and Telecommunications, China
{ccchenhao997, zepeng, fxfeng, rfli, xjwang}@bupt.edu.cn

code: https://github.com/CCChenhao997/EMCGCN-ASTE.

—— ACL 2022

2022.10.15 • ChongQing

Leibniz-Institut für Sozialwissenschaften

Reported by Junhao Cao

1.Introduction

2.Method

3.Experiments

Introduction

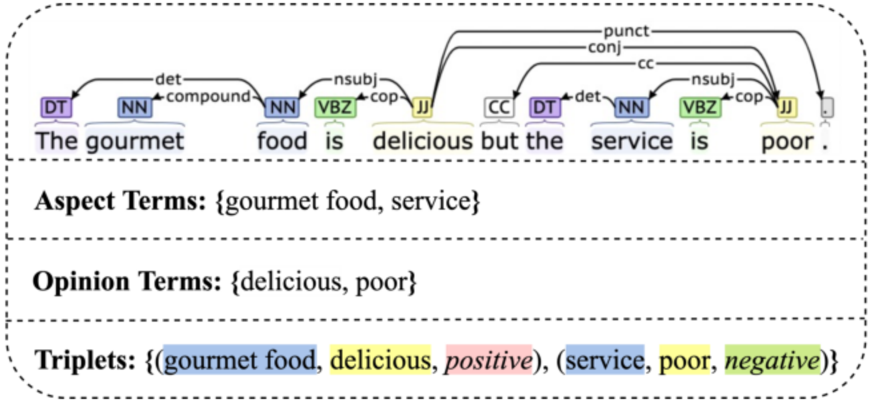


Figure 1: A sentence with its dependency tree is given to illustrate ASTE task. In the triplet set, aspect terms, opinion terms are highlighted in blue and yellow, respectively. The *positive* sentiment polarity is highlighted in red, while the *negative* in green.

Given an input sentence $X = \{w_1, w_2, \cdots, w_n\}$ with n words, the goal of our model is to output a set of triplets $\mathcal{T} = \{(a, o, s)_m\}_{m=1}^{|\mathcal{T}|}$ from the sentence X, where a and o denote aspect term and opinion term,

Introduction

#	Relation	Meaning
1	B-A	beginning of aspect term.
2	I-A	inside of aspect term.
3	A	word pair (w_i, w_j) belongs to the same aspect term.
4	B-O	beginning of opinion term.
5	I-O	inside of opinion term.
6	O	word pair (w_i, w_j) belongs to the same opinion term.
7	POS	w_i and w_j of the word pair (w_i, w_j) respectively belong to an aspect term and an opinion term, and they form aspect-opinion pair with positive/neutral/negative sentiment.
8	NEU	
9	NEG	
10	⊥	no above relations between word pair (w_i, w_j).

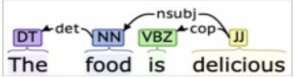
Table 1: The meanings of our defined ten relations. Note that these relations can also be seen as labels.

	The	gourmet	food	is	delicious	but	the	service	is	poor
The	⊥	⊥	⊥	⊥	⊥	⊥	⊥	⊥	⊥	⊥
gourmet	⊥	B-A	A	⊥	POS	⊥	⊥	⊥	⊥	⊥
food	⊥	A	I-A	⊥	POS	⊥	⊥	⊥	⊥	⊥
is	⊥	⊥	⊥	⊥	⊥	⊥	⊥	⊥	⊥	⊥
delicious	⊥	POS	POS	⊥	B-O	⊥	⊥	⊥	⊥	⊥
but	⊥	⊥	⊥	⊥	⊥	⊥	⊥	⊥	⊥	⊥
the	⊥	⊥	⊥	⊥	⊥	⊥	⊥	⊥	⊥	⊥
service	⊥	⊥	⊥	⊥	⊥	⊥	⊥	B-A	⊥	NEG
is	⊥	⊥	⊥	⊥	⊥	⊥	⊥	⊥	⊥	⊥
poor	⊥	⊥	⊥	⊥	⊥	⊥	⊥	NEG	⊥	B-O

Figure 3: Table filling for triplet extraction in a sentence is illustrated. Each cell denotes a word pair with a relation or label. Refer Table 1 for definitions of relations.

1.How to utilize various relations between words to help ASTE task?
2.How to utilize the linguistic features to help ASTE task?

Introduction



Part-of-Speech Combination

	The	food	is	delicious
The	DT-DT	DT-NN	DT-VBZ	DT-JJ
food	NN-DT	NN-NN	NN-VBZ	NN-JJ
is	VBZ-DT	VBZ-NN	VBZ-VBZ	VBZ-JJ
delicious	DT-JJ	NN-JJ	VBZ-JJ	JJ-JJ

Syntactic Dependency Type

	The	food	is	delicious
The	self	det	-	-
food	det	self	-	nsubj
is	-	-	self	cop
delicious	-	nsubj	cop	self

Tree-based Distance

	The	food	is	delicious
The	0	1	3	2
food	1	0	2	1
is	3	2	0	1
delicious	2	1	1	0

Relative Position Distance

	The	food	is	delicious
The	0	1	2	3
food	1	0	1	2
is	2	1	0	1
delicious	3	2	1	0

Figure 4: Four types of features for a sentence.

1.How to utilize various relations between words to help ASTE task?
2.How to utilize the linguistic features to help ASTE task?

Method

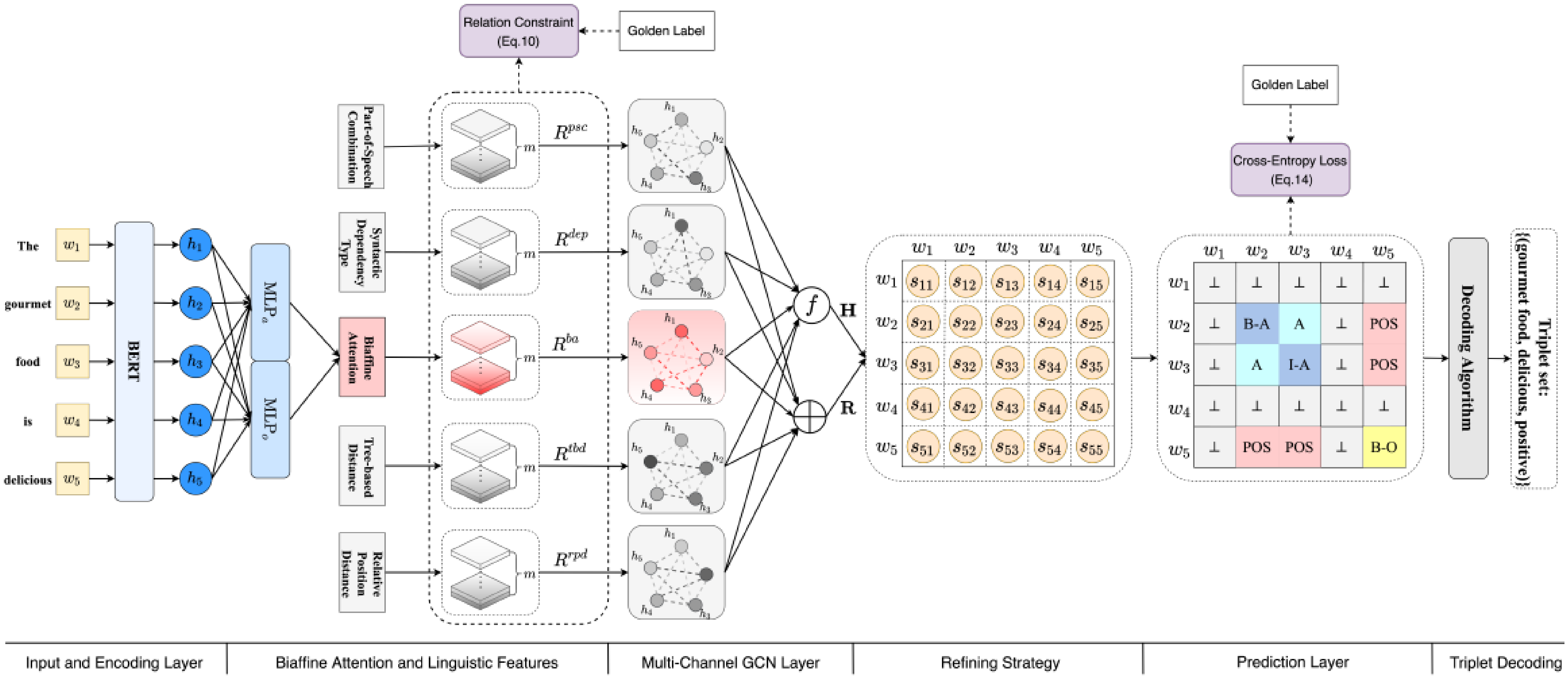


Figure 2: The overall architecture of our end-to-end model EMC-GCN.

Method

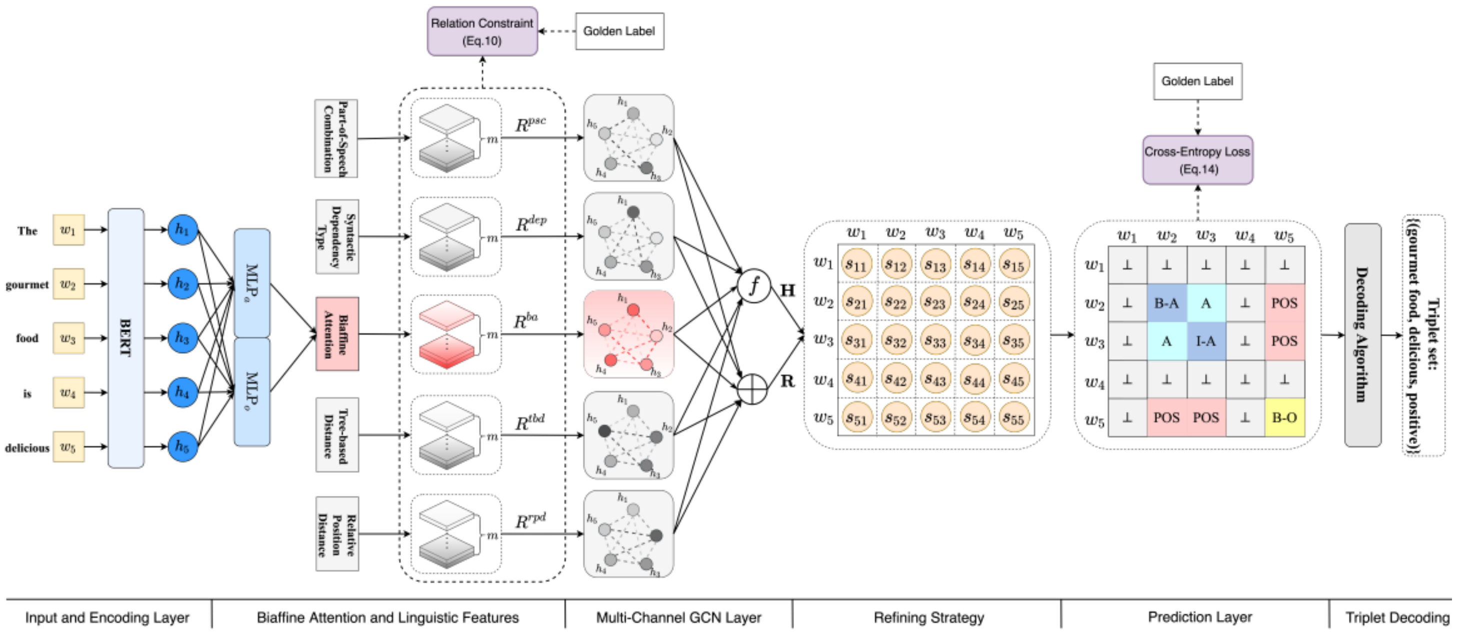


Figure 2: The overall architecture of our end-to-end model EMC-GCN.

Input and Encoding Layer.

$$X = \{w_1, w_2, ..., w_n\}$$

$$H = \{h_1, h_2, ..., h_n\}$$

Biaffine Attention Module

$$h_i^a = \mathrm{MLP}_a(h_i) \tag{1}$$

$$h_j^o = \mathrm{MLP}_o(h_j) \tag{2}$$

$$g_{i,j} = h_i^{a\mathrm{T}} U_1 h_j^o + U_2 \left(h_i^a \oplus h_j^o\right) + b \tag{3}$$

$$r_{i,j,k} = \frac{\exp\left(g_{i,j,k}\right)}{\sum_{l=1}^{m} \exp\left(g_{i,j,l}\right)} \tag{4}$$

$$R = \mathrm{Biaffine}\left(\mathrm{MLP}_a(H), \mathrm{MLP}_o(H)\right) \tag{5}$$

Method

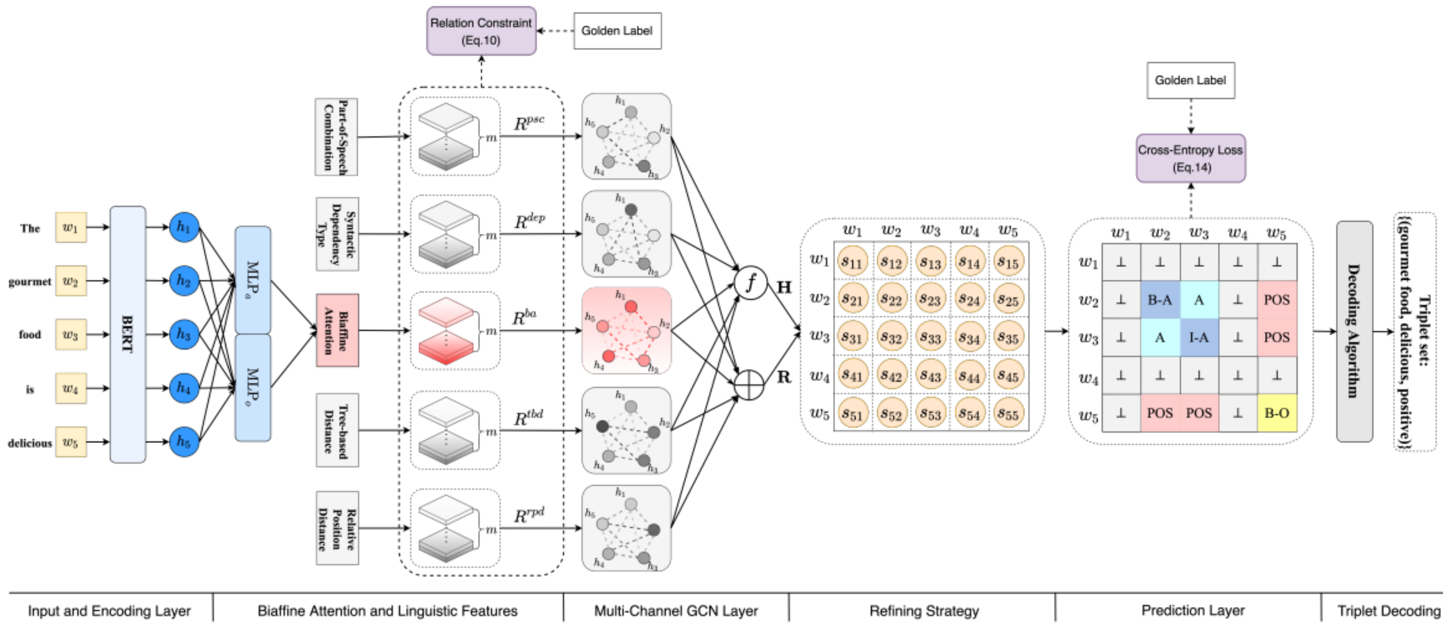


Figure 2: The overall architecture of our end-to-end model EMC-GCN.

Multi-Channel GCN

$$\widetilde{H}_k^{ba} = \sigma\left(R^{ba}_{:,:,k} H W_k + b_k\right) \quad (6)$$

$$\hat{H}^{ba} = f(\widetilde{H}_1^{ba}, \widetilde{H}_2^{ba}, ..., \widetilde{H}_m^{ba}) \quad (7)$$

Linguistic Features

$$\mathbf{H} = f\left(\hat{H}^{ba}, \hat{H}^{psc}, \hat{H}^{dep}, \hat{H}^{tbd}, \hat{H}^{rpd}\right) \quad (8)$$

$$\mathbf{R} = R^{ba} \oplus R^{psc} \oplus R^{dep} \oplus R^{tbd} \oplus R^{rpd} \quad (9)$$

where $\mathbf{H} = \{\mathbf{h}_1, \mathbf{h}_2, ..., \mathbf{h}_n\}$ and $\mathbf{R} = \{\mathbf{r}_{1,1}, \mathbf{r}_{1,2}, ..., \mathbf{r}_{n,n}\}$ denote node representations and edge representations of word pairs.

Relation Constraint

$$\mathcal{L}_{ba} = -\sum_{i}^{n}\sum_{j}^{n}\sum_{c\in\mathcal{C}} \mathbb{I}(y_{ij} = c)\log(r_{i,j|c}) \quad (10)$$

Method

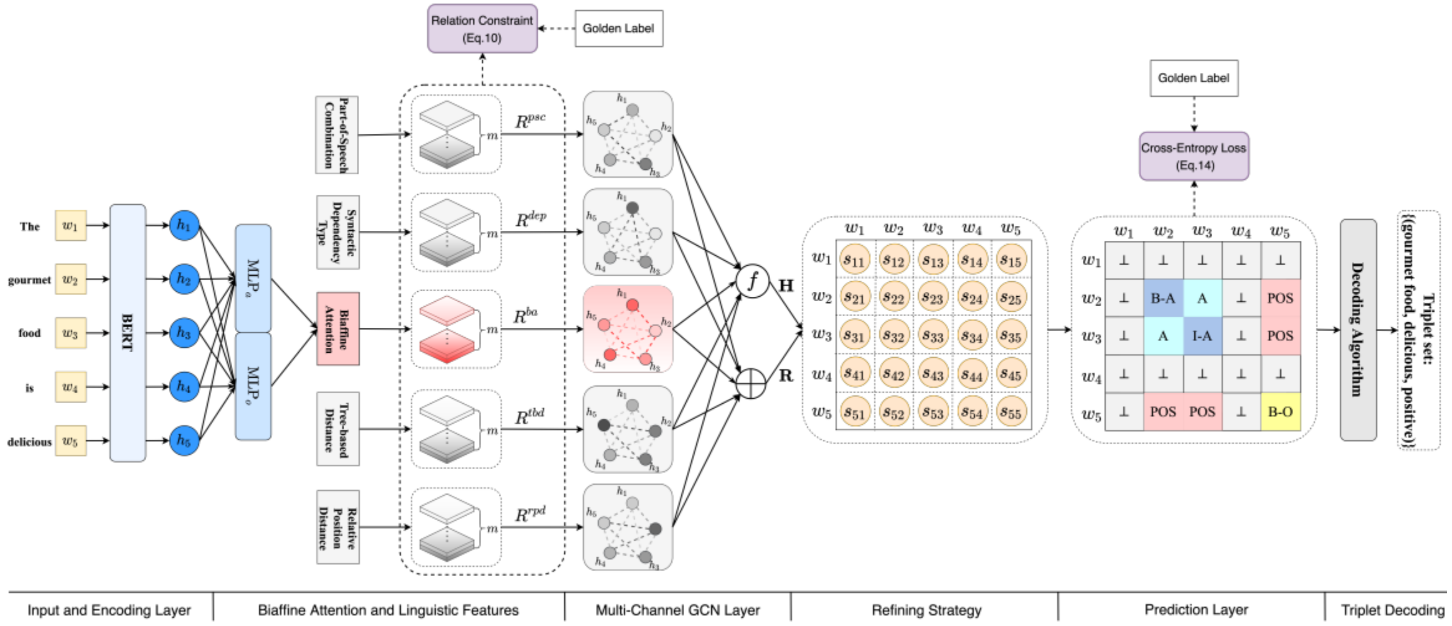


Figure 2: The overall architecture of our end-to-end model EMC-GCN.

Refining Strategy and Prediction Layer

$$s_{ij} = \mathbf{h}_i \oplus \mathbf{h}_j \oplus \mathbf{r}_{ij} \oplus \mathbf{r}_{ii} \oplus \mathbf{r}_{jj} \qquad (11)$$

$$p_{ij} = \mathrm{softmax}(W_p s_{ij} + b_p) \qquad (12)$$

Loss Function

$$\mathcal{L} = \mathcal{L}_p + \alpha\mathcal{L}_{ba} + \beta\left(\mathcal{L}_{psc} + \mathcal{L}_{dep} + \mathcal{L}_{tbd} + \mathcal{L}_{rpd}\right) \quad (13)$$

$$\mathcal{L}_p = -\sum_i^n \sum_j^n \sum_{c\in\mathcal{C}} \mathbb{I}(y_{ij} = c)\log(p_{i,j|c}). \quad (14)$$

Experiments

Dataset		14res		14lap		15res		16res	
		#S	#T	#S	#T	#S	#T	#S	#T
$\mathcal{D}_1$	train	1,259	2,356	899	1,452	603	1,038	863	1,421
	dev	315	580	225	383	151	239	216	348
	test	493	1,008	332	547	325	493	328	525
$\mathcal{D}_2$	train	1266	2338	906	1460	605	1013	857	1394
	dev	310	577	219	346	148	249	210	339
	test	492	994	328	543	322	485	326	514

Table 2: Statistics for two groups of experiment datasets.

Model	**14res**			**14lap**			**15res**			**16res**		
	P	**R**	**F1**	**P**	**R**	**F1**	**P**	**R**	**F1**	**P**	**R**	**F1**
Peng-two-stage+IOG	58.89	60.41	59.64	48.62	45.52	47.02	51.70	46.04	48.71	59.25	58.09	58.67
IMN+IOG	59.57	63.88	61.65	49.21	46.23	47.68	55.24	52.33	53.75	-	-	-
GTS-CNN	70.79	61.71	65.94	55.93	47.52	51.38	60.09	53.57	56.64	62.63	66.98	64.73
GTS-BiLSTM	67.28	61.91	64.49	59.42	45.13	51.30	63.26	50.71	56.29	66.07	65.05	65.56
S^3E^2	69.08	64.55	66.74	59.43	46.23	52.01	61.06	56.44	58.66	71.08	63.13	66.87
GTS-BERT	70.92	69.49	70.20	57.52	51.92	54.58	59.29	58.07	58.67	68.58	66.60	67.58
BMRC	-	-	70.01	-	-	57.83	-	-	58.74	-	-	67.49
Our EMC-GCN	71.85	72.12	**71.98**	61.46	55.56	**58.32**	59.89	61.05	**60.38**	65.08	71.66	**68.18**

Table 3: Experimental results on $\mathcal{D}_1$ (Wu et al., 2020a). All baseline results are from the original papers.

Experiments

Model	14res			14lap			15res			16res		
	P	**R**	**F1**	**P**	**R**	**F1**	**P**	**R**	**F1**	**P**	**R**	**F1**
CMLA+♮	39.18	47.13	42.79	30.09	36.92	33.16	34.56	39.84	37.01	41.34	42.10	41.72
RINANTE+♮	31.42	39.38	34.95	21.71	18.66	20.07	29.88	30.06	29.97	25.68	22.30	23.87
Li-unified-R♮	41.04	67.35	51.00	40.56	44.28	42.34	44.72	51.39	47.82	37.33	54.51	44.31
Peng-two-stage♮	43.24	63.66	51.46	37.38	50.38	42.87	48.07	57.51	52.32	46.96	64.24	54.21
OTE-MTL†	62.00	55.97	58.71	49.53	39.22	43.42	56.37	40.94	47.13	62.88	52.10	56.96
JET-BERT♮	70.56	55.94	62.40	55.39	47.33	51.04	64.45	51.96	57.53	70.42	58.37	63.83
GTS-BERT†	68.09	69.54	68.81	59.40	51.94	55.42	59.28	57.93	58.60	68.32	66.86	67.58
BMRC†	75.61	61.77	67.99	70.55	48.98	57.82	68.51	53.40	60.02	71.20	61.08	65.75
BART-ABSA†	65.52	64.99	65.25	61.41	56.19	58.69	59.14	59.38	59.26	66.60	68.68	67.62
Our EMC-GCN	71.21	72.39	**71.78**	61.70	56.26	**58.81**	61.54	62.47	**61.93**	65.62	71.30	**68.33**

Table 4: Experimental results on $\mathcal{D}_2$ (Xu et al., 2020). The "♮" denotes that results are retrieved from Xu et al. (2020). The "†" means that we reproduce the models using released code with original parameters on the dataset.

Model	**14res**	**14lap**	**15res**	**16res**
EMC-GCN	**71.78**	**58.81**	**61.93**	**68.33**
w/o Ten Relations	70.68	57.71	59.85	66.48
w/o Linguistic Features	71.22	58.38	60.62	67.15
w/o Relation Constraint	70.59	57.28	59.83	67.89
w/o Refining Strategy	70.62	56.72	60.23	67.31

Table 5: F1 scores of ablation study on $\mathcal{D}_2$.

Thank you!