Aspect Feature Distillation and Enhancement Network for Aspect-based Sentiment Analysis

Rui Liu
Institute of Information Engineering, Chinese Academy of Sciences
School of Cyber Security, University of Chinese Academy of Sciences
Beijing, China
liurui3221@iie.ac.cn

Jiahao Cao
Institute of Information Engineering, Chinese Academy of Sciences
School of Cyber Security, University of Chinese Academy of Sciences
Beijing, China
caojiahao@iie.ac.cn

Nannan Sun
Institute of Information Engineering, Chinese Academy of Sciences
School of Cyber Security, University of Chinese Academy of Sciences
Beijing, China
sunnannan@iie.ac.cn

Lei Jiang*
Institute of Information Engineering, Chinese Academy of Sciences
School of Cyber Security, University of Chinese Academy of Sciences
Beijing, China
jianglei@iie.ac.cn

（SIGIR-2022)

gesis Leibniz-Institut für Sozialwissenschaften

Forschungszentrum · Research Center L3S

Reported by Zhaoze Gao

1. Introduction
2. Approach
3. Experiments

Introduction

The **ambience** was nice, but **service** wasn't so great.

two challenges:

the attention mechanism paying partial attention to aspect-unrelated words inevitably introduces irrelevant noise.

the cross-entropy loss lacks discriminative learning of features, which makes it difficult to exploit the implicit information of intra-class compactness and inter-class separability.

Approach

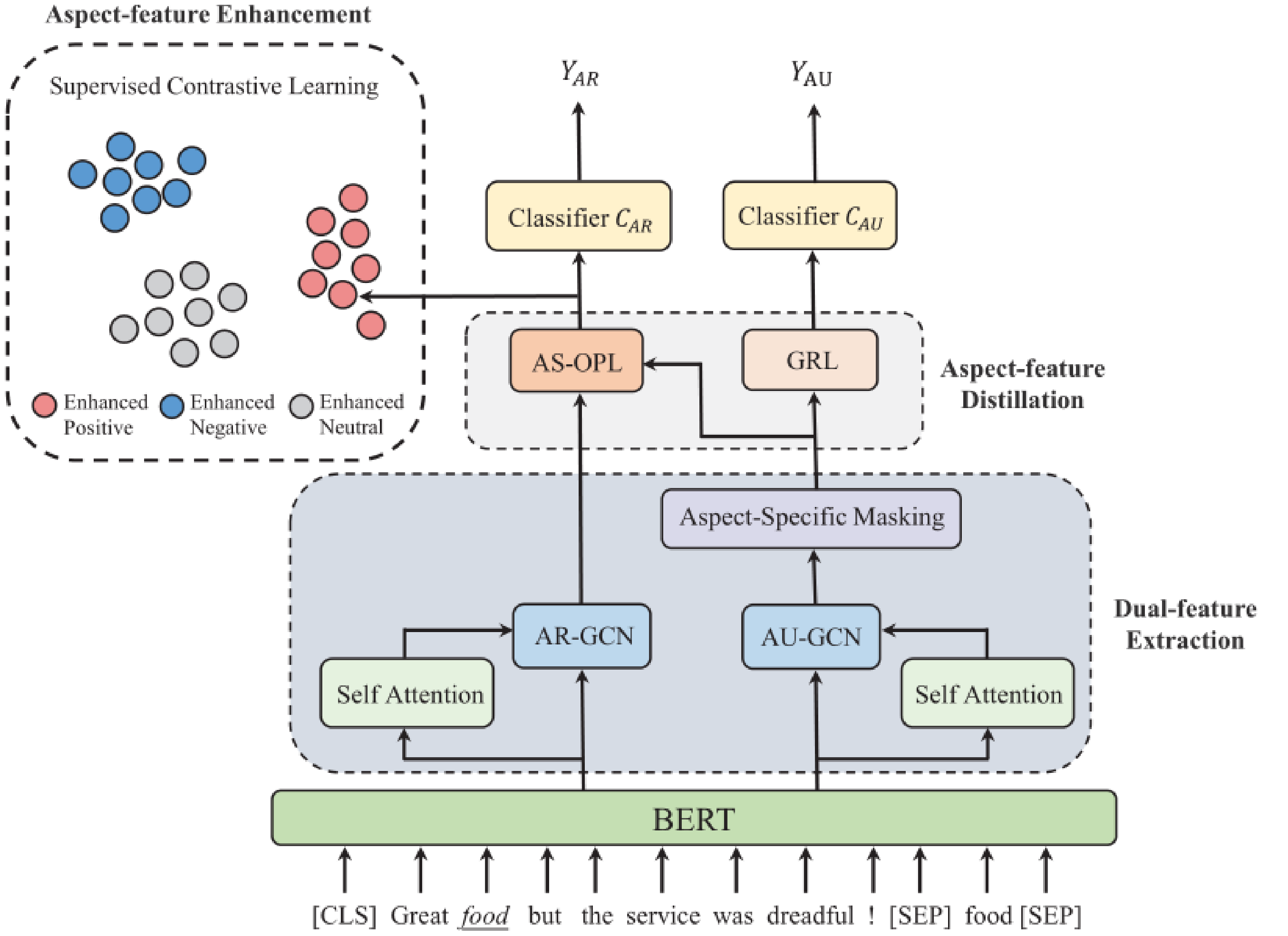


Figure 1: The overall architecture of AFDEN, which is mainly composed of dual-feature extraction module, aspect-feature distillation module and aspect-feature enhancement module. The details of our model are described in the main text.

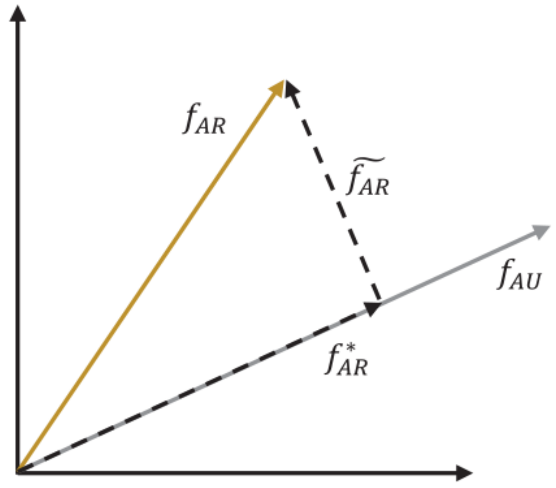
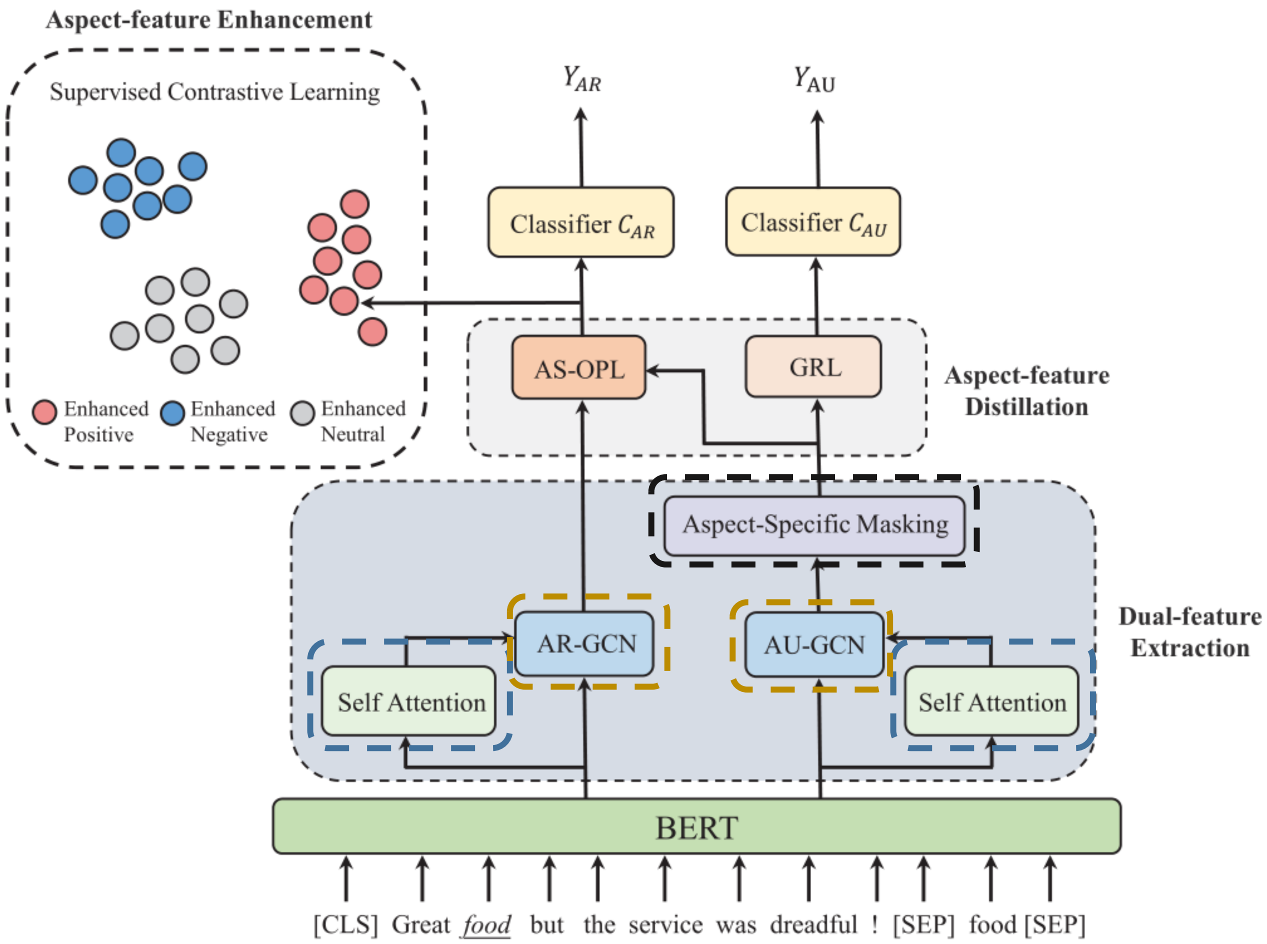


Figure 2: The principle of AS-OPL.

Approach



$$S = [\omega_1, \ \ldots, \ \omega_{a+1}, \ \ldots, \ \omega_{a+m}, \ \ldots, \ \omega_n]$$

$$A = [\omega_{a+1}, \ \ldots, \ \omega_{a+m}]$$

$$y \in \{positive, negative, neutral\}$$

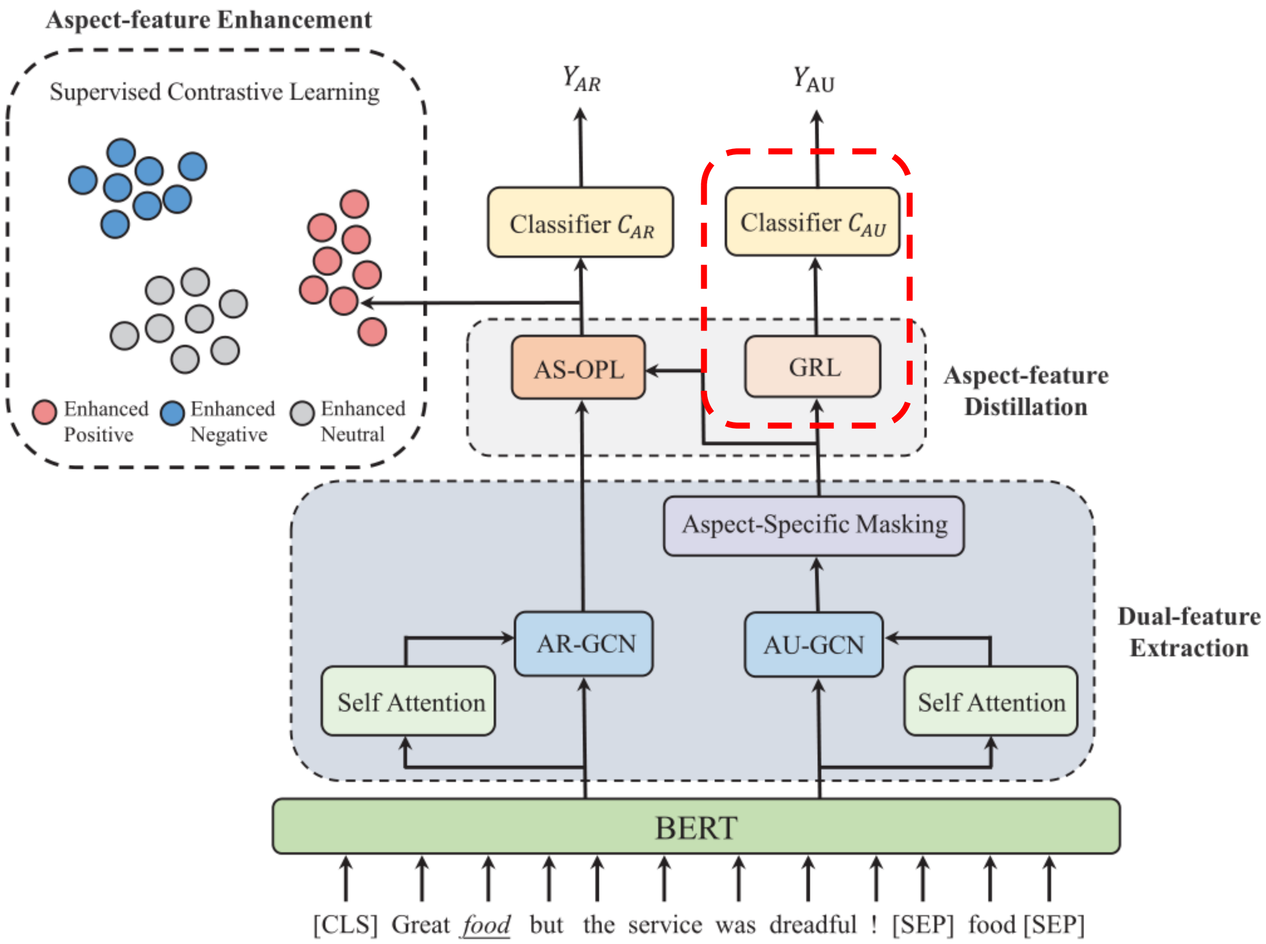
$$A = softmax(\frac{QW^Q \times (KW^K)^T}{\sqrt{d}}) \qquad (1)$$

$$h_i^l = \sigma(\sum_{j=1}^{n} A_{ij} W^l h_j^{l-1} + b^l) \qquad (2)$$

$$h_t^L = \begin{cases} 0 & 1 \le t < a+1, \ a+m < t \le n \\ h_t^L & a+1 \le t \le a+m \end{cases} \qquad (3)$$

$$H_{mask}^L = \{0, \ \cdots, \ h_{a+1}^L, \ \cdots, \ h_{a+m}^L, \ \cdots, \ 0\}.$$

Approach



$$GRL(x) = x \tag{4}$$

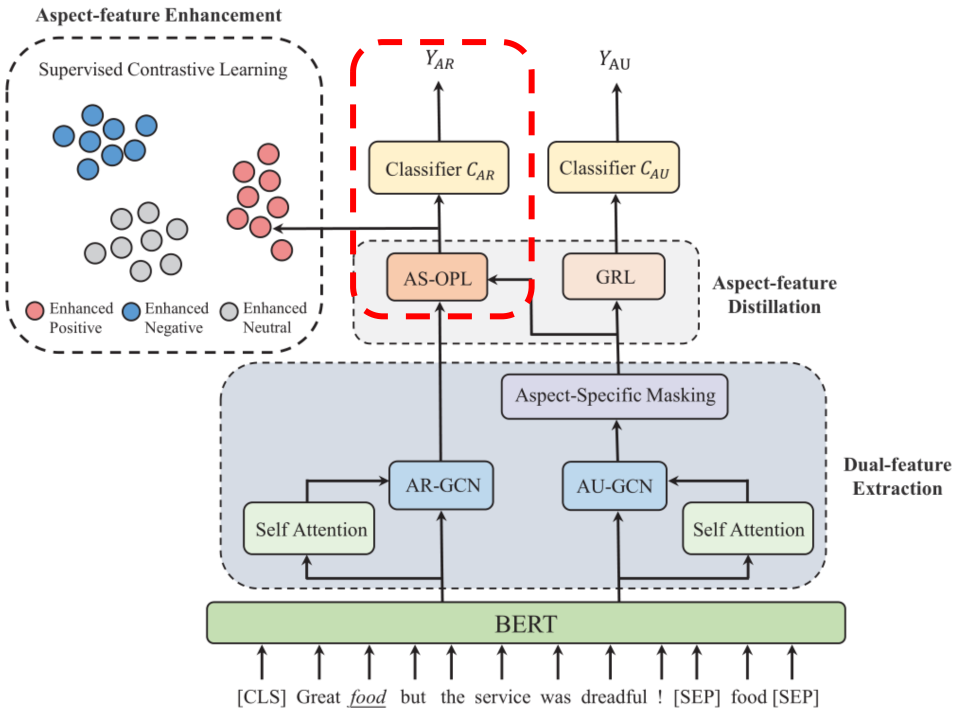
$$\frac{\partial GRL(x)}{\partial x} = -\lambda I \tag{5}$$

$$Y_{AU} = softmax(\widetilde{f_{AU}} \cdot W_{AU} + b_{AU}) \tag{6}$$

$$\mathcal{L}_{AU} = CrossEntropy(y_{truth}, y_{AU}) \tag{7}$$

the output $\widetilde{f_{AU}}$ of GRL is sent to the classifier C_{AU} to obtain the prediction result:

Approach



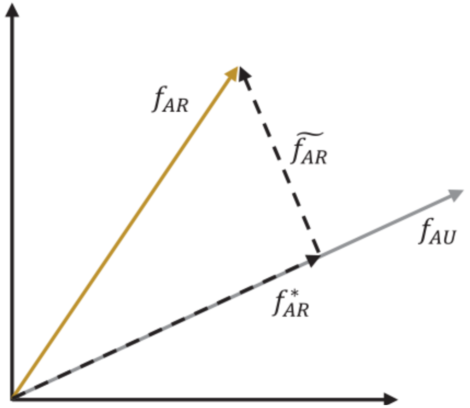


Figure 2: The principle of AS-OPL.

$$f_{AR}^{*} = Proj(f_{AR}, f_{AU}) \tag{8}$$

$$Proj(x, y) = \frac{x \cdot y}{|y|}\frac{y}{|y|} \tag{9}$$

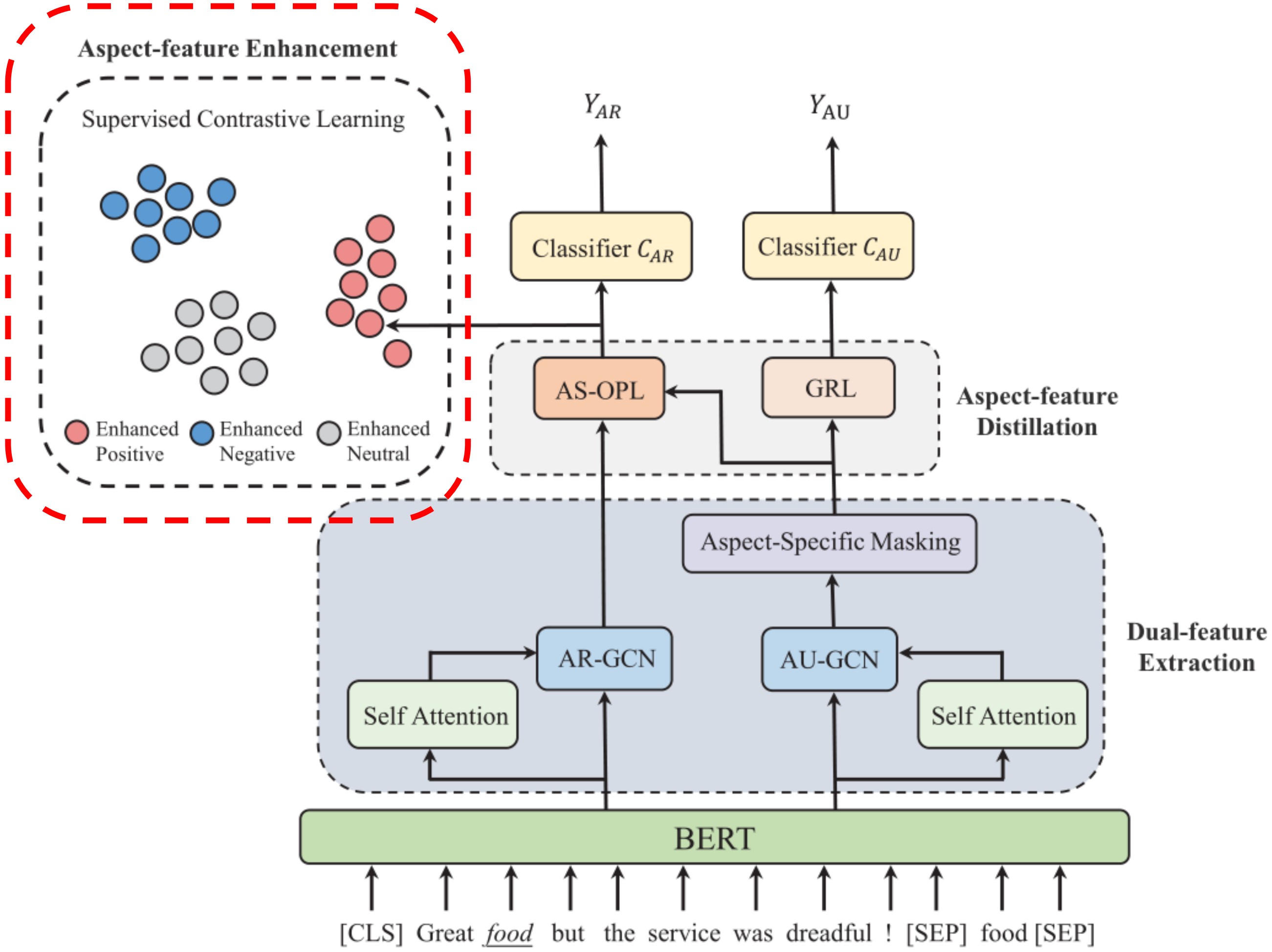
$$\widetilde{f_{AR}} = Proj(f_{AR}, (f_{AR} - f_{AR}^{*})) \tag{10}$$

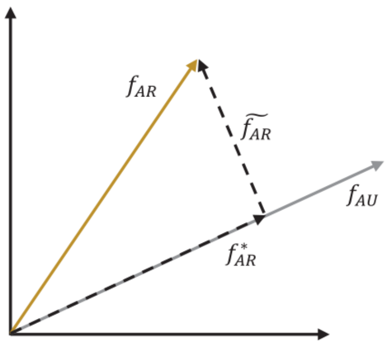
$$Y_{AR} = softmax(\widetilde{f_{AR}} \cdot W_{AR} + b_{AR}) \tag{11}$$

$$\mathcal{L}_{AR} = CrossEntropy(y_{truth}, y_{AR}) \tag{12}$$

- |x| · cosΘ称为向量x在向量y上的投影
- x · y= |x| · |y| · cosΘ
- |x| · cosΘ=$\frac{x \cdot y}{|y|}$

Approach





$$z_i = \widetilde{f_{AR}}$$

$$P_B^{sup}(i,c) = \frac{exp(sim(z_i, z_c)/\tau)}{\sum_{b\in B, b\neq i} exp(sim(z_i, z_b)/\tau)} \qquad (13)$$

$$\mathcal{L}_B^{sup} = \sum_{i\in B} -\log \frac{1}{C_i} \sum_{y_i=y_c, c\neq i} P_B^{sup}(i,c) \qquad (14)$$

Experiments

Table 1: Statistics on four datasets of ABSA.

Dataset	Division	#Positive	#Negative	#Neutral
Restaurant	Train	2164	807	637
	Test	728	196	196
Laptop	Train	994	870	464
	Test	341	128	169
Twitter	Train	1561	1560	3127
	Test	173	173	346
MAMS	Train	3380	2764	5042
	Test	400	329	607

Experiments

Table 2: Experimental results comparison on three publicly available datasets.

Models	Rest14		Lap14		Twitter	
	Accuracy	Macro-F1	Accuracy	Macro-F1	Accuracy	Macro-F1
BERT-SPC	84.46	76.98	78.99	75.03	73.55	72.14
AEN+BERT	83.12	73.76	79.93	76.31	74.71	73.13
BERT-PT	84.95	76.96	78.07	75.08	-	-
TD-BERT	85.10	78.40	78.90	74.40	76.70	74.30
CapsNet+BERT	85.09	77.75	78.21	73.34	-	-
SDGCN-BERT	83.57	76.47	81.35	78.34	-	-
R-GAT+BERT	86.60	81.35	78.21	74.07	76.15	74.88
DGEDT+BERT	86.30	80.00	79.80	75.60	77.90	75.40
BERT-ADA	87.14	80.05	79.19	74.18	-	-
DualGCN+BERT	87.13	81.16	81.80	78.10	77.40	76.02
Our AFDEN	**87.41**	**82.21**	**82.13**	**78.81**	**78.47**	**77.27**

Experiments

Table 3: Experimental results of ablation study

Models	Rest14		Lap14		Twitter	
	Accuracy	Macro-F1	Accuracy	Macro-F1	Accuracy	Macro-F1
AFDEN w/o AFE	86.16	80.14	79.62	76.19	75.72	73.75
AFDEN w/o DFE	86.16	80.25	78.68	74.71	75.87	74.82
AFDEN w/o AFD	86.07	79.67	79.93	76.34	76.30	74.99
AFDEN	**87.41**	**82.21**	**82.13**	**78.81**	**78.47**	**77.27**

Experiments

Table 4: Case studies of our AFDEN model compared with other baselines.

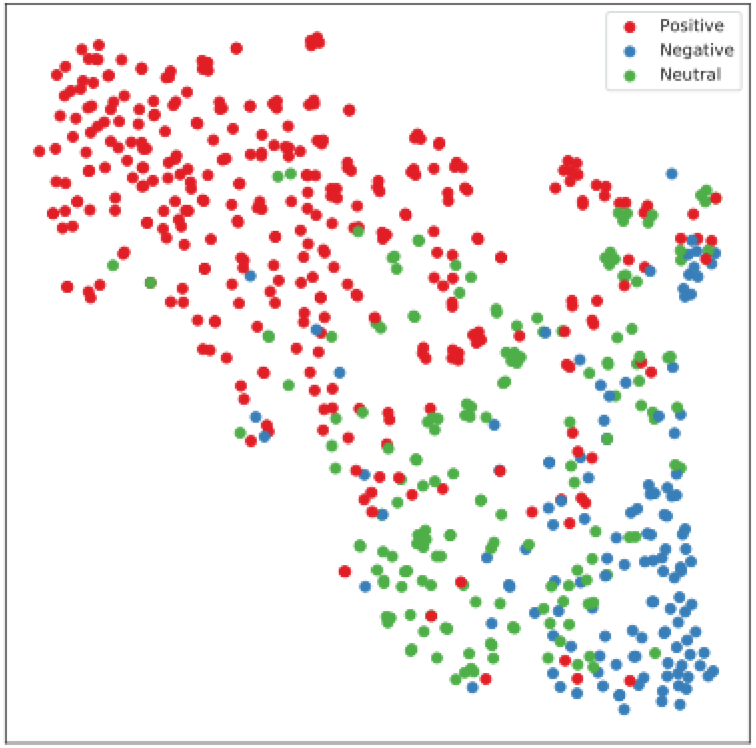
#	Review	AEN+BERT	BERT-SPC	DualGCN+BERT	AFDEN
1	They are served on focacchia bread and are to die for .	($P_{\times}$, $P_{\times}$)	($P_{\times}$, $P_{\times}$)	($P_{\times}$, $P_{\times}$)	($O_{\checkmark}$, $O_{\checkmark}$)
2	Great beer selection too , something like 50 beers .	$P_{\times}$	$P_{\times}$	$P_{\times}$	$O_{\checkmark}$
3	I do not like too much windows 8 .	$P_{\times}$	$P_{\times}$	$P_{\times}$	$N_{\checkmark}$
4	A beautiful atmosphere , perfect for drinks and / or appetizers .	($P_{\times}$, $P_{\times}$)	($P_{\times}$, $P_{\times}$)	($P_{\times}$, $P_{\times}$)	($P_{\times}$, $O_{\checkmark}$)
5	It's good to go there for drinks if you don't want to get drunk because you'll be lucky if you can get one drink an hour .	($N_{\times}$, $N_{\times}$)	($P_{\times}$, $P_{\times}$)	($P_{\times}$, $O_{\checkmark}$)	($O_{\checkmark}$, $O_{\checkmark}$)

Experiments

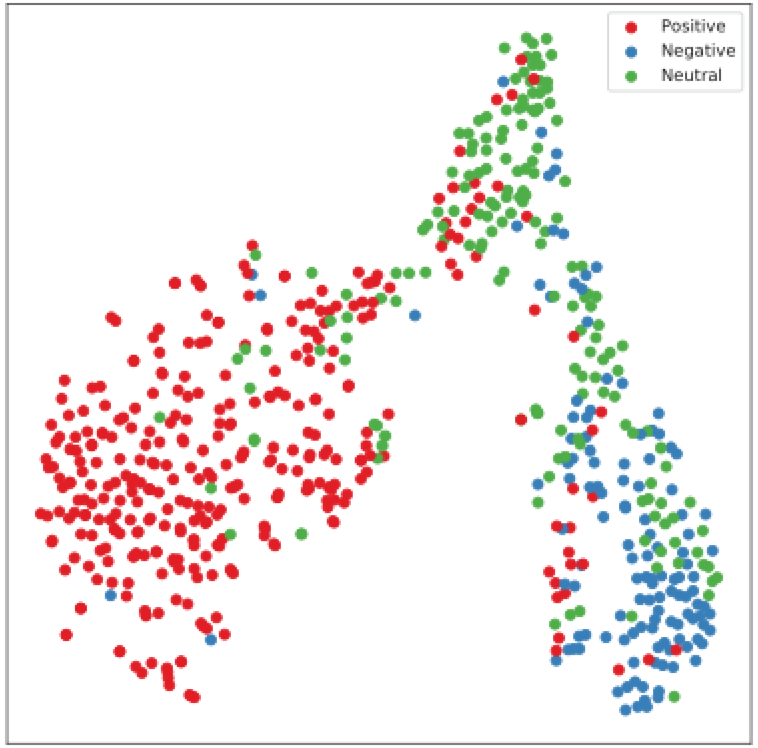
Table 5: Model performance on Aspect Robustness Test Set (ARTS). We compare the model accuracy on the original and new testsets, and calculate the accuracy decline of prediction between them.

Models	Restaurant-ARTS		Laptop-ARTS	
	Ori→New	Decline	Ori→New	Decline
AEN+BERT	83.12→25.45	-57.67	79.93→30.09	-49.84
BERT-SPC	83.04→54.82	-29.22	77.59→50.94	-26.65
CapsNet+BERT	83.48→55.36	-28.12	77.12→25.86	-51.46
BERT-PT	86.70→59.29	-27.41	78.53→53.29	-25.24
DualGCN+BERT	87.13→63.57	-23.56	81.80→57.99	-23.81
AFDEN	**87.41→65.18**	**-22.23**	**82.13→59.87**	**-22.26**

Experiments



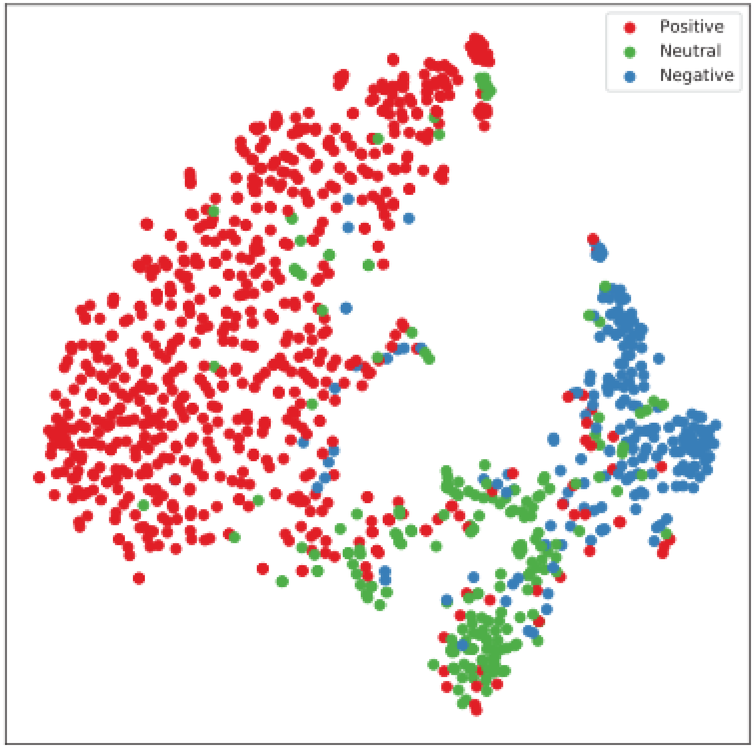
(a) AFDEN w/o AFE



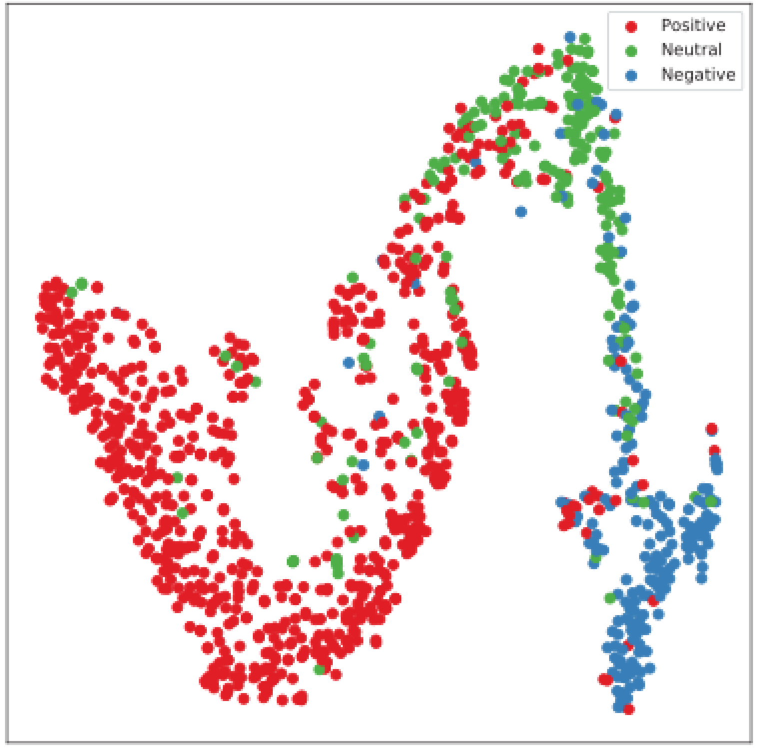
(b) Our AFDEN

Figure 3: The visualization of aspect-related embeddings on Laptop dataset.

Experiments



(a) AFDEN w/o AFE



(b) Our AFDEN

Figure 4: The visualization of aspect-related embeddings on Restaurant dataset.

Experiments

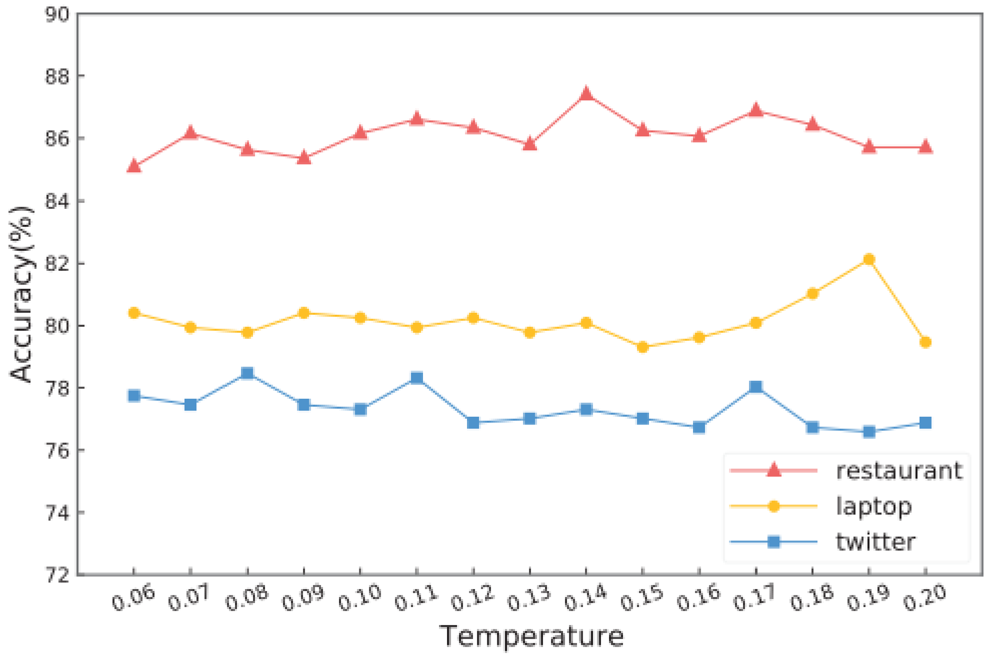


Figure 5: Effect of different temperatures on three datasets.

Experiments

Table 6: Model performance on MAMS

Models	MAMS	
	Accuracy	Macro-F1
AEN	66.72	-
CapsNet	79.78	-
AEN+BERT	72.08	71.46
BERT-SPC	82.22	-
CapsNet+BERT	83.39	-
AFDEN	**85.33**	**84.73**

Thank you !