YNU-HPCC at SIGHAN-2024 dimABSA Task: Using PLMs with a Joint Learning Strategy for Dimensional Intensity Prediction

Zehui Wang, You Zhang*, Jin Wang, Dan Xu, and Xuejie Zhang
School of Information Science and Engineering
Yunnan University
Kunming, China
Contact: `wangzehui@stu.ynu.edu.cn`, `yzhang0202@ynu.edu.cn`

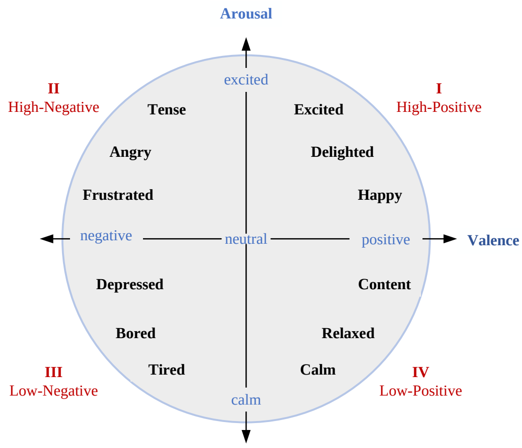
Abstract

The dimensional approach can represent more fine-grained emotional information than discrete affective states. In this paper, a pre-trained language model (PLM) with a joint learning strategy is proposed for the SIGHAN-2024 shared task on Chinese dimensional aspect-based sentiment analysis (dimABSA), which requires submitted models to provide fine-grained multi-dimensional (Valence and Arousal) intensity predictions for given aspects of a review. The proposed model consists of three parts: an input layer that concatenates both given aspect terms and input sentences; a Chinese PLM encoder that generates aspect-specific review representation; and separate linear predictors that jointly predict Valence and Arousal sentiment intensities. Moreover, we merge simplified and traditional Chinese training data for data augmentation. Our system ranked 2nd place out of 5 participants in subtask 1-intensity prediction. The code is publicly available at https://github.com/WZH5127/2024_subtask1_intensity_prediction.

1 Introduction

Aspect-based sentiment analysis (ABSA) (Pontiki et al., 2014, 2015, 2016) is used to identify the sentiment polarity regarding specific aspects within a sentence. In recent years, ABSA tasks have gradually extended into diverse subtasks, including aspect sentiment triplet extraction (ASTE) (Chen et al., 2022; Xu et al., 2021; Zhao et al., 2022; Zhang et al., 2023) and aspect sentiment quadruple prediction (ASQP) (Hu et al., 2022a; Wang et al., 2023; Zhou et al., 2023; Zhang et al., 2021). In contrast to these tasks, which consider affective states as discrete classes (positive, neutral, and negative), the dimensional approach provides more fine-grained emotional information (Lee et al., 2022).

Dimensional sentiment analysis represents affective states as continuous numerical values in multiple dimensions, such as Valence and Arousal space (Yu et al., 2016), as illustrated in Figure 1(a). The Valence dimension indicates the degree of positive or negative sentiments, while the Arousal dimension refers to the degree of calmness or excitement. Valence and Arousal are represented by continuous real-valued scores ranging from 1 to 9, with lower scores indicating stronger negative or calm sentiments, higher scores indicating stronger positive or excited sentiments, and mid-range scores, such as 5, indicating neutral states. Combining aspect-based and multi-dimensional sentiment analysis, a shared task of Chinese dimensional ABSA shared task (dimABSA) (Lee et al., 2024) is proposed in SIGHAN-2024, which primarily includes



(a) Valence and Arousal Spaces

(b) An example of intensity prediction

Figure 1: The diagram of Valence and Arousal space and dimABSA.

*Corresponding author.

Proceedings of the 10th SIGHAN Workshop on Chinese Language Processing (SIGHAN-10), pages 96–101

intensity prediction, triplet extraction, and quadruple extraction. In subtask 1 intensity prediction, given a sentence and an aspect, the system is required to predict the Valence and Arousal intensities of the sentence regarding the aspect. For instance, in the sentence shown in Figure 1(b), there are two aspects, “柠檬酱”(lemon sauce) and “塔皮”(tart crust), with required two-dimensional (Valence#Arousal) intensity predictions of 5.67#5.5 and 4.83#5.0, respective.

Recently, pre-trained language models (PLMs) (Devlin et al., 2018; Li et al., 2020; Hu et al., 2022b) have achieved significant success in various natural language processing (NLP) tasks, including sentiment analysis. For instance, using Chinese-based PLMs such as BERT-base-Chinese and BERT-wwm-ext (Cui et al., 2021) for intensity prediction on Chinese EmoBank (Lee et al., 2022) has yielded superior results compared to traditional methods. However, when these methods tackle dimABSA tasks, they continue to encounter challenges in (1) integrating traditional and simplified Chinese for robust review representation and (2) capturing internal relatedness across multiple dimensions for a comprehensive understanding of semantics.

To address these issues, we utilized whole-word masking (wwm) (Cui et al., 2021; Pandey et al., 2022) PLM of BERT-wwm-ext with a joint learning strategy for dimensional intensity prediction in ABSA. The model consists of the input layer, PLM encoder, and dimensional linear layer (dimLinear). Initially, we concatenate one aspect term and the review sentence as model sequence input. Then, we utilize BERT-wwm-ext as the PLM encoder to generate robust text representation. Finally, the dimLinear layer contains separate linear predictors that jointly predict Valence and Arousal sentiment intensities. Moreover, we merge traditional and simplified Chinese training samples into an augmented training set for generalized optimization. In experiments, we found that the integrations between two types of Chinese corpora and the joint optimization of multiple dimensions resulted in better performance. Consequently, our team ranked 2nd out of 5 participants in subtask 1 of the shared dimABSA task.

The remainder of this paper is structured as follows: Section 2 describes the architecture of our model in detail. Section 3 presents extensive experiments, analysis, and results. Finally, conclusions and future work are discussed in Section 4.

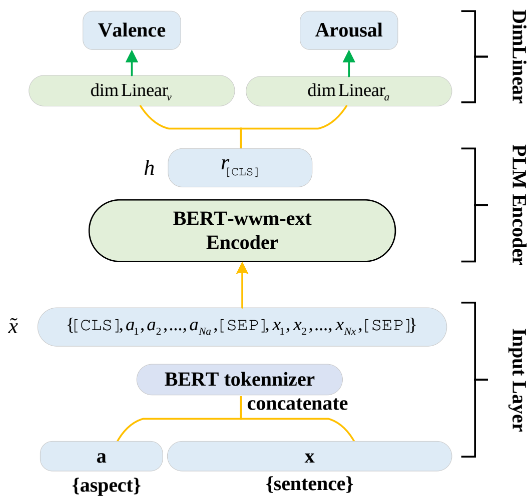


Figure 2: The overview of model architecture.

2 System Description

In this section, we primarily describe the architecture of our model. As depicted in Figure 2, the model comprises three main components: input layer, PLM encoder, and dimLinear layer.

2.1 Input Layer

Both an aspect $\mathbf{a}$ and a review text $\mathbf{x}$ are first tokenized into discrete tokens, denoted by $\mathbf{a} = \{a_1, a_2, ..., a_{Na}\}$ and $\mathbf{x} = \{x_1, x_2, ..., x_{Nx}\}$, respectively, where Na and Nx represent the length of the aspect and the review. To feed both the aspect and the review into models, we concatenate the aspect and the review tokens, denoted as $\tilde{\mathbf{x}} = \{\texttt{[CLS]}, \mathbf{a}, \texttt{[SEP]}, \mathbf{x}, \texttt{[SEP]}\}$, where `[CLS]` and `[SEP]` are special tokens for syntactic separation.

2.2 Chinese PLM Encoder

To learn the hidden aspect-specific review representation h, we use the Chinese PLM of BERT-wwm-ext to encode the concatenated input sequence formally:

$$\mathbf{r} = f(\tilde{\mathbf{x}}; \theta_{\mathrm{BERT-wwm-ext}}) \in \mathbb{R}^{N \times d} \quad (1)$$

where $f(\cdot)$ represents the encoder propagation; $\theta_{\mathrm{BERT-wwm-ext}}$ is the trainable parameters initialized from a pre-trained checkpoint and fine-tuned for the specific task; and $N = (Na + Nx + 3)$ and d indicates the input length and hidden dimensionality, respectively.

Similar to BERT families, we take as the final review representation the first token representation, i.e., $h = r_{\texttt{[CLS]}} \in \mathbb{R}^d$.

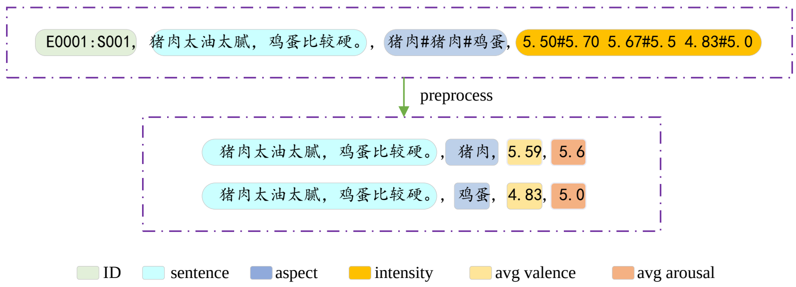


Figure 3: An example of data preprocessing.

2.3 Dimensional Intensity Prediction

To predict the intensities of Valence $v \in \mathbb{R}^1$ and Arousal $a \in \mathbb{R}^1$, we use the dimLinear that contains two linear projections to estimate $p(v|h)$ and $p(a|h)$ simultaneously. This multi-task learning strategy could facilitate the model in learning robust review representation. The prediction is as follows:

$$\begin{aligned} v &= \dim \mathrm{Linear}_v(h) \\ a &= \dim \mathrm{Linear}_a(h) \end{aligned} \tag{2}$$

where each $\mathrm{dimLinear\_}(\cdot)$ is implemented via two stacked fully connected layers.

During model optimization, we employ mean absolute error (MAE) as the cost function to maximize the likelihood of model performance $p_\theta(v, a|\mathbf{a}, \mathbf{x})$ in an end-to-end manner.

3 Experimental Results

In this section, we present the comparative results of the proposed methods.

3.1 Datasets

Throughout the competition, we utilized datasets exclusively provided by the organizers of the shared dimABSA task. These datasets were formally partitioned into Train, Dev, and Test sets. Since golden labels were not provided for participants, an additional Dev* set was created by randomly sampling 10% of the Train sample to aid in model selections.

Furthermore, the dataset encompassed traditional and simplified Chinese versions, the only difference being the language. To enhance the generalization performance of our models, we augmented the Train set by integrating both versions. For more detailed statistics on the datasets, please refer to Table 1.

Dataset	# samples	Max length
Train	6050	56
Merged-Train	12100	56
Dev	100	40
Dev*	1210	46
Test	2000	59

Table 1: Detailed statistics of the datasets.

3.2 Evaluation Metrics

To evaluate the performance of participant systems for Subtask 1, the organizers furnished MAE and the Pearson Correlation Coefficient (PCC) as evaluation metrics.

- MAE

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n} |\hat{p}_i - p_i| \tag{3}$$

where $\hat{p}_i$ and p_i respectively denoted the ith actual value and predicted value, n is the number of test samples.

- PCC

$$\mathrm{PCC} = \frac{1}{n-1}\sum_{i=1}^{n} (\frac{\hat{p}_i - \hat{\mu}}{\hat{\sigma}})(\frac{p_i - \mu_p}{\sigma_p}) \tag{4}$$

where $\hat{\mu}$ and $\hat{\sigma}$ respectively represent the mean value and the standard deviation of all predictions, while μ_p and σ_p respectively represent that of golden labels. A lower MAE and a higher PCC indicate more accurate prediction performance.

3.3 Implementation Details

Data preprocessing. We observed that certain samples in the Train set contained multiple intensity values for a single aspect, reflecting different opinions, as illustrated in Figure 3. As Subtask 1 only concerns the overall sentiment towards an aspect in a review, we computed the average intensity across various opinions to derive an overall sentiment score.

Model	Dev*				Test			
	Valence		Arousal		Valence		Arousal	
	MAE↓	PCC↑	MAE↓	PCC↑	MAE↓	PCC↑	MAE↓	PCC↑
BERT-base-chinese	0.222	0.951	0.280	0.817	0.299	0.911	**0.318**	0.767
BERT-wwm	**0.221**	**0.956**	**0.277**	**0.822**	0.298	0.912	0.319	0.766
RoBERTa-wwm-ext	0.241	0.954	0.294	0.808	0.306	0.913	0.327	0.766
MacBERT-wwm-ext	0.253	0.936	0.296	0.803	0.312	0.905	0.327	0.761
BERT-wwm-ext (Ranked 2nd)	0.227	0.952	0.284	0.814	**0.294**	**0.917**	**0.318**	**0.771**
BERT-wwm-ext†	0.247	0.948	0.283	0.803	0.311	0.910	0.323	0.748
BERT-wwm-ext‡	0.249	0.934	0.300	0.787	0.311	0.904	**0.318**	0.760

Table 2: Comparative Dev* and Test results for subtask 1. **Bold figures** meant the best performance regarding various metrics.

Hyperparameters. The maximum length of the longest sentence in a batch sample was the maximum. We employed the base version of BERT-wwm-ext as the backbone model. Specifically, the model consisted of 12 transformer layers with a hidden representation dimensionality (d) of 768 (Vaswani et al., 2017). For optimization, we utilized the Adam optimizer with a linear warmup schedule. The base learning rate was set to 3e-5, with a batch size of 32.

Baselines. We implemented several baseline models to evaluate the performance of BERT-wwm-ext in dimensional prediction. Initially, we employed various PLMs as our backbones, including BERT-base-chinese, BERT-wwm, RoBERTa-wwm-ext, and MacBERT-wwm-ext (Cui et al., 2021). Subsequently, we introduce two variants of models: (1) BERT-wwm-ext†, which independently predicted Valence and Arousal intensities, and (2) BERT-wwm-ext‡, trained solely on the traditional Chinese-based Train set.

3.4 Result and Analysis

As depicted in Table 2, our proposed system's comparable Dev* and Test results against several baselines in terms of MAE and PCC were reported. With different PLMs as backbones, models achieved varying performances. In contrast to RoBERTa and MacBERT, Chinese-based BERT families achieved relatively lower MAE and higher PCC. This performance discrepancy could be attributed to the utilization of a next-sentence prediction task during the pretraining phase of BERT PLMs. This task's alignment with our input structure, which combines aspect and review texts, likely facilitated the model in better understanding the relationship between aspects and sentences during the fine-tuning phase.

In our settings, BERT-wwm performed relatively better than BERT-Chinese in Dev* and on par in Test. This is because the adopted wwm strategy could enhance Chinese sentence modeling. In contrast to BERT-wwm, BERT-wwm-ext leveraged a larger pretraining corpus to acquire more comprehensive language knowledge, demonstrating better generalization in the Test set.

Furthermore, we conducted ablation studies to examine the effect of our joint learning strategy. Without the joint optimization of Valence and Arousal dimensions, the performance of BERT-wwm-ext† degraded, and so did BERT-wwm-ext‡ when only the traditional Chinese version was utilized as the Train set. These phenomena underscored the effectiveness of our joint learning strategy in facilitating robust aspect-specific review representation.

In conclusion, our proposed methods attained the best Test scores in both MAE and PCC metrics, which ranked 2nd place out of 5 participants. This result highlights the competitiveness and effectiveness of our system in the shared subtask 1.

4 Conclusions

This paper presented our system developed for the SIGHAN-2024 shared Task dimABSA. Our experimental results in subtask 1 demonstrated that our proposed model achieved significant performance in the dimensional intensity prediction of ABSA. As a result, our team ranked 2nd place in subtask 1.

Future works will explore incorporating our model with the existing dimensional sentiment analysis corpus and investigating a unified model handling multiple targets in dimABSA.

Limitations

Although our proposed system enhances the quality of dimensional intensity prediction, there are still several limitations. First, we have only validated the effectiveness of the joint learning strategy for intensity prediction and have not tested it for fine-grained aspect extraction. Second, our current approach uses traditional PLMs as backbones. In the future, we plan to explore the use of large-scale PLMs, such as ChatGPT and LLaMA, for dimABSA tasks.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (NSFC) under Grant Nos. 61966038 and 62266051, and Caiyun Postdoctoral Foundation under Grant No. C615300504090.

References

Yuqi Chen, Chen Keming, Xian Sun, and Zequn Zhang. 2022. A Span-level Bidirectional Network for Aspect Sentiment Triplet Extraction. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, pages 4300–4309, Abu Dhabi, United Arab Emirates. Association for Computational Linguistics.

Yiming Cui, Wanxiang Che, Ting Liu, Bing Qin, and Ziqing Yang. 2021. Pre-Training with Whole Word Masking for Chinese BERT. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 29:3504–3514.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. Bert: Pre-training of Deep Bidirectional Transformers for Language Understanding. *arXiv preprint arXiv:1810.04805*.

Mengting Hu, Yike Wu, Hang Gao, Yinhao Bai, and Shiwan Zhao. 2022a. Improving Aspect Sentiment Quad Prediction via Template-Order Data Augmentation. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, pages 7889–7900, Abu Dhabi, United Arab Emirates. Association for Computational Linguistics.

Yibo Hu, MohammadSaleh Hosseini, Erick Skorupa Parolin, Javier Osorio, Latifur Khan, Patrick Brandt, and Vito D'Orazio. 2022b. ConfliBERT: A Pre-trained Language Model for Political Conflict and Violence. In *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 5469–5482, Seattle, United States. Association for Computational Linguistics.

Lung-Hao Lee, Jian-Hong Li, and Liang-Chih Yu. 2022. Chinese EmoBank: Building Valence-Arousal Resources for Dimensional Sentiment Analysis. *Transactions on Asian and Low-Resource Language Information Processing*, 21(4):1–18.

Lung-Hao Lee, Liang-Chih Yu, Suge Wang, and Jian Liao. 2024. Overview of the SIGHAN 2024 shared task for Chinese dimensional aspect-based sentiment analysis. In *Proceedings of the 10th SIGHAN Workshop on Chinese Language Processing*. Association for Computational Linguistics.

Bohan Li, Hao Zhou, Junxian He, Mingxuan Wang, Yiming Yang, and Lei Li. 2020. On the Sentence Embeddings from Pre-trained Language Models. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 9119–9130, Online. Association for Computational Linguistics.

Amit Pandey, Swayatta Daw, Narendra Unnam, and Vikram Pudi. 2022. Multilinguals at SemEval-2022 Task 11: Complex NER in Semantically Ambiguous Settings for Low Resource Languages. In *Proceedings of the 16th International Workshop on Semantic Evaluation (SemEval-2022)*, pages 1469–1476, Seattle, United States. Association for Computational Linguistics.

Maria Pontiki, Dimitris Galanis, Haris Papageorgiou, Ion Androutsopoulos, Suresh Manandhar, Mohammad AL-Smadi, Mahmoud Al-Ayyoub, Yanyan Zhao, Bing Qin, Orphée De Clercq, Véronique Hoste, Marianna Apidianaki, Xavier Tannier, Natalia Loukachevitch, Evgeniy Kotelnikov, Nuria Bel, Salud María Jiménez-Zafra, and Gülşen Eryiğit. 2016. SemEval-2016 Task 5: Aspect Based Sentiment Analysis. In *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016)*, pages 19–30, San Diego, California. Association for Computational Linguistics.

Maria Pontiki, Dimitris Galanis, Haris Papageorgiou, Suresh Manandhar, and Ion Androutsopoulos. 2015. SemEval-2015 Task 12: Aspect Based Sentiment Analysis. In *Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015)*, pages 486–495, Denver, Colorado. Association for Computational Linguistics.

Maria Pontiki, Dimitris Galanis, John Pavlopoulos, Harris Papageorgiou, Ion Androutsopoulos, and Suresh Manandhar. 2014. SemEval-2014 Task 4: Aspect Based Sentiment Analysis. In *Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014)*, pages 27–35, Dublin, Ireland. Association for Computational Linguistics.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention Is All You Need. *Advances in neural information processing systems*, 30.

An Wang, Junfeng Jiang, Youmi Ma, Ao Liu, and Naoaki Okazaki. 2023. Generative Data Augmentation for Aspect Sentiment Quad Prediction. In *Proceedings of the 12th Joint Conference on Lexical and Computational Semantics (*SEM 2023)*, pages 128–140, Toronto, Canada. Association for Computational Linguistics.

Lu Xu, Yew Ken Chia, and Lidong Bing. 2021. Learning Span-Level Interactions for Aspect Sentiment Triplet Extraction. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 4755–4766, Online. Association for Computational Linguistics.

Liang-Chih Yu, Lung-Hao Lee, Shuai Hao, Jin Wang, Yunchao He, Jun Hu, K. Robert Lai, and Xuejie Zhang. 2016. Building Chinese Affective Resources in Valence-Arousal Dimensions. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 540–545, San Diego, California. Association for Computational Linguistics.

Wenxuan Zhang, Yang Deng, Xin Li, Yifei Yuan, Lidong Bing, and Wai Lam. 2021. Aspect Sentiment Quad Prediction as Paraphrase Generation. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 9209–9219, Online and Punta Cana, Dominican Republic. Association for Computational Linguistics.

Yice Zhang, Yifan Yang, Meng Li, Bin Liang, Shiwei Chen, and Ruifeng Xu. 2023. Target-to-Source Augmentation for Aspect Sentiment Triplet Extraction. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pages 12165–12177, Singapore. Association for Computational Linguistics.

Yichun Zhao, Kui Meng, Gongshen Liu, Jintao Du, and Huijia Zhu. 2022. A Multi-Task Dual-Tree Network for Aspect Sentiment Triplet Extraction. In *Proceedings of the 29th International Conference on Computational Linguistics*, pages 7065–7074, Gyeongju, Republic of Korea. International Committee on Computational Linguistics.

Junxian Zhou, Haiqin Yang, Yuxuan He, Hao Mou, and Junbo Yang. 2023. A Unified One-Step Solution for Aspect Sentiment Quad Prediction. *arXiv preprint arXiv:2306.04152*.