Supplementary Materials to "Triple Dimensional Valence-Arousal-Dominance Encouraging Graph Attention Networks to Exploit Aspect-based Sentiment Analysis"

Xuefeng Shi[1], Weiping Ding[1*], Min Hu[2,3], Xin Kang[4], Fuji Ren[5,6*]
1 School of Computer Science and Artificial Intelligence, Nantong University, Seyuan Road, Nantong, 226019, Jiangsu, China.
2 School of Computer and Information, Hefei University of Technology, Danxia Road, Hefei, 230601, Anhui, China.
3 Anhui Province Key Laboratory of Affective, Hefei University of Technology, Danxia Road, Hefei, 230601, Anhui, China.
4 Department of Computer Science, Tokushima University, 2-24 Shin kura-cho, Tokushima, 770-8501, Tokushima, Japan.
5 Shenzhen Institute for Advanced Study, University of Electronic Science and Technology of China, Guanlan Road, Shenzhen, 518000, Guangdong, China.
6 School of Computer Science and Engineering, University of Electronic Science and Technology of China, Xiyuan Avenue, Chengdu, 611731, Sichuan, China.

*Corresponding author(s). E-mail(s): ding.wp@ntu.edu.cn; renfuji@uestc.edu.cn;
Contributing authors: xuefeng shi@ntu.edu.cn; jsjxhumin@hfut.edu.cn;
kang-xin@is.tokushima-u.ac.jp;

In this supplementary file, we provide:

1. The Impact of Dropout;
2. The Impact of Learning Rate;
3. The Impact of Batch Size;
4. Robust Analysis;

1 The Impact of Dropout

Initially, the dropout is devised to mitigate the pressure caused by over-fitting while learning the material, which can reduce the co-adaptability between neurons and increase the generalization ability of the model. In this work, this subsection is organized to analyze the influence caused by the dropout in the fine-tuning of VADGAT, and the results are shown in Figure 1. In the figure, this phenomenon is apparent that each dataset should be learned with different dropout values in the training process of VADGAT. Concretely, while the dropout is set to 0.4, VADGAT achieves the optimal result on Lap14. Additionally, it is noted that the performances of the other four datasets (Rest14, Rest15, Rest16, and Twitter) are improved largely with the dropout 0.1 in the current state, which suggests that discarding the appropriate proportion of the textual information can promote the robustness of VADGAT effectively.

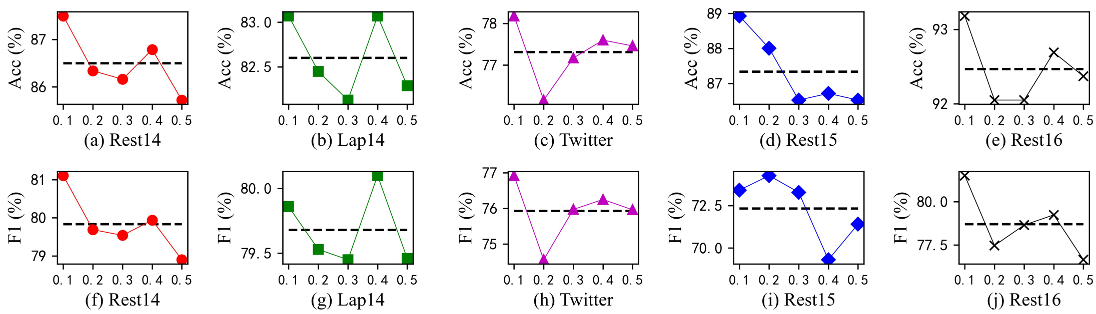


Fig.1 The impact of dropout.

2 The Impact of Learning Rate

The appropriate learning rate is the key hyper-parameter to learning a successful ABSA model, which has been proven significant in fine-tuning the PLM for downstream tasks. In this subsection, this work also provides a relevant study to attempt to explain the significance of the learning rate to the proposed VADGAT. Figure 2 is depicted to show the experimental results achieved via six learning rates ({1e-5, 2e-5, 3e-5, 4e-5, 5e-5, 6e-5}). From the figure, it is obvious that VADGAT achieves the optimal results on the five mentioned benchmarks simultaneously, while the learning rate is set to 1e-5. Besides, the larger the learning rate is, the lower the result is achieved. This finding indicates that VADGAT can be fine-tuned to the optimal state for extracting the sentiment features, which implies the fine-grained features can optimize the restored parameters in the PLM.

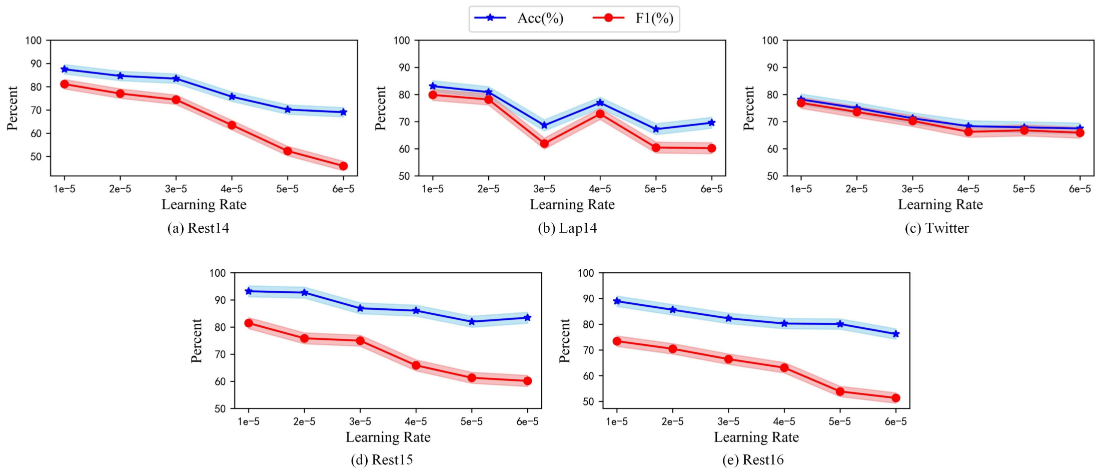
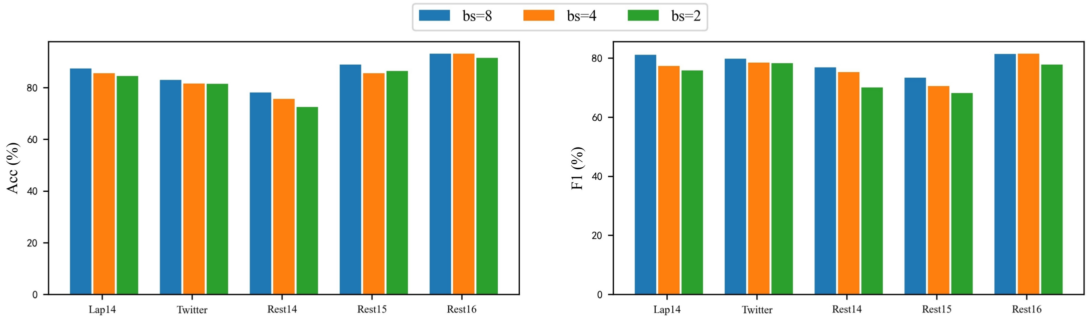


Fig.2 The impact of learning rate.

3 The Impact of Batch Size

Batch size means the granularity of the input data to train the deep learning model, which impacts the learning consequence largely. In this subsection, this work discusses the influence caused by batch size on the performance of VADGAT in ABSA. For intuitive observation, the related results are depicted in Figure 3. From the figure, it can be learned that this work mainly conducts the relevant experiments with three groups of batch size, which are 2, 4, and 8, respectively. In the scope of investigation, while the batch size is set to 8, VADGAT could learn abundant sentiment information from the input textual materials, and this is the reason that batch size is set to 8 for further validation study in this work.



(a) Accuracy on five benchmarks with different batch sizes. (b) F1 on five benchmarks with different batch sizes.

Fig. 3 The impact of batch size.

4 Robust Analysis

Furthermore, to investigate the generalization of the proposed approach VADGAT, this work conducts the corresponding experiments with different contextual representations obtained by different PLMs, which are RoBERTa, BERT, GPT, and BART, respectively. For convenient comparison, the related experimental results are reported in Table 1. From the table, it is obvious that RoBERTa can encourage VADGAT's capability to recognize the sentence's sentiment polarity, which is embodied in the five mentioned benchmarks clearly. And, the experimental results based on BERT, GPT, and BART are lower than the ones achieved on RoBERTa. On the other perspective, the generalization of VADGAT is still limited to the special PLM, which denotes the robustness of this work should be studied further in the future. Additionally, this work believes that the manually designed prompt template also limits the generalization of VADGAT in fact. Conclusively, this analysis shows that VADGAT based on RoBERTa can understand the contextual sentiment information comprehensively and capture the key features in predicting the sentiment polarity.

Table 4. The generalization ability of VADGAT.

PLM	Lap14		Twitter		Rest14		Rest15		Rest16	
	Acc(%)	F1(%)	Acc(%)	F1(%)	Acc(%)	F1(%)	Acc(%)	F1(%)	Acc(%)	F1(%)
RoBERTa	**83.07**	**80.10**	**78.18**	**76.92**	**87.50**	**81.11**	**88.93**	**73.42**	**93.18**	**81.51**
BERT	77.43	72.99	74.57	72.57	84.02	77.08	82.47	72.63	89.77	76.12
GPT	79.78	76.34	72.11	69.48	82.14	73.39	83.03	65.66	88.96	73.02
BART	79.94	76.31	71.79	71.58	83.13	73.83	84.87	70.01	90.26	73.24

Note: The best result is in **bold**.