SCUZANE at SemEval-2026 Task 3: Dimensional Aspect-based Sentiment Analysis with Supervised Contrastive Regression and R-Drop Regularization

Ziang Zhou[1], Xiangmei He[2], Chenhongyi Bai[3]
[1]College of Computer Science, Sichuan University
[2]College of Water Resources & Hydropower
[3]College of Electrical Engineering, Sichuan University
Correspondence: zhouzane86@gmail.com

Abstract

Current Aspect-Based Sentiment Analysis (ABSA) often relies on coarse-grained categorical labels, such as Positive and Negative, and this often leads to fail capturing the subtle intensity of emotional expression in real-world text. To address this issue, the SemEval-2026 Shared Task 3: Dimensional ABSA (DimABSA) extends the traditional ABSA by replacing categorical sentiment polarity with continuous valence-arousal (VA) scores. In this paper, we describe our system for Subtask 1 (Dimensional Aspect Sentiment Regression) of Track A (DimABSA). Our system utilizes a DeBERTa-v3-large backbone, enhanced by a prompt-based learning strategy that concatenates aspect information with the context. And we employ multi-sample dropout and a weighted aggregation of the hidden states from the last four layers to capture rich semantic representations. Our experimental results across all provided domains on different languages demonstrate the effectiveness of integrating consistency regularization with dimensional contrastive learning for fine-grained sentiment regression.

1 Introduction

Aspect-Based Sentiment Analysis (ABSA) has long been an important task in Natural Language Processing (NLP), providing a granular understanding of customers' opinions by identifying specific aspect terms and their associated sentiment polarities (Zhang et al., 2023). However, the traditional tri-partite classification (positive, negative, and neutral) is increasingly viewed as an oversimplification of human opinions. Drawing on foundational psychological models, such as Russell's circumplex model (Russell, 1980, 2003), affective states are more accurately represented as continuous coordinates in a multi-dimensional space.

Particularly, Valence (which measures the degree of positivity or negativity) and Arousal (which measures the intensity of emotion) provide a fine-grained, real-valued framework for capturing the nuances of emotional expression. The SemEval-2026 Task 3: Dimensional ABSA (DimABSA) (Yu et al., 2026) addresses this gap by shifting the ABSA paradigm from categorical classification to dimensional regression. This task requires participants to predict continuous VA scores within the range $[1.00, 9.00]$ for specific aspects across the domains Restaurant, Laptop, Hotel, and Finance. This change leads to significant challenges: participants must not only identify the relationship between an aspect and its context but also determine the sentiment intensity with high numerical precision. In this paper, we present our system designed for Subtask 1: Dimensional Aspect Sentiment Regression (DimASR) of Track A (DimABSA). Our system is built upon the DeBERTa-v3-large model (He et al., 2021), which leverages disentangled attention and ELECTRA-style pre-training to achieve superior linguistic representations. Based on the fact that regression tasks in VA space are highly sensitive to noise and data sparsity, we move beyond standard fine-tuning by incorporating three core architectural improvements: (1) Deep feature fusion: We aggregate the hidden states from the final four layers of the encoder, capturing a mixture of syntactic and high-level semantic information that is often lost when using only the last layer. (2) Consistency regularization via R-Drop: In order to stabilize the model's predictions and prevent overfitting, we employ a simple regularization strategy upon dropout in model training (R-Drop) (Liang et al., 2021), forcing the model to produce consistent outputs for the same input under different dropout masks. (3) Supervised dimensional contrastive learning: We introduce a contrastive objective specifically adapted for regression. By using the Euclidean distance between VA labels to weight the loss, we encourage the model to cluster features with similar emotional intensities in the latent

Proceedings of the The 20th International Workshop on Semantic Evaluation (2026), pages 1992–1996

space. Our system is evaluated using Root Mean Square Error (RMSE) and Pearson Correlation Coefficient (PCC). We demonstrate that the synergy of transformer-based prompts, multi-sample dropout, and correlation-aware loss functions (combining CCC and MSE) provides a robust solution for the regression of dimensional Aspect-Based sentiment analysis.

2 Related Work

2.1 Aspect-Based Sentiment Analysis (ABSA)

ABSA provides a coarse understanding of opinions by identifying specific targets and their associated sentiments (Zhang et al., 2023). Early research primarily addressed Aspect Sentiment Classification (ASC), categorizing sentiment into discrete polarities (Pontiki et al., 2014). Then ABSA evolved into more complex tasks such as Aspect Sentiment Triplet Extraction (ASTE) (Peng et al., 2020) and Aspect Sentiment Quad Prediction (ASQP) (Zhang et al., 2021). Despite their success, these tasks are limited by their categorical nature, which cannot distinguish between different intensities of the same sentiment category (e.g., "fine" vs. "great").

2.2 Dimensional Aspect-Based Sentiment Analysis (DimABSA)

The integration of dimensional sentiment into the ABSA framework is an emerging paradigm that aims to capture fine-grained emotional nuances. Unlike traditional ABSA, DimABSA reformulates the sentiment element as real-valued coordinates in the Valence-Arousal (VA) space (Russell, 1980, 2003). While dimensional sentiment analysis has been explored at the sentence or document level (Mohammad, 2018; Lee et al., 2022), DimABSA presents a unique challenge: a system must attribute specific intensity scores to multiple aspects within the same context. Recent works (Lee et al., 2024, 2026) have begun to establish benchmarks for this task, indicating that dimensional representations can better support subtle applications such as social network opinion analysis and sophisticated customer feedback systems.

2.3 Pre-trained Models and Consistency Regularization

Mainstream research in sentiment analysis is currently dominated by Pre-trained Language Models (PLMs). DeBERTa-v3 (He et al., 2021) is particularly notable for its disentangled attention mechanism, which improves the model's ability to relate aspects to their descriptors. To mitigate the instability inherent in regression tasks, techniques such as R-Drop (Liang et al., 2021) have been introduced to enforce consistency across different dropout sub-networks. Additionally, the adaptation of supervised contrastive learning for continuous labels has enabled models to learn more discriminative feature spaces by using the Euclidean distance between labels as a similarity metric, thereby moving beyond the limitations of standard point-wise loss functions.

3 Method

Our proposed system for the Dimensional Aspect Sentiment Regression (DimASR) subtask follows a unified regression framework. We transform the task into a sentence-pair regression problem by utilizing a prompt-based input and an ensemble of regularization techniques to handle the continuous nature of the Valence-Arousal (VA) space.

3.1 Model Architecture

We employ DeBERTa-v3-large (He et al., 2021) as the encoder backbone. DeBERTa-v3 improves upon traditional Transformers through disentangled attention, which separately encodes content and relative position, and Gradient-Disentangled Embedding Sharing via ELECTRA-style pre-training.To maximize the semantic richness of our representations, we do not rely solely on the final hidden state. Instead, we extract the hidden states of the last four layers (L_{21} to L_{24}) and perform weighted pooling on the $[CLS]$ tokens:

$$h_{fused} = \frac{1}{4} \sum_{i=21}^{24} [CLS]_i \tag{1}$$

This fused feature vector h_{fused} is then passed through a multi-sample dropout layer consisting of five independent dropout masks (with $p = 0.2$) and a shared linear regressor to produce the final output vector $\hat{y} = [V, A]$.

3.2 Consistency Regularization with R-Drop

Due to the high sensitivity of regression to minor feature perturbations, we introduce R-Drop (Liang et al., 2021). During training, each input batch is fed through the model twice. Due to the stochastic nature of the dropout layers, the model generates two slightly different predictions, $\hat{y}_1$ and $\hat{y}_2$, for the same input x. We minimize the Mean Squared

Error (MSE) between these two outputs to enforce consistency:

$$L_{rdrop} = \text{MSE}(\hat{y}_1, \hat{y}_2) \quad (2)$$

3.3 Supervised Dimensional Contrastive Loss

In order to better structure the latent feature space, we introduce a regression-aware contrastive loss. Unlike the traditional contrastive learning for classification, we use the Euclidean distance between gold labels in the VA space to weight sample similarity. For a pair of features $\{f_i, f_j\}$ with corresponding labels $\{y_i, y_j\}$, the weight is defined as $w_{ij} = \exp(-\|y_i - y_j\|_2)$. The loss promotes features with similar VA values to cluster together:

$$L_{contrast} = -\frac{1}{N}\sum_{i=1}^{N} \log \Bigg(\frac{\sum_{j\neq i} w_{ij} \cdot \exp(\text{sim}(f_i, f_j)/\tau)}{\sum_{j\neq i} \exp(\text{sim}(f_i, f_j)/\tau)} \Bigg) \quad (3)$$

where $\text{sim}(\cdot)$ denotes cosine similarity and τ is the temperature hyperparameter.

3.4 Multi-Objective Optimization

The model is trained with a weighted combination of three losses. The regression loss L_{reg} is a hybrid of MSE and Concordance Correlation Coefficient (CCC) loss. The CCC loss is particularly effective for dimensional sentiment analysis as it accounts for the covariance and mean shifts between predicted and gold values:

$$L_{CCC} = 1 - \frac{2\rho\sigma_{\hat{y}}\sigma_y}{\sigma_{\hat{y}}^2 + \sigma_y^2 + (\mu_{\hat{y}} - \mu_y)^2} \quad (4)$$

While MSE ensures point-wise accuracy, it often ignores the linear correlation between predictions and ground truth. We introduce the CCC loss into the loss function:

$$L_{reg} = 0.3 \cdot L_{MSE} + 0.7 \cdot L_{CCC} \quad (5)$$

The final objective function is:

$$L_{total} = L_{reg} + \alpha \cdot L_{rdrop} + \beta \cdot L_{contrast} \quad (6)$$

where $\alpha = 1.0$ and $\beta = 0.2$ are empirically determined scaling factors.

4 Experiment Setups

In this section, we describe the implementation details, hyperparameter configurations, and evaluation protocol used to develop and validate our system.

4.1 Dataset and Preprocessing

Our experiments are conducted on datasets of Subtask 1 (Dimensional Aspect Sentiment Regression) of Track A (DimABSA). Each instance in the dataset consists of a text segment and a list of target aspects. To adapt the input for the DeBERTa-v3 encoder, we transform each aspect into a prompt-based query:

Aspect: [Aspect_Name].
Intensity of valence and arousal: [Text]

The inputs are truncated or padded to a maximum sequence length of 128 tokens. We utilize the ‘DeBERTa-v3-large‘ tokenizer.

Table 1: Hyperparameter settings

Hyperparameter	**Value**
Backbone Model	DeBERTa-v3-large
Max Sequence Len.	128
Batch Size	32
LR (Backbone)	1×10^{-5}
LR (Head)	5×10^{-5}
Optimizer	AdamW
Weight Decay	0.01
Warmup Ratio	0.1
Epochs	15
Dropout Rate	0.2

4.2 Implementation Details

The system is implemented using PyTorch and the HuggingFace Transformers library. The specific hyperparameter settings used for our best-performing submission are summarized in Table 1.

We apply a differential learning rate strategy: the pre-trained DeBERTa backbone is fine-tuned with a conservative rate (1×10^{-5}), while the task-specific regression heads and contrastive projection layers use a higher rate (5×10^{-5}) to accelerate convergence on the dimensional sentiment task. We employ a linear learning rate scheduler with a 10% warmup phase.

4.3 Training Strategy and Loss Weights

To optimize the model's stability, we set the R-Drop α coefficient to 1.0 and the contrastive α to

Table 2: Performance evaluation of team SCUZANE across multiple languages and domains.

Language	Domain	RMSE	PCC-V	PCC-A	Rank
English	Laptop	1.4242	0.8408	0.5143	21 / 36
	Restaurant	1.3483	0.8579	0.6075	22 / 40
Japanese	Finance	0.9580	0.8061	0.5471	12 / 23
	Hotel	0.7129	0.9281	0.7268	11 / 22
Russian	Restaurant	1.5572	0.8969	0.6073	11 / 24
Tatar	Restaurant	2.3199	0.5707	0.2698	19 / 22
Ukrainian	Restaurant	1.5730	0.8891	0.5930	11 / 20
Chinese	Finance	0.5117	0.8738	0.6603	7 / 21
	Laptop	0.6981	0.9080	0.7446	9 / 24
	Restaurant	0.9636	0.8573	0.6353	9 / 25

Table 3: Performance comparisons of team SCUZANE vs. baselines.

Dataset	Team	Results
eng-laptop	SCUZANE(Ours)	**1.4242**
	Kimi-K2 Thinking	2.1893
	Qwen-3 14B	2.8089
eng-restaurant	SCUZANE(Ours)	**1.3483**
	Kimi-K2 Thinking	2.1461
	Qwen-3 14B	2.6427
jpn-finance	SCUZANE(Ours)	**0.958**
	Kimi-K2 Thinking	1.6396
	Qwen-3 14B	1.8964
jpn-hotel	SCUZANE2(Ours)	**0.7129**
	Kimi-K2 Thinking	1.7553
	Qwen-3 14B	2.2906
rus-restaurant	SCUZANE(Ours)	**1.5572**
	Kimi-K2 Thinking	1.7768
	Qwen-3 14B	2.1528
tat-restaurant	SCUZANE(Ours)	2.3199
	Kimi-K2 Thinking	**1.938**
	Qwen-3 14B	2.6367
ukr-restaurant	SCUZANE(Ours)	**1.573**
	Kimi-K2 Thinking	1.7805
	Qwen-3 14B	2.2121
zho-finance	SCUZANE(Ours)	**0.5117**
	Qwen-3 14B	1.4707
	Kimi-K2 Thinking	1.9652
zho-laptop	SCUZANE(Ours)	**0.6981**
	Kimi-K2 Thinking	1.644
	Qwen-3 14B	1.7706
zho-restaurant	SCUZANE(Ours)	**0.9636**
	Kimi-K2 Thinking	1.8959
	Qwen-3 14B	2.0073

0.2. The regression loss is a weighted sum of 30% MSE and 70% CCC, ensuring the model prioritizes the correlation between dimensions over simple point-wise error.

Training is monitored using an early stopping mechanism with a patience of 15 epochs, based on the Average Pearson Correlation Coefficient (Avg PCC) on the validation set. This prevents overfitting to the training distribution and ensures the best-performing weights are saved for final inference.

4.4 Evaluation Metrics

Following the official SemEval-2026 guidelines, our primary evaluation metric is Root Mean Square Error (RMSE). Additionally, we report the Pearson Correlation Coefficient (PCC) for both Valence ($PCC\text{-}V$) and Arousal ($PCC\text{-}A$) to measure the linear relationship between our predictions and the ground truth.

5 Results

We performed experiments on all six languages and four domains provided by the organizer. Table 2 demonstrates our system's RMSE, $PCC\text{-}V$, $PCC\text{-}A$ and rank for each dataset. Table 3 showcases performance comparisons of our system with baseline models.

6 Conclusion

In this paper, we presented our system submitted for the SemEval-2026 Task 3 - Dimensional Aspect-Based Sentiment Analysis (Track A), Subtask 1 — DimASR (Dimensional Aspect Sentiment

Regression). By combining a DeBERTa-v3-large backbone with a multi-task learning objective, and comprising CCC-informed regression, R-Drop consistency, and dimensional contrastive learning. Our system effectively maps textual aspect-level opinions to the continuous Valence-Arousal space. Our experiments on six languages and four domains demonstrate that our system achieves competitive performance and outperforms the baseline models.

References

Pengcheng He, Jianfeng Gao, and Weizhu Chen. 2021. Debertav3: Improving deberta using electra-style pre-training with gradient-disentangled embedding sharing. *Preprint*, arXiv:2111.09543.

Lung-Hao Lee, Jian-Hong Li, and Liang-Chih Yu. 2022. Chinese emobank: Building valence-arousal resources for dimensional sentiment analysis. *ACM Transactions on Asian and Low-Resource Language Information Processing*, 21(4):1–18.

Lung-Hao Lee, Liang-Chih Yu, Natalia Loukashevich, Ilseyar Alimova, Alexander Panchenko, Tzu-Mi Lin, Zhe-Yu Xu, Jian-Yu Zhou, Guangmin Zheng, Jin Wang, Sharanya Awasthi, Jonas Becker, Jan Philip Wahle, Terry Ruas, Shamsuddeen Hassan Muhammad, and Saif M. Mohammad. 2026. Dimabsa: Building multilingual and multidomain datasets for dimensional aspect-based sentiment analysis. *Preprint*, arXiv:2601.23022.

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 12th Workshop on Chinese Language Processing (SIGHAN-12)*, pages 180–189, Miami, Florida. Association for Computational Linguistics.

Xiaobo Liang, Lijun Wu, Juntao Li, Yue Wang, Qi Meng, Tao Qin, Wei Chen, Min Zhang, and Tie-Yan Liu. 2021. R-drop: Regularized dropout for neural networks. *Advances in Neural Information Processing Systems*, 34:10890–10905.

Saif M. Mohammad. 2018. Obtaining reliable human ratings of valence, arousal, and dominance for 20,000 english words. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 174–184, Melbourne, Australia. Association for Computational Linguistics.

Haiyun Peng, Lu Xu, Lidong Bing, Fei Huang, Wei Lu, and Luo Si. 2020. Knowing what, how and why: A near complete solution for aspect-based sentiment analysis. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 8600–8607.

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.

James A. Russell. 1980. A circumplex model of affect. *Journal of Personality and Social Psychology*, 39(6):1161–1178.

James A. Russell. 2003. Core affect and the psychological construction of emotion. *Psychological Review*, 110(1):145–172.

Liang-Chih Yu, Jonas Becker, Shamsuddeen Hassan Muhammad, Idris Abdulmumin, Lung-Hao Lee, Ying-Lung Lin, Jin Wang, Jan Philip Wahle, Terry Ruas, Alexander Panchenko, Ilseyar Alimova, Kai-Wei Chang, Lilian Wanzare, Nelson Odhiambo, Bela Gipp, and Saif M. Mohammad. 2026. SemEval-2026 task 3: Dimensional aspect-based sentiment analysis (DimABSA). In *Proceedings of the 20th International Workshop on Semantic Evaluation (SemEval-2026)*.

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. Association for Computational Linguistics.

Wenxuan Zhang, Xin Li, Yang Deng, Lidong Bing, and Wai Lam. 2023. A survey on aspect-based sentiment analysis: Tasks, methods, and challenges. *IEEE Transactions on Knowledge and Data Engineering*, 35(11):11019–11038.