



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, rfl1, xjwang}@bupt.edu.cn

Code&dataset:<https://github.com/CCChenhao997/EMCGCN-ASTE>

ACL-2022



gesis
Leibniz-Institut
für Sozialwissenschaften



Reported by Dongdong Hu

Introduction

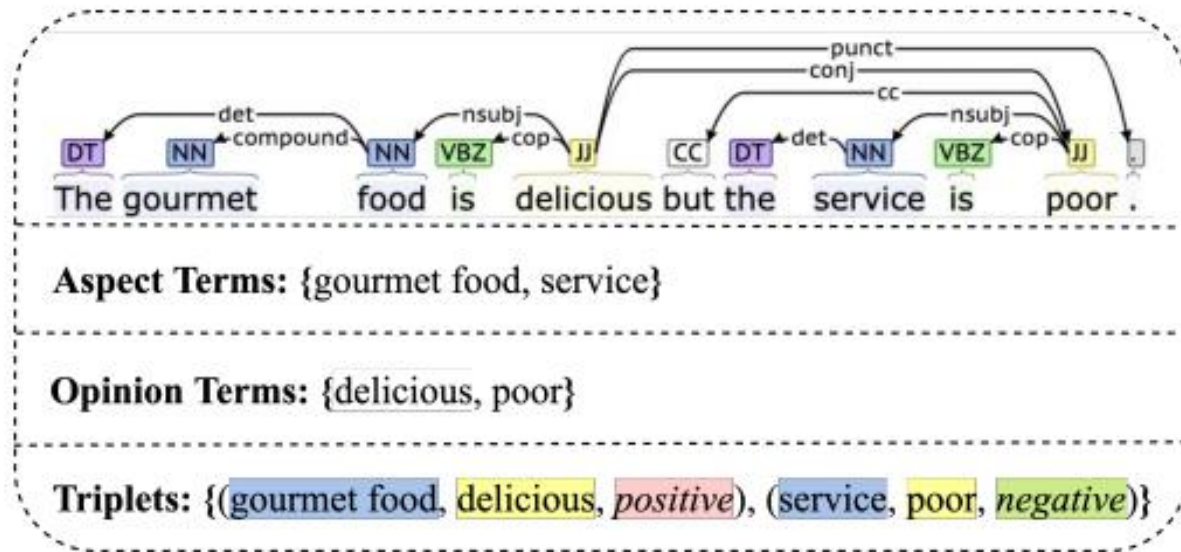


Figure 1: A sentence with its dependency tree is given to illustrate ASTE task. In the triplet set, aspect terms, opinion terms are highlighted in blue and yellow, respectively. The *positive* sentiment polarity is highlighted in red, while the *negative* in green.

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

$$\mathcal{T} = \{(a, o, s)_m\}_{m=1}^{|\mathcal{T}|}$$

$$S = \{POS, NEU, NEG\}$$

However, these methods ignore the **relations** between words for ASTE task

Method

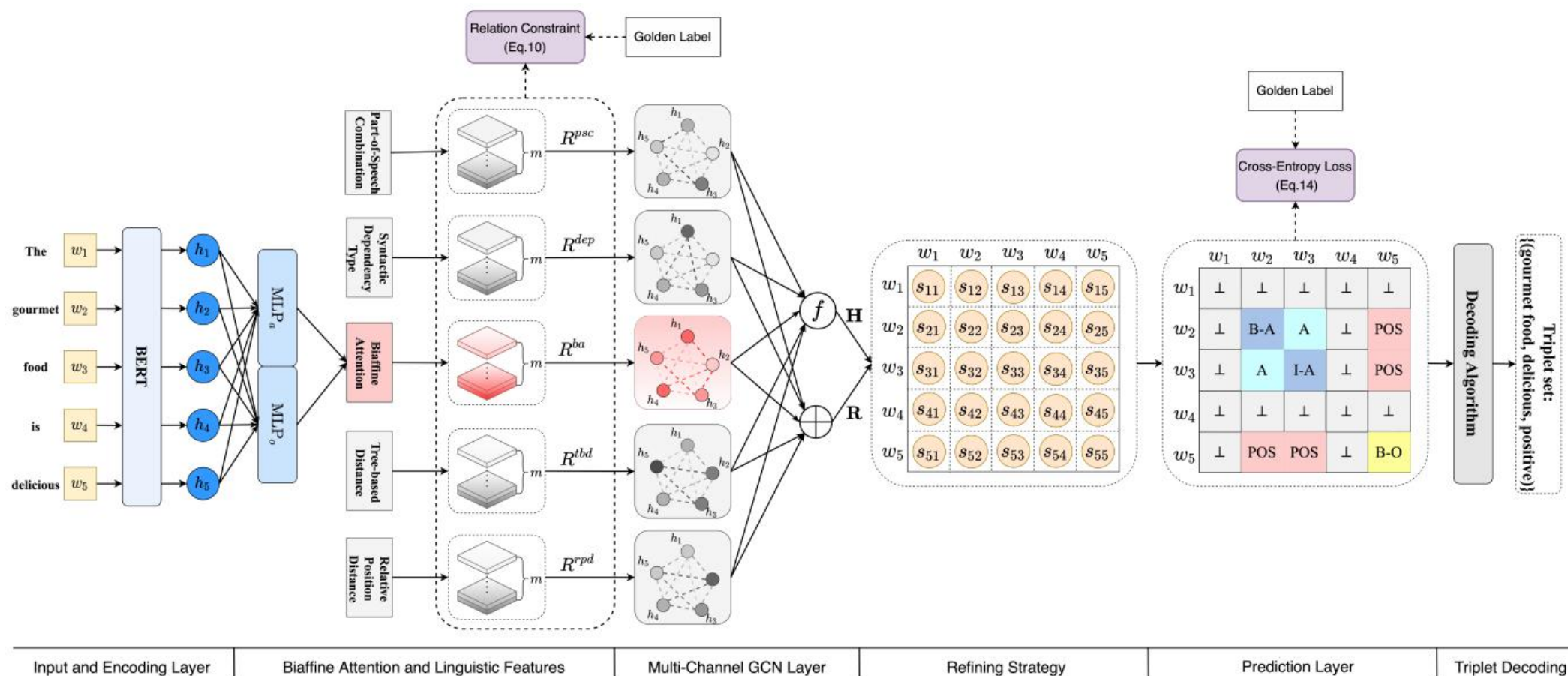


Figure 2: The overall architecture of our end-to-end model EMC-GCN.

Method

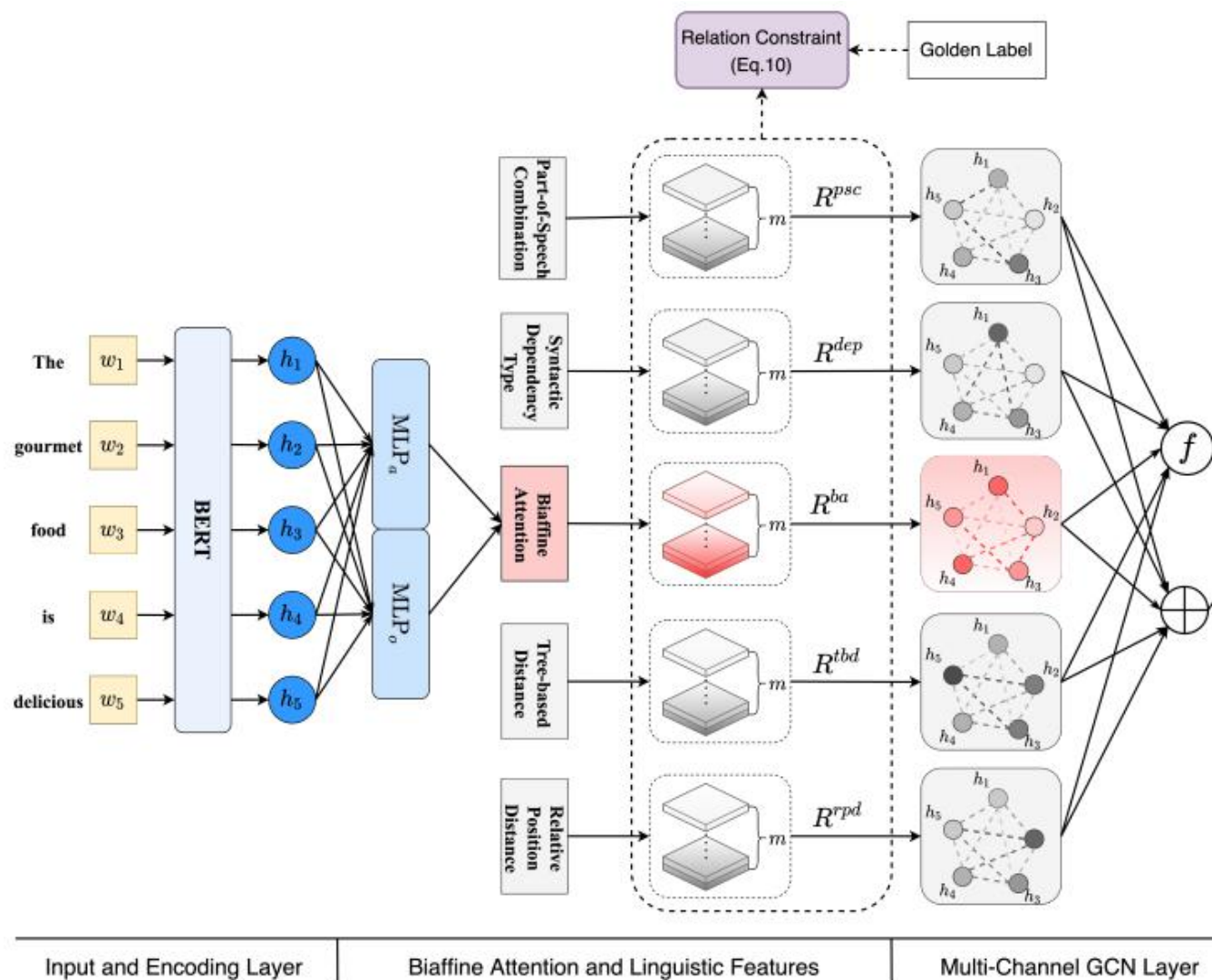
	The	gourmet	food	is	delicious	but	the	service	is	poor
The	⊥	⊥	⊥	⊥	⊥	⊥	⊥	⊥	⊥	⊥
gourmet	⊥	B-A	A	⊥	POS	⊥	⊥	⊥	⊥	⊥
food	⊥	A	I-A	⊥	POS	⊥	⊥	⊥	⊥	⊥
is	⊥	⊥	⊥	⊥	⊥	⊥	⊥	⊥	⊥	⊥
delicious	⊥	POS	POS	⊥	B-O	⊥	⊥	⊥	⊥	⊥
but	⊥	⊥	⊥	⊥	⊥	⊥	⊥	⊥	⊥	⊥
the	⊥	⊥	⊥	⊥	⊥	⊥	⊥	⊥	⊥	⊥
service	⊥	⊥	⊥	⊥	⊥	⊥	⊥	B-A	⊥	NEG
is	⊥	⊥	⊥	⊥	⊥	⊥	⊥	⊥	⊥	⊥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

#	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
8	NEU	an aspect term and an opinion term, and they form aspect-
9	NEG	opinion pair with positive/neutral/negative sentiment.
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.

Method



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

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

$$h_i^a = \text{MLP}_a(h_i) \quad (1)$$

$$h_j^o = \text{MLP}_o(h_j) \quad (2)$$

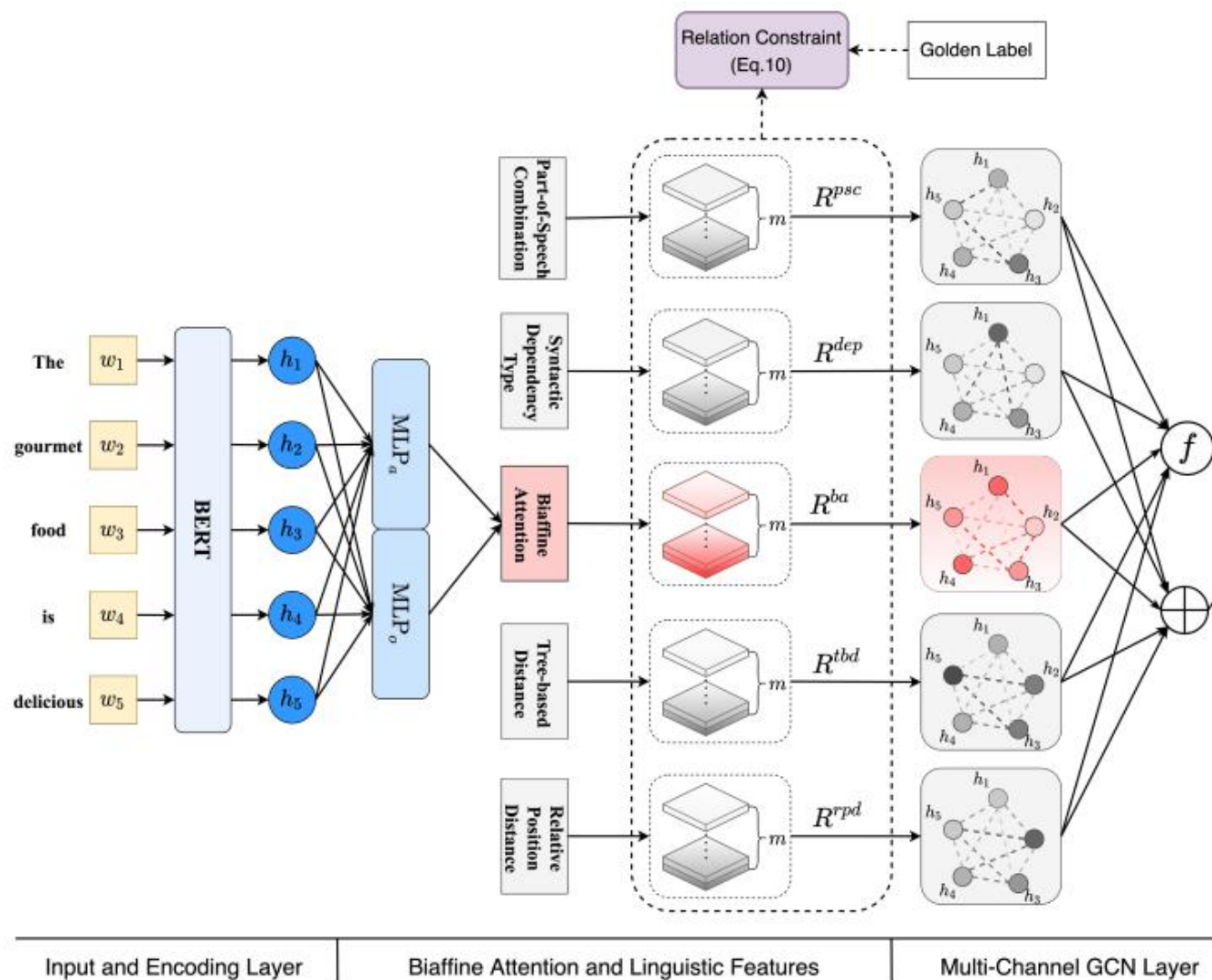
$$g_{i,j} = h_i^{aT} U_1 h_j^o + U_2 (h_i^a \oplus h_j^o) + b \quad (3)$$

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

$$R = \text{Biaffine}(\text{MLP}_a(H), \text{MLP}_o(H)) \quad (5)$$

The score vector $r_{i,j} \in \mathbb{R}^{1 \times m}$ models relations between w_i and w_j , m is the number of relation types and $r_{i,j,k}$ denotes the score of the k -th relation type for word pair (w_i, w_j) . The adjacency tensor $R \in \mathbb{R}^{n \times n \times m}$

Method



$R^{ba} \in \mathbb{R}^{n \times n \times m}$ which is constructed by the aforementioned biaffine attention module.

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

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

Method

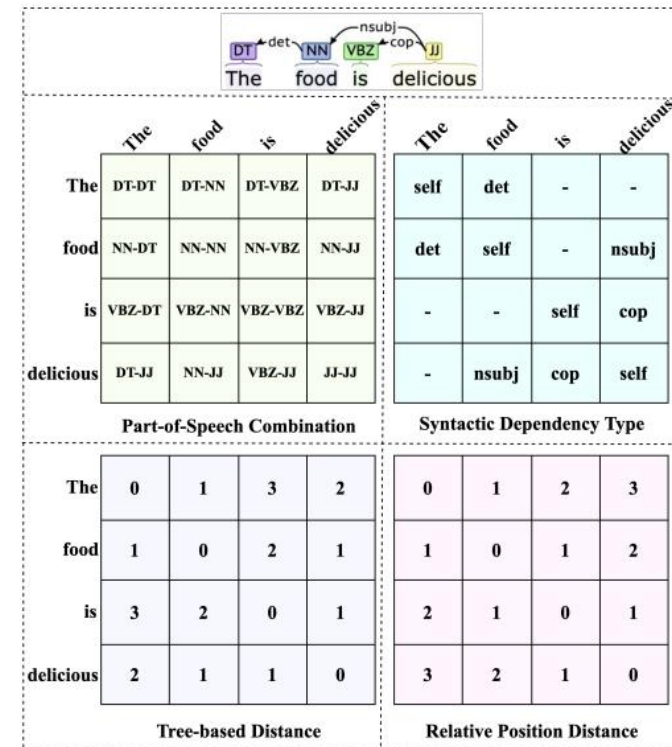
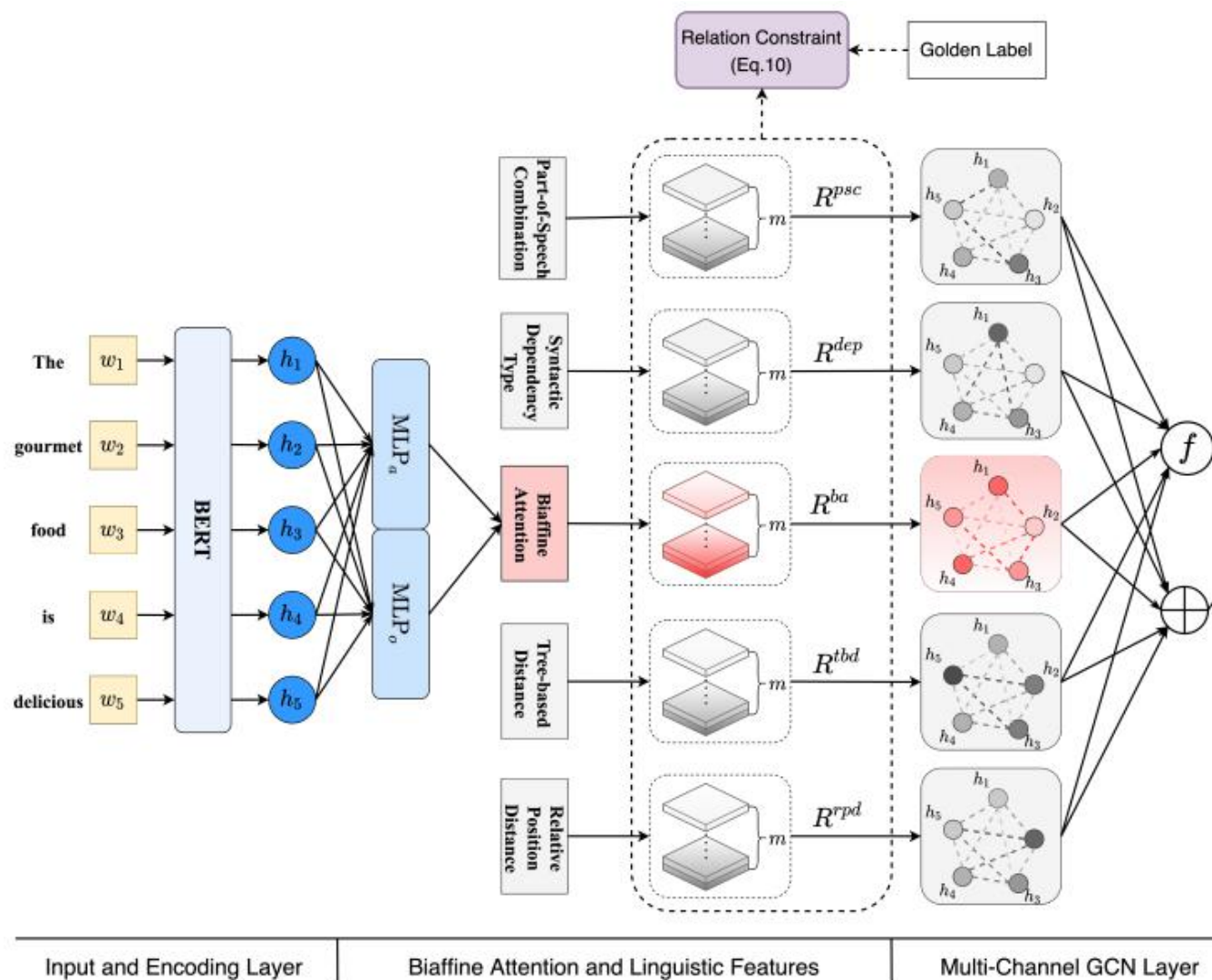


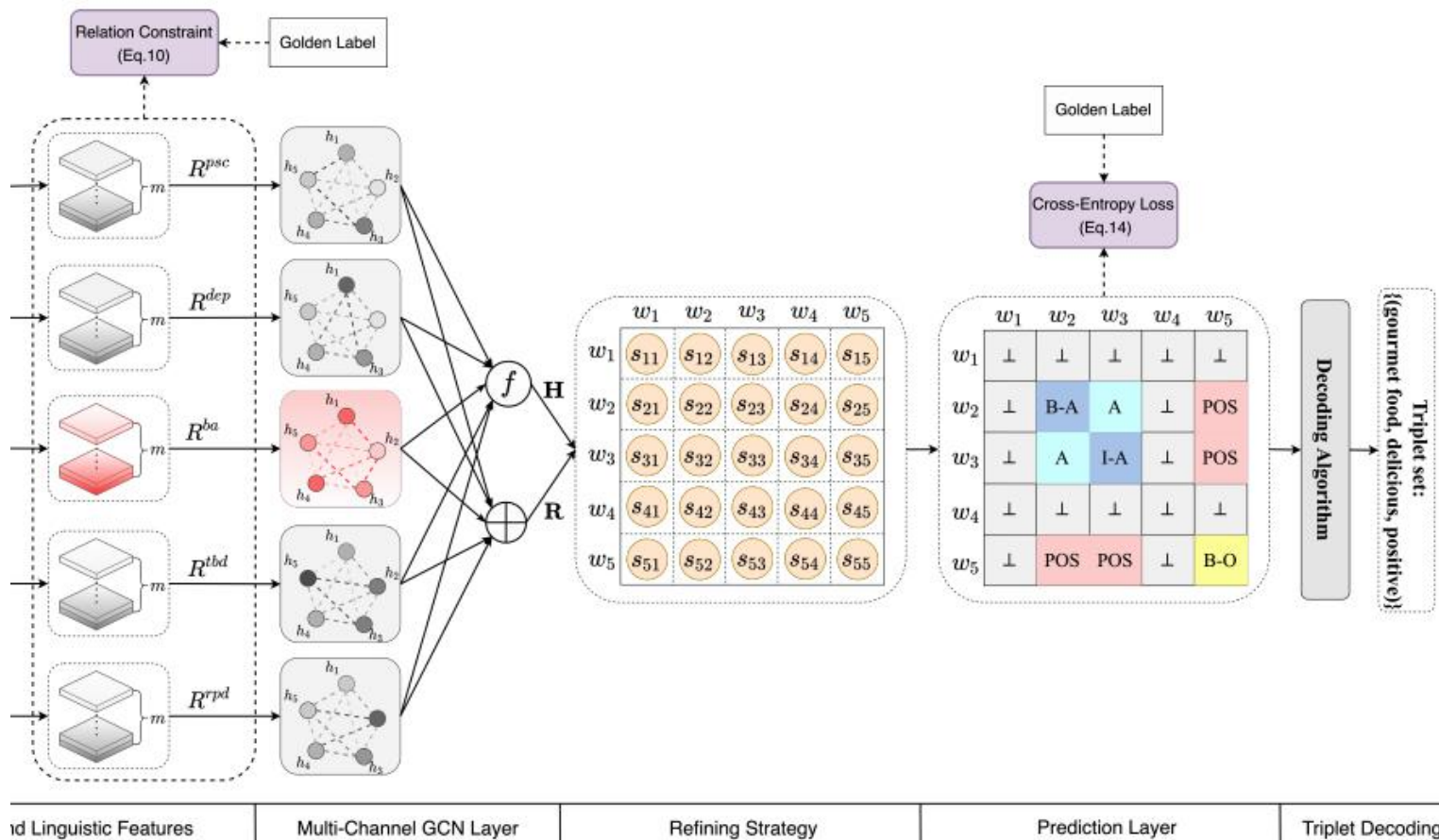
Figure 4: Four types of features for a sentence.

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

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

where $\mathbf{H} = \{h_1, h_2, \dots, h_n\}$ and $\mathbf{R} = \{r_{1,1}, r_{1,2}, \dots, r_{n,n}\}$ denote node representations and edge representations of word pairs.

Method



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

where $\mathbb{I}(\cdot)$ denotes the indicator function, y_{ij} is the ground truth of word-pair (w_i, w_j) , and \mathcal{C} denotes the relation set

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

$$p_{ij} = \text{softmax}(W_p s_{ij} + b_p) \quad (12)$$

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

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

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.

Method

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 ³ E 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
			73.92			57.20			60.19			68.85

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+ ^h	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+ ^h	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 ^h	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 ^h	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 ^h	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
			72.10			58.00			61.31			67.34

Table 4: Experimental results on \mathcal{D}_2 (Xu et al., 2020). The “^h” 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.

Experiments

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

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

Model	14res			14lap		
	POS	NEU	NEG	POS	NEU	NEG
EMC-GCN	74.69	19.65	62.43	67.74	19.14	56.20
<i>w/o</i> Refining Strategy	74.98	17.39	59.87	67.31	16.08	52.74

Table 6: F1 scores of three sentiment relations on \mathcal{D}_2 .

Experiments

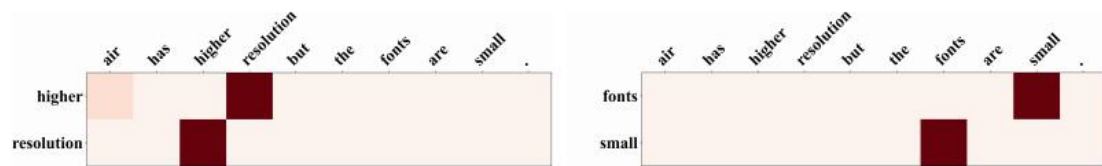


Figure 5: Visualization of *POS* and *NEG* relation channels of adjacency tensor R^{ba} obtained from the biaffine attention.

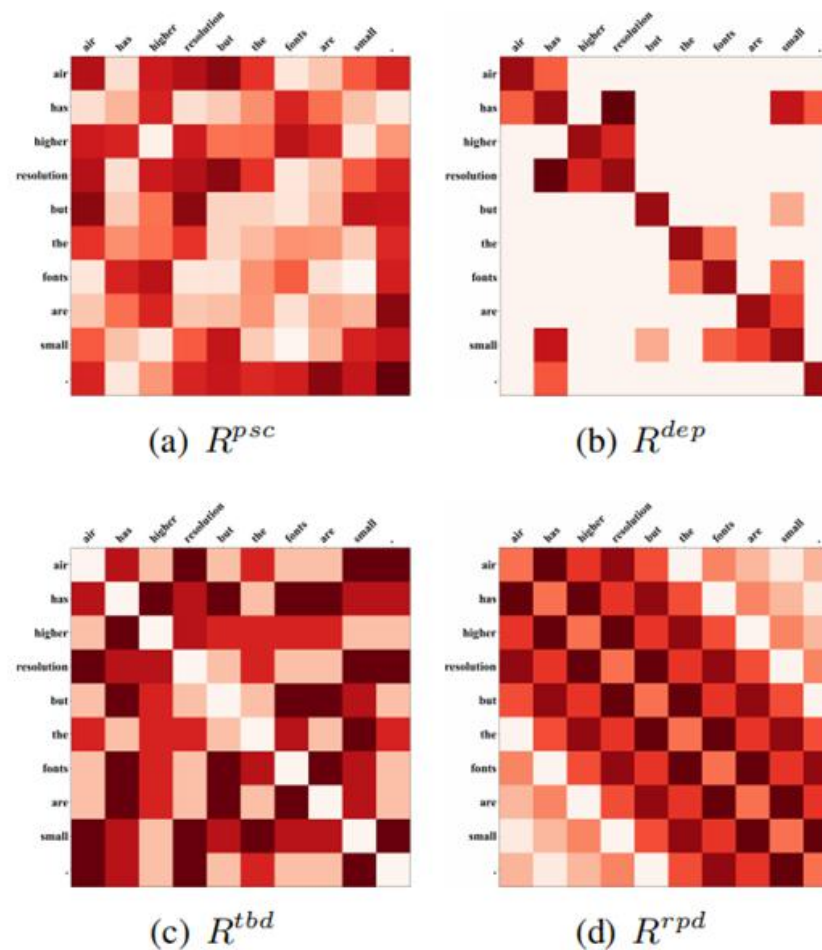


Figure 6: Visualization of adjacency tensors of four linguistic features.

Experiments

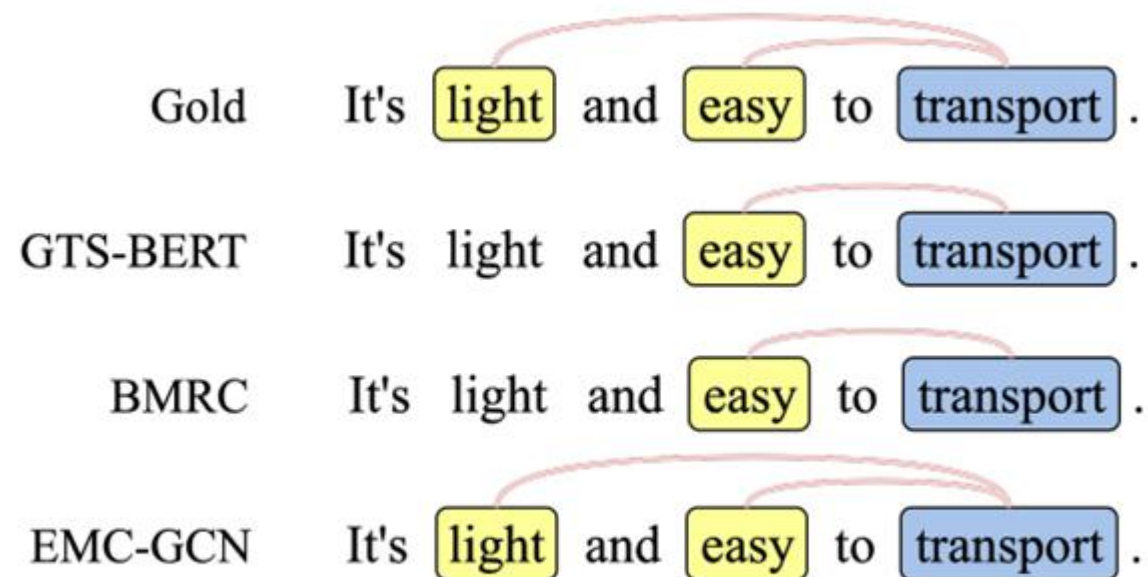


Figure 7: Different models outputs for a given sentence.



Thanks