Chongqing University of Technology

ATAI
Advanced Technique of Artificial Intelligence

BiSyn-GAT+: Bi-Syntax Aware Graph Attention Network for Aspect-based Sentiment Analysis

Shuo Liang [1] , Wei Wei [2,✉], Xian-Ling Mao [3], Fei Wang [4], Zhiyong He [5]

[1,2] Cognitive Computing and Intelligent Information Processing (CCIIP) Laboratory, School of Computer Science and Technology, Huazhong University of Science and Technology

[3] School of Computer Science and Technology, Beijing Institute of Technology

[4] Institute of Computig Technology, Chinese Academy of Sciences

[5] Naval University of Engineering

[1] shuoliang@hust.edu.cn, [2] weiw@hust.edu.cn, [3] maoxl@bit.edu.cn, [4] wangfei@ict.ac.cn, [5] moonmon_pub@outlook.com

code:https://github.com/CCIIPLab/BiSyn_GAT_plus

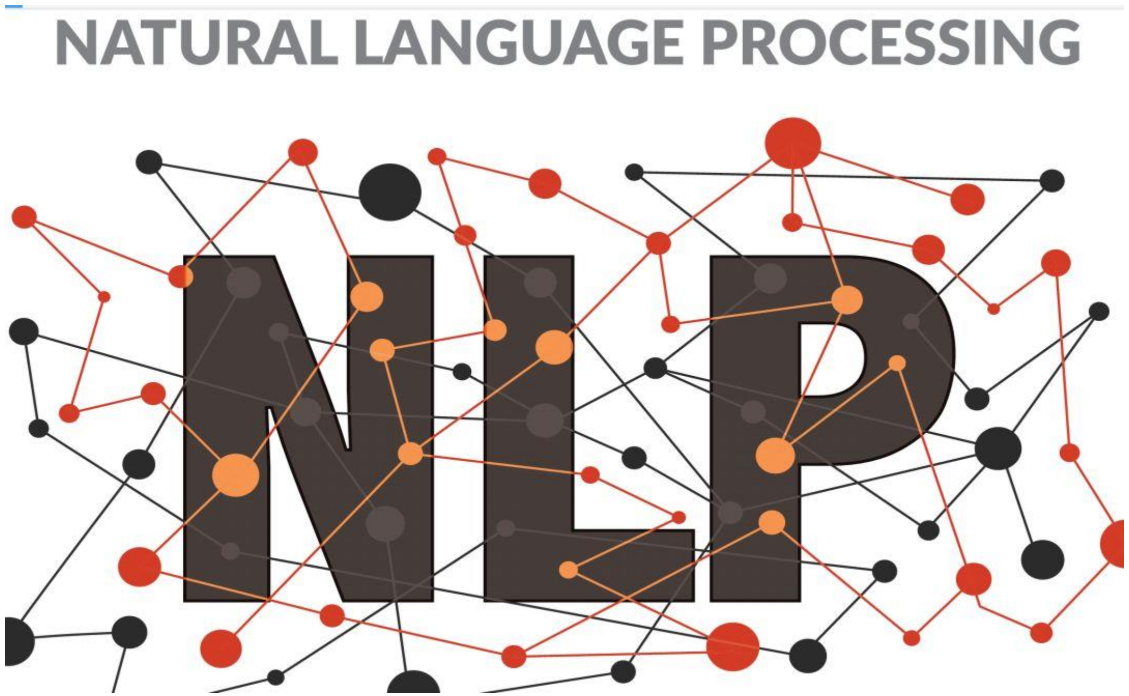
2022. 12. 15 • ChongQing

2022_ACL

Reported by Junhao Cao

gesis Leibniz-Institut für Sozialwissenschaften

Forschungszentrum L3S Research Center

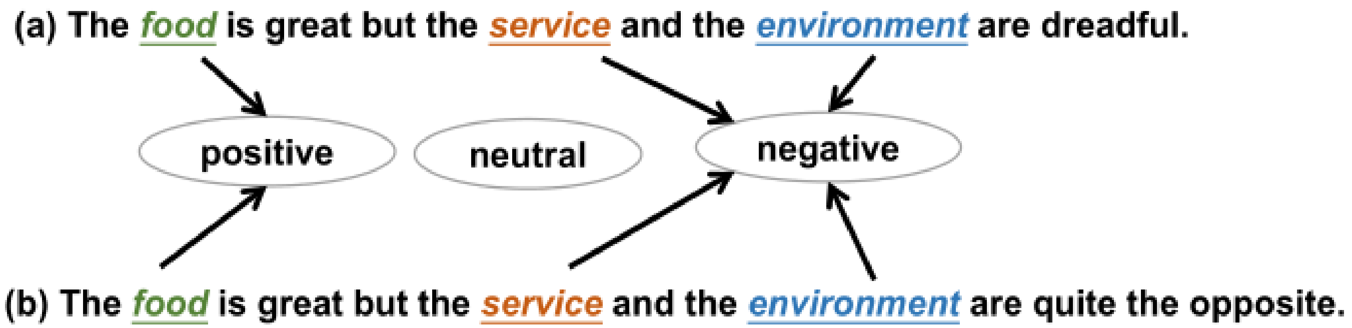


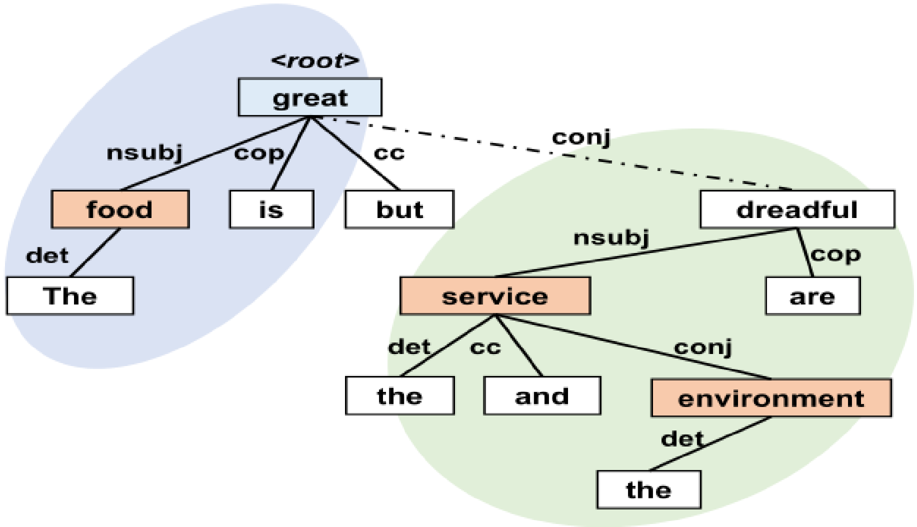
1.Introduction

2.Method

3.Experiments

Introduction

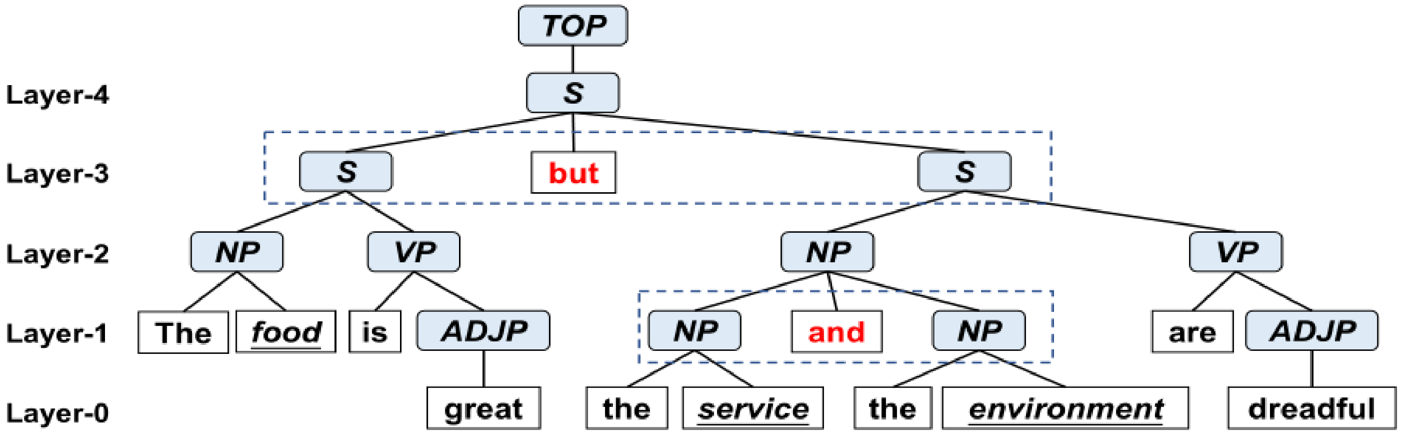




Many previous methods often assume that words closer to the target aspect are more likely to be related to its sentiment.

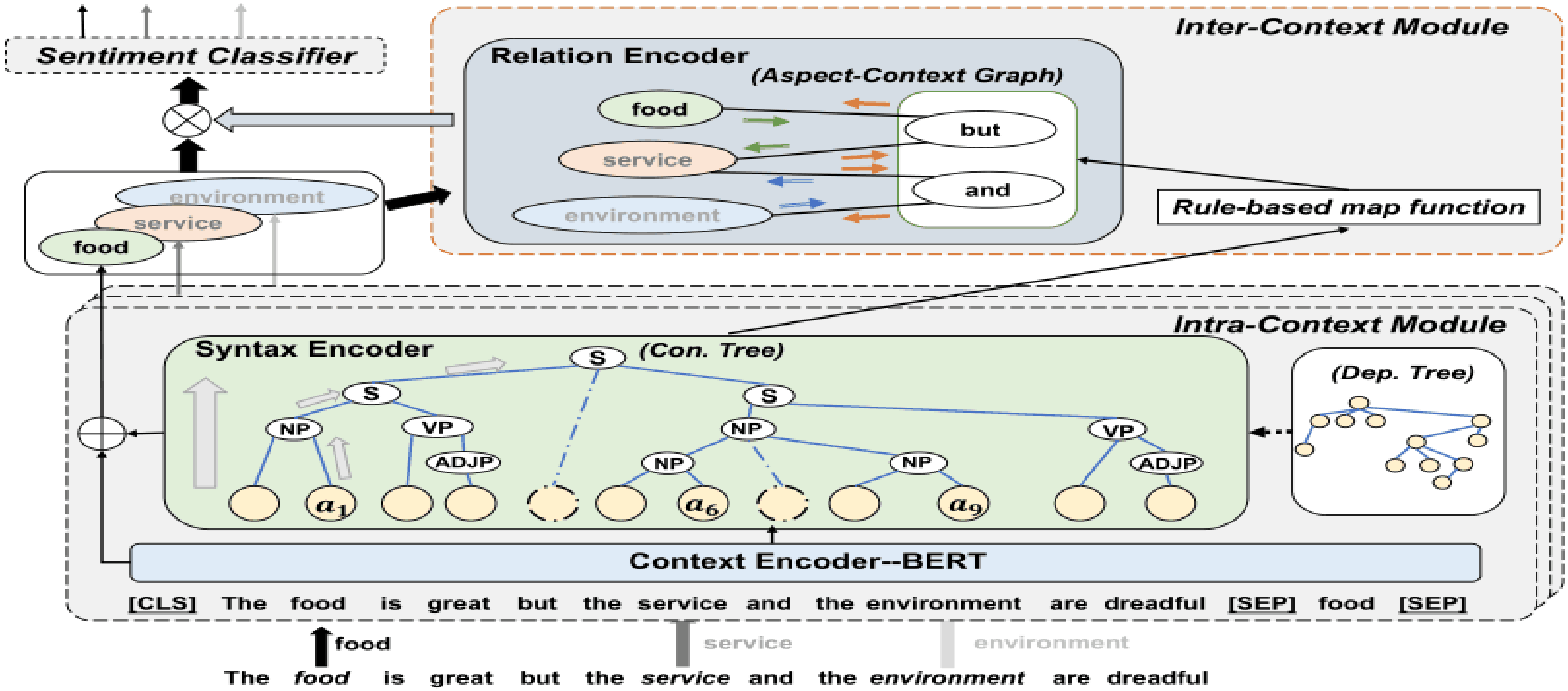
the inherent properties of dependency structure tree may introduce noise.

Introduction

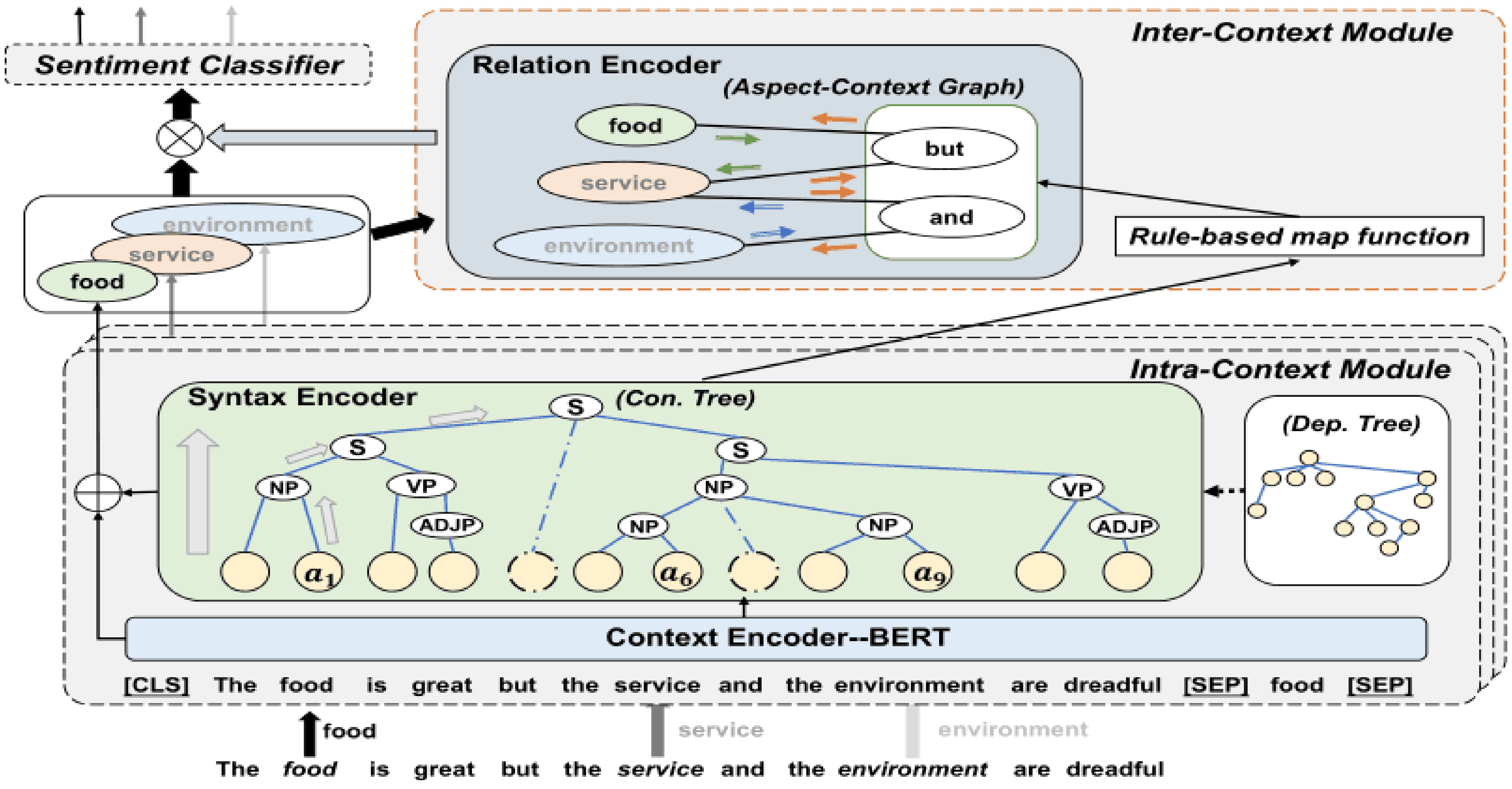


a constituent tree often contains phrase segmentation and hierarchical composition structure.

Method


Sentiment Classifier
Inter-Context Module
Relation Encoder
(Aspect-Context Graph)
food
service
environment
but
and
Rule-based map function
environment
service
food
Intra-Context Module
Syntax Encoder
(Con. Tree)
S
S
S
NP
VP
ADJP
NP
NP
NP
VP
ADJP
a_1
a_6
a_9
(Dep. Tree)
Context Encoder--BERT
[CLS] The food is great but the service and the environment are dreadful [SEP] food [SEP]
food
service
environment
The food is great but the service and the environment are dreadful

Method



Context Encoder

$$BERT\_seq_t = [\text{CLS}]+\{w_i\}+[\text{SEP}]+a_t+[\text{SEP}] \,, \quad (1)$$

$$h^t = \left\{h_0^t, h_1^t, \ldots, h_{n'}^t, \ldots, h_{n'+2+m_t'}^t\right\} \quad (2)$$

$$\hat{\mathbf{h}}_i^t = \frac{1}{|BertT(w_i)|} \sum_{k \in BertT(w_i)} h_k^t, \quad (3)$$

Method

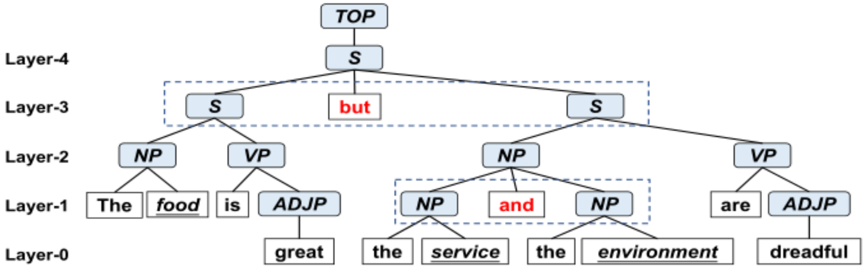
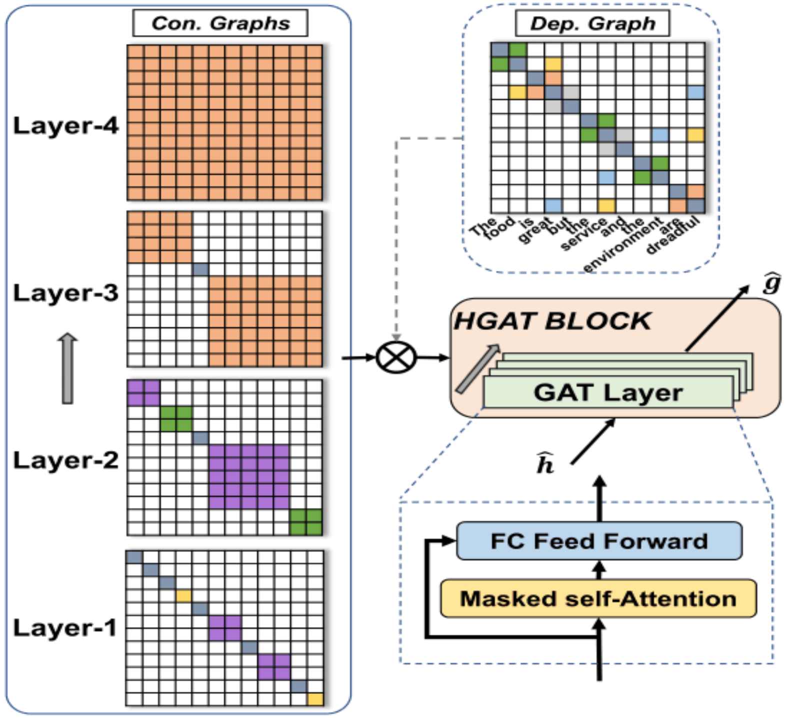


Figure 3: Constituent tree of the sentence "The food is great but the service and the environment are dreadful". Context words are in rectangles and parsed phrase types are in rounded rectangles.



Syntax Encoder

Graph construction

$$\mathbf{CA}_{i,j}^{l} = \begin{cases} 1 & \text{if } w_i, w_j \text{ in same phrase of } \{ph_u^l\} \\ 0 & \text{otherwise} \end{cases}, \tag{4}$$

HGAT block

$$\hat{\mathbf{g}}_i^{t,l} = FC(\mathbf{g}_i^{t,l} + \hat{\mathbf{g}}_i^{t,l-1}), \tag{5}$$

$$\mathbf{g}_i^{t,l} = \|_{z=1}^{Z} \sigma \left(\sum_{j \in \mathcal{N}^{t,l}(i)} \alpha_{ij}^{lz} \mathbf{W}_g^{lz} \hat{\mathbf{g}}_j^{t,l-1} \right), \tag{6}$$

$$\alpha_{ij}^{lz} = \frac{\exp\left(f\left(\hat{\mathbf{g}}_i^{t,l-1}, \hat{\mathbf{g}}_j^{t,l-1}\right)\right)}{\sum_{j' \in \mathcal{N}^l(i)} \exp\left(f\left(\hat{\mathbf{g}}_i^{t,l-1}, \hat{\mathbf{g}}_{j'}^{t,l-1}\right)\right)}, \tag{7}$$

Method

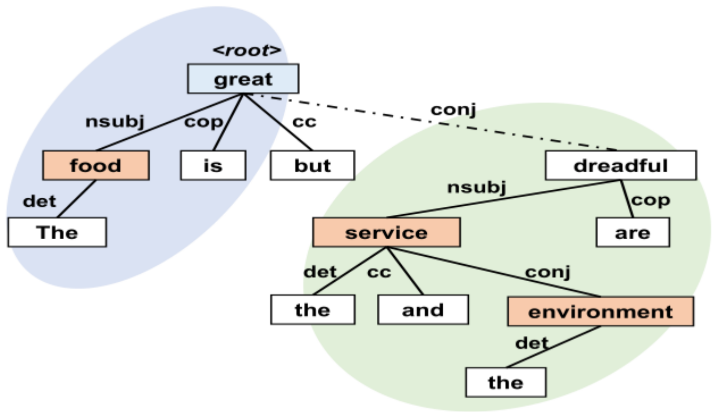
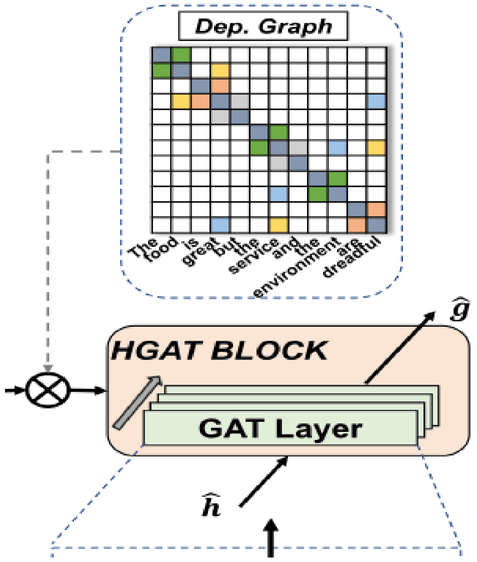
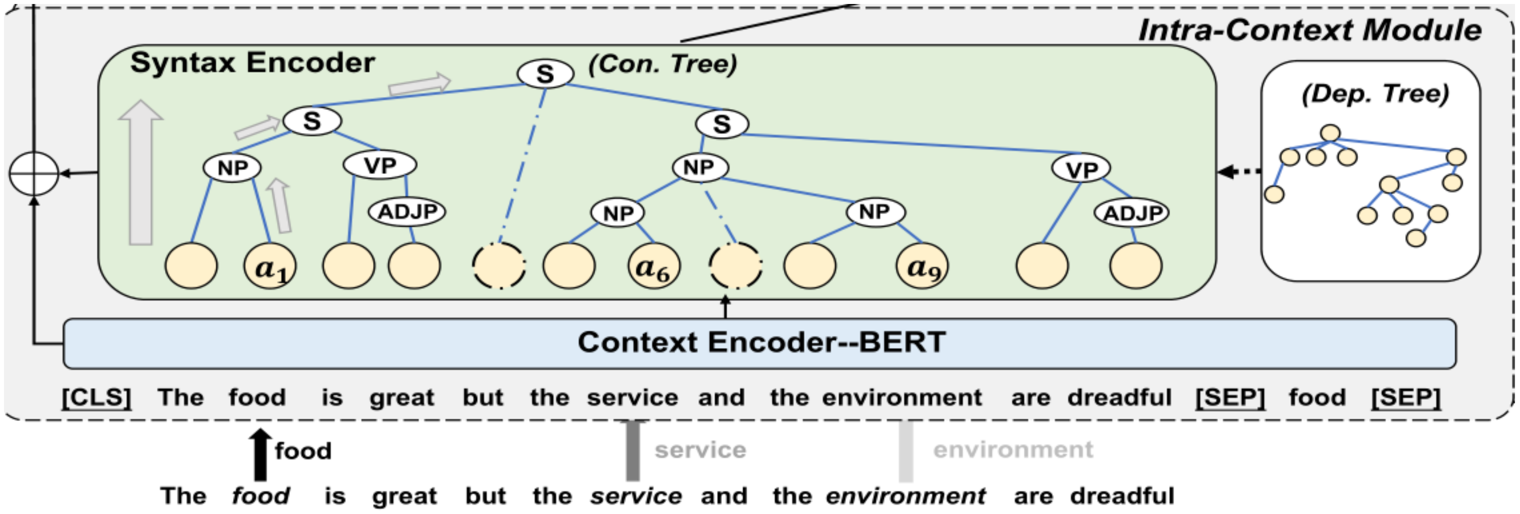


Figure 2: Dependency tree of “The food is great but the





With dependency information

$$\mathbf{DA}_{i,j} = \begin{cases} 1 & \text{if } w_i, w_j \text{ link directly in Dep.Tree} \\ 0 & \text{otherwise} \end{cases} \tag{8}$$

position-wise dot

$$\mathbf{FA} = \mathbf{CA} \cdot \mathbf{DA} \tag{9}$$

position-wise add

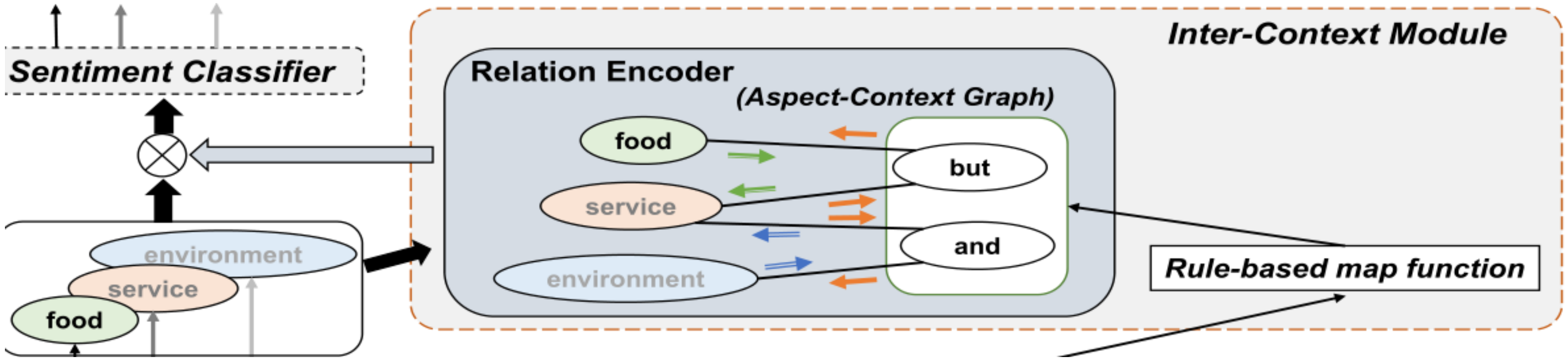
$$\mathbf{FA} = \mathbf{CA} + \mathbf{DA} \tag{10}$$

conditional position-wise add

$$\mathbf{FA} = \mathbf{CA} \oplus \mathbf{DA} \tag{11}$$

$$\mathbf{v}_t^{as} = \left[\hat{\mathbf{h}}_t^t + \hat{\mathbf{g}}_t^t ; h_0^t\right] \tag{12}$$

Method



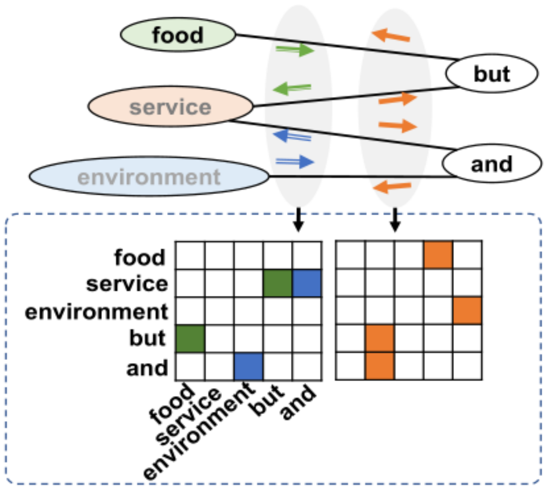


Figure 6: Example of an aspect-context graph and corresponding two adjacent matrices for distinguishing the bi-directional relations.

Inter-Context Module

Phrase segmentation

$$PS(a_i, a_j) = \begin{cases} \{w_k\}, & \text{if } |Br(a_i, a_j)| = 0 \\ Br(a_i, a_j), & otherwise \end{cases}, \tag{13}$$

Training

$$\mathbf{o_t} = \mathbf{v}_t^{as} + \mathbf{v}_t^{aa}, \tag{14}$$

$$\mathbf{p(t)} = softmax(\mathbf{W_p o_t} + \mathbf{b_p}), \tag{15}$$

$$L(\theta)^{Sentiment} = -\sum_{s} \sum_{a_t \in A_s} loss(\mathbf{p(t)}, y(t)), \tag{16}$$

Experiment

Dataset		Sentence-Level			Aspect-Level		
		Multi-Asp.	Single-Asp.	All	Pos.	Neg.	Neu.
Rest-aurant	Train	971	1009	1980	2164	807	637
	Test	315	284	599	727	196	196
Laptop	Train	538	916	1454	937	851	455
	Test	150	259	409	337	128	167
MAMS	Train	4297	0	4297	3380	2764	5042
	valid	500	0	500	403	325	604
	Test	500	0	500	400	329	607
Twitter	Train	0	6051	6051	1507	1528	3016
	Test	0	677	677	172	169	336

Table 1: Statistics of datasets. Multi-Asp., Single-Asp. indicate the number of sentences with multiple or single aspect; Pos., Neg., and Neu. show the number of aspects towards positive, negative and neutral label.

Category	Model	Dataset							
		Restaurant		Laptop		MAMS		Twitter	
		Acc.(%)	F1.(%)	Acc.(%)	F1.(%)	Acc.(%)	F1.(%)	Acc.(%)	F1.(%)
w/o Syn.	BERT-SPC	84.46	76.98	78.99	75.03	82.82	81.90	73.55	72.14
	AEN-BERT	83.12	73.76	79.93	76.31	-	-	74.71	73.13
w/ Syn.	R-GAT	86.60	81.35	78.21	74.07	-	-	76.15	74.88
	RGAT+	86.68	80.92	80.94	78.20	84.52	83.74	76.28	75.25
	DGEDT	86.30	80.00	79.80	75.60	-	-	<u>77.90</u>	75.40
	DualGCN	87.13	81.16	81.80	78.10	-	-	77.40	<u>76.02</u>
	SDGCN	83.57	76.47	81.35	78.34	-	-	-	-
	InterGCN	87.12	81.02	<u>82.87</u>	<u>79.32</u>	-	-	-	-
Ours	BiSyn-GAT	<u>87.49</u>	<u>81.63</u>	82.44	79.15	<u>84.90</u>	<u>84.43</u>	**77.99**	**76.80**
	BiSyn-GAT+	**87.94**	**82.43**	**82.91**	**79.38**	**85.85**	**85.49**		

Table 2: Performance comparison of models on four datasets. The best are in **bold**, and second-best are <u>underlined</u>.

Experiment

Category	Ablation	Dataset							
		Restaurant		Laptop		MAMS		Twitter	
		Acc.(%)	F1.(%)	Acc.(%)	F1.(%)	Acc.(%)	F1.(%)	Acc.(%)	F1.(%)
w/o AA	w/o syn. & dep.(BERT+)	84.99	78.51	79.11	75.76	82.71	82.22	75.48	74.54
	w/o con.	86.42	80.10	80.22	76.42	83.38	82.90	76.51	75.29
	w/o dep.	86.60	81.51	81.80	78.48	84.58	84.09	76.81	75.86
	con.$\times$dep.	86.86	80.82	80.85	77.27	84.21	83.76	76.51	75.37
	con.+dep.	86.86	81.59	82.12	78.93	84.73	84.14	<u>77.40</u>	<u>76.39</u>
	con.$\oplus$dep. (BiSyn-GAT)	87.49	81.63	82.44	79.15	84.90	84.43	**77.99**	**76.80**
w/ AA	con.+dep.	<u>87.76</u>	<u>82.18</u>	<u>82.75</u>	<u>79.16</u>	<u>85.48</u>	<u>85.05</u>	-	-
	con.$\oplus$dep. (BiSyn-GAT+)	**87.94**	**82.43**	**82.91**	**79.38**	**85.85**	**85.49**	-	-

Table 3: Ablation study. Notations "con." and "dep." represent syntax information from constituent tree and dependency tree, respectively. $\times, +, \oplus$ represent the position-wise dot, position-wise add, conditional position-wise add operations, respectively, when fusing two syntax information. "AA" represents modeling aspect-aspect relations. The best performances are in **bold**, and second-best are <u>underlined</u>.

Experiment

Model		Dataset			
		Restaurant		MAMS	
		Acc.(%)	F1.(%)	Acc.(%)	F1.(%)
BiSyn-GAT		87.49	81.63	84.90	84.43
aspect-context graph	w/ Bi-relation	**87.94**	**82.43**	**85.85**	**85.49**
	w/o Bi-relation	87.85	82.27	85.10	84.69
aspect graph	adjacent	87.49	81.69	85.10	84.61
	Bi-adjacent	87.40	81.53	85.18	84.74
	global	87.49	81.70	85.32	84.88

Table 4: Performance comparison of aspect-context graph variants on Restaurant and MAMS dataset. The best performances are in **bold**.

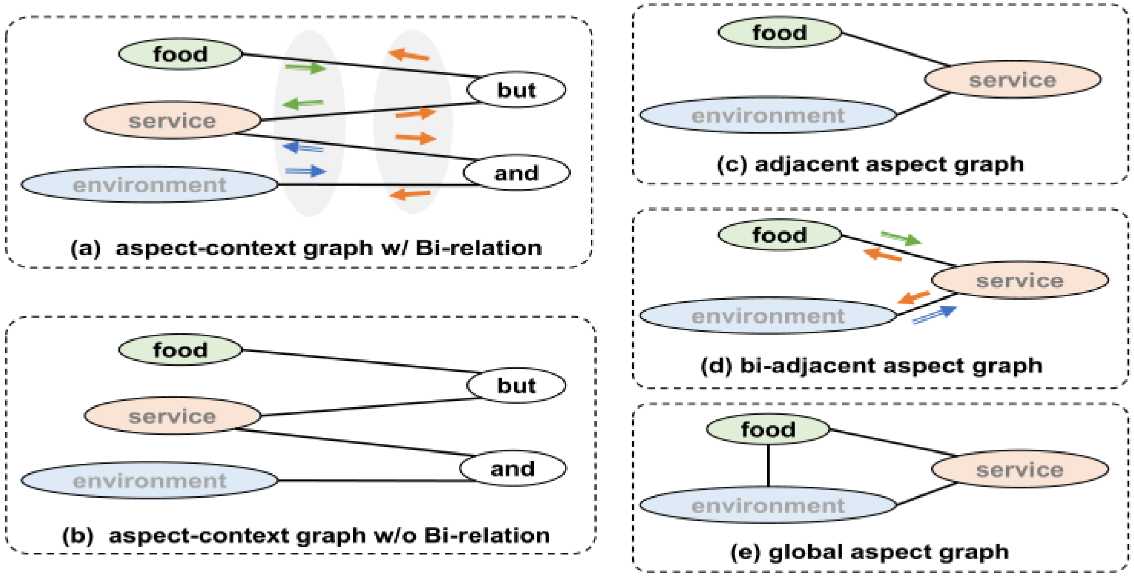


Figure 7: Illustrations of variants when investigating the effects of aspect-context graph.

Experiment

Model	Parser	Restaurant Acc.(%)	Restaurant F1.(%)	MAMS Acc.(%)	MAMS F1.(%)
Base		84.99	78.51	82.71	82.22
w/o dep.	Stanford Parser	86.51	81.34	84.51	84.06
	SuPar	86.60	81.51	84.58	84.09
BiSyn-GAT	Stanford Parser	86.66	81.56	84.88	84.31
	SuPar	87.49	81.63	84.90	84.43
BiSyn-GAT+	Stanford Parser	87.84	82.39	85.78	85.40
	SuPar	87.94	82.43	85.85	85.49

Table 5: Experiments results with different parsers. w/o dep. is one variant of BiSyn-GAT, only using constituent information.

Experiment

Sentences	Aspects	BiSyn-GAT	BiSyn-GAT+
it doesn't look like much on the **outside**$_{neg}$, but the minute you walk inside, it's a whole other **atmosphere**$_{pos}$.	outside	neu ✗	neg ✓
	atmosphere	pos ✓	pos ✓
while the **service**$_{neg}$ and **setting**$_{neg}$ were average , the **food**$_{pos}$ was excellent.	service	neg ✓	neg ✓
	setting	neu ✗	neg ✓
	food	pos ✓	pos ✓
food was average, the **appetizers**$_{pos}$ were better than the **main courses**$_{neu}$.	appetizers	pos ✓	pos ✓
	main courses	pos ✗	neu ✓
i have no complaints about the **wait**$_{pos}$ or the **service**$_{pos}$ but the **pizza**$_{neg}$ was bit at all something to write home about.	wait	neu ✗	pos ✓
	service	neg ✗	pos ✓
	pizza	neg ✓	neg ✓

Table 6: Predictions from *BiSyn-GAT* and *BiSyn-GAT+*. The notations pos, neg, and neu in the table represent positive, negative, and neutral. For each sentence, the aspects are displayed in bold, with golden sentiment polarities as the subscripts. The phrase segmentation words are shown underline between the corresponding two aspects. False predictions are marked with ✗ while true predictions are marked with ✓.

Experiment

Con. Tree Depth	Dataset								
	Restaurant		Laptop		MAMS			Twitter	
	Train	Test	Train	Test	Train	Valid	Test	Train	Test
1	177	68	206	84	208	16	19	1215	117
2	369	135	724	247	1301	152	141	1066	**147**
3	**462**	**148**	**936**	**312**	**2265**	244	261	**1186**	123
4	363	108	612	202	2085	**276**	**292**	947	96
5	311	75	429	116	1761	203	194	677	79
6	237	40	266	73	1211	141	157	414	57
7	136	27	205	41	901	99	117	246	23
8	108	10	106	18	545	81	65	145	22
9	59	8	43	14	380	57	34	86	8
≥ 10	60	13	81	12	529	63	56	69	5
MAX.	18	13	17	13	19	17	15	14	11

Table 7: Depth distribution of parsed constituent trees on four datasets. The maximums are in **bold**. The last row lists the max tree depth of each dataset.

Multi. Aspect Distribution	Dataset						
	Restaurant		Laptop		MAMS		
	Train	Test	Train	Test	Train	Valid	Test
2	555	192	343	101	2568	285	264
3	261	73	137	33	1169	136	173
4	103	31	40	9	364	55	45
5	32	14	9	6	126	16	10
6	11	3	5	1	48	5	5
7	5	1	3	-	13	2	-
8	3	-	-	-	6	-	1
9	1	-	-	-	1	-	-
10	-	-	-	-	1	1	1
11	-	-	-	-	1	1	1
13	-	1	1	-	-	-	

Table 8: Multi.aspect distribution of three datasets.

Thank you!

Leibniz-Institut für Sozialwissenschaften