

# PICT at SemEval-2026 Task 3: A Transformer-Based System for Dimensional Aspect-Aware Sentiment Regression with Weighted Layer Pooling

**Aditya Bhalgat**  
PICT, Pune, India  
adityabhalgat81@gmail.com

**Omkar Jagtap**  
PICT, Pune, India  
omkardjagtap2@gmail.com

**Anupama Phakatkar**  
PICT, Pune, India  
agphakatkar@pict.edu

## Abstract

Standard sentiment analysis may overlook the target-specific emotional intensity. The SemEval-2026 Task 3 (DimASR) tackles this problem by requiring the prediction of continuous valence and arousal scores for particular aspects. To counteract the negative transfer issue when training monolithic models on diverse datasets, the PICT team designed separate domain-specific pipelines for the Laptop and Restaurant domains. Our model design includes RoBERTa-Large with weighted layer pooling to retain syntax, aspect-focused multi-head attention to decouple context, and a deep residual regression head to improve the mapping to the continuous valence-arousal space. Regularized with R-Drop and Stochastic Weight Averaging (SWA), our model is able to effectively reduce the variance of the regression task. On the official English test data, our system ranked 3rd in the Restaurant domain (RMSE = 1.1958) and 9th in the Laptop domain (RMSE = 1.3261), significantly outperforming the organizer LLM baselines.

## 1 Introduction

Traditional, coarse-grained sentiment analysis poses a significant limitation to accurate opinion mining, as it often compresses complex, multi-faceted feedback into a single generic label. Extracting and isolating sentiments tied to specific targets—known as Aspect-Based Sentiment Analysis (ABSA)—is crucial for ensuring accurate feedback interpretation and facilitating effective decision-making in industry. One of the fastest-growing use cases for natural language processing (NLP) in e-commerce and hospitality is automated review analysis, often generated from massive volumes of heterogeneous customer feedback. However, capturing the true intensity and quality of these opinions is challenging, and simply categorizing sentiments as “positive” or “negative” artificially

limits their real-world utility. The reliability of standard categorical models in nuanced applications is a major concern due to their inability to distinguish between subtle emotional gradations (e.g., distinguishing a “warm” dish from a “dangerously hot” one). Robust solutions mapping sentiment to a continuous dimensional space are essential for addressing these concerns and enabling the highly accurate use of NLP in opinion mining.

The SemEval-2026 Task 3 (Yu et al., 2026) shared task focuses on identifying and predicting these continuous emotional intensities in user reviews. Each text contains subjective feedback tied to one or more specific aspect terms. The task involves moving beyond discrete classification; specifically, Track A, Subtask 1 (Dimensional Aspect Sentiment Regression, or DimASR) focuses on predicting real-valued Valence (degree of positivity) and Arousal (emotional intensity) scores bounded between 1.00 and 9.00 across distinct datasets, including the Laptop and Restaurant domains (Lee et al., 2026).

In this paper, we present our approach for the SemEval-2026 Task 3 competition. We develop a highly replicable, domain-isolated architecture using RoBERTa-Large, extending it with weighted layer pooling, aspect-aware attention, and a deep residual regression head to stabilize continuous affect mapping. We provide a detailed description of our methodology and results, followed by an ablation study and error analysis in the field of dimensional sentiment regression.

## 2 Background and Related Work

### 2.1 The Evolution of Aspect-Based Sentiment Analysis

Opinion mining has undergone a significant structural evolution over the past decade. Early foundational benchmarks, heavily driven by the SemEval-2014, 2015, and 2016 shared tasks (Pontiki et al.,

2014, 2015, 2016), formulated Aspect-Based Sentiment Analysis (ABSA) primarily as a discrete classification problem. The objective was to assign a categorical polarity to explicitly identified aspect terms. As the field matured, the community shifted toward unified extraction paradigms. Peng et al. (2020) introduced near-complete solutions for Aspect Sentiment Triplet Extraction (ASTE), while subsequent works tackled Aspect Sentiment Quad Prediction (ASQP), extracting the full interaction between aspects, categories, opinions, and discrete sentiments (Cai et al., 2021; Zhang et al., 2021). Despite these advancements, as noted in recent comprehensive surveys (Zhang et al., 2023), the reliance on discrete categorical labels remained a persistent bottleneck, artificially compressing the natural continuum of human feedback.

## 2.2 The Shift to Dimensional Affect

To capture the true intensity and gradation of emotion, psychological frameworks—most notably the Circumplex Model of Affect (Russell, 1980, 2003)—proposed mapping emotions onto a continuous, real-valued coordinate space comprising Valence (pleasure/displeasure) and Arousal (activation/deactivation). This dimensional paradigm was gradually adopted into natural language processing through the creation of continuous affective lexicons, such as EmoBank (Buechel and Hahn, 2017) and its multilingual extensions (Yu et al., 2016; Lee et al., 2022). Recent shared tasks have increasingly emphasized this continuous spectrum, progressing from document-level emotion intensity (Mohammad and Bravo-Marquez, 2017; Mohammad et al., 2018) to granular, aspect-level dimensional regression (Lee et al., 2024). Transitioning from categorical cross-entropy to continuous regression poses a distinct modeling challenge: networks must move beyond recognizing simple polarity to accurately capturing fine-grained linguistic modifiers.

## 2.3 Contextual Pooling and Regression Stabilization

Pre-trained Language Models (PLMs), particularly RoBERTa (Liu et al., 2019), currently dominate ABSA benchmarks. However, the standard adaptation paradigm—extracting the final hidden state of the [CLS] token—often discards crucial local syntactic dependencies. Representation learning literature demonstrates that the topmost layers of deep transformers are heavily biased toward global semantics, while lower layers retain the syntax nec-

essary for resolving local aspect-modifier interactions (Jawahar et al., 2019; Tenney et al., 2019). To counteract this, our architecture utilizes *weighted layer pooling*, aggregating a multi-level hierarchy of features, and focuses the network using multi-head attention mechanisms (Vaswani et al., 2017) tailored to the target aspect.

Furthermore, mapping these high-dimensional embeddings to a constrained 1-to-9 continuous scale is notoriously susceptible to high variance. To address this, we replace the standard shallow linear classifier with a deep residual network, leveraging skip-connections to ensure stable gradient flow while modeling complex, non-linear feature interactions (He et al., 2016). Finally, to aggressively regularize the network against the heterogeneous vocabularies of disparate domains, we integrate R-Drop (Wu et al., 2021), which minimizes the bidirectional KL-divergence between sub-models during dropout passes, and Stochastic Weight Averaging (SWA) (Izmailov et al., 2018), which targets wider, flatter minima in the loss landscape to significantly enhance out-of-distribution generalization.

# 3 System Overview

## 3.1 Architectural Approach

When designing our system for continuous affective regression, our primary objective was mitigating prediction variance. As illustrated in Figure 1, our approach diverges from standard transformer-based classifiers in three key ways. First, to prevent the negative transfer observed between heterogeneous lexicons, we abandoned cross-domain training, instead instantiating strictly independent pipelines for the Laptop and Restaurant datasets. Second, recognizing that standard [CLS] token classification discards the lower-level syntactic modifiers crucial for determining emotional intensity, we bypass the standard extraction method in favor of multi-level layer pooling and targeted attention. Finally, to stabilize the inherently volatile mapping of high-dimensional embeddings to a constrained 1-to-9 scalar space, we replace the traditional shallow linear classifier with a deep, regularized residual network.

## 3.2 Input Encoding and Layer Pooling

Given a context sentence of  $N$  tokens and a target aspect of  $M$  tokens, we format the input sequence

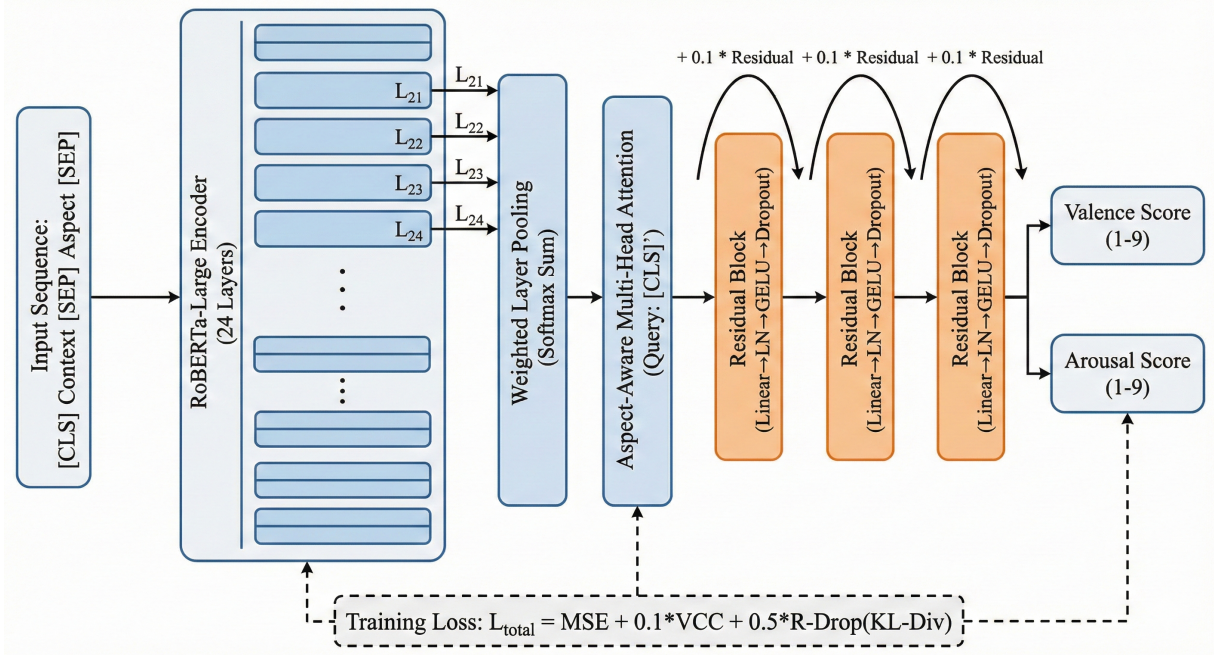


Figure 1: RoBERTa architecture for continuous Valence-Arousal prediction.

$X$  using standard RoBERTa separation:

$$X = \begin{matrix} [\text{CLS}] & w_1 & \dots & w_N \\ [\text{SEP}] & a_1 & \dots & a_M \\ & & & [\text{SEP}] \end{matrix} \quad (1)$$

Passing  $X$  through the 24-layer RoBERTa-Large encoder yields a sequence of hidden states  $H^l \in \mathbb{R}^{T \times d}$  for each layer  $l$ , where  $T$  is the sequence length and  $d = 1024$ . Standard adaptations rely solely on the final layer  $H^{24}$ , which tends to overfit to pre-training objectives and dilute local syntactic dependencies. To retain a richer feature hierarchy that captures both global semantics and local intensity modifiers (e.g., distinguishing between “good” and “exceptionally good”), we pool representations from the top four transformer layers ( $l \in \{21, 22, 23, 24\}$ ). Using learnable scalar weights  $\alpha_l$ , we compute the pooled sequence representation:

$$H_{pool} = \sum_{l=21}^{24} \text{softmax}(\alpha_l) H^l \quad (2)$$

This mechanism allows the network to dynamically balance its attention between high-level meaning and low-level structure during optimization.

### 3.3 Aspect-Aware Multi-Head Attention

With a rich, multi-level representation established, the network must next isolate the context specific to the target aspect. To explicitly ground the representation, we apply a 4-head attention mechanism

over the pooled layers. Treating the pooled [CLS] representation,  $h_{CLS}$ , as the query ( $Q$ ), and the full token sequence  $H_{pool}$  as the keys ( $K$ ) and values ( $V$ ):

$$\text{Attention}(Q, K, V) = \text{softmax} \left( \frac{QK^T}{\sqrt{d_k}} \right) V \quad (3)$$

Unlike standard approaches that use the mean of the aspect tokens as the Query, we explicitly use [CLS]. Because topmost layers encode global semantics, [CLS] serves as an information bottleneck summarizing overall sentence polarity. Using it as a Query forces the attention module to effectively filter the sequence for local modifiers that actively align with or contradict this global context.

### 3.4 Deep Residual Regression Head

The final architectural bottleneck is the regression phase. Standard fine-tuning often employs a shallow linear layer for prediction. We found this insufficient for modeling the complex, non-linear feature interactions required to map rich transformer embeddings to a continuous, bounded Valence-Arousal space.

Instead, we project the attention output through a 3-block residual network. For a given block input  $x^{(i)}$ , the forward pass integrates Dropout ( $p = 0.15$ ), Layer Normalization, and a GELU

activation:

$$x^{(i+1)} = x^{(i)} + 0.1 \cdot \text{Dropout}(\text{GELU}(\text{LayerNorm}(Wx^{(i)} + b))) \quad (4)$$

The highly scaled residual connection (0.1) is strictly enforced to maintain stable gradient flow, allowing the head to deeply process the interactions before making the final linear projection to the scalar values  $\hat{y}_v$  (Valence) and  $\hat{y}_a$  (Arousal).

### 3.5 Optimization and Variance Reduction

Fine-tuning a 355-million parameter model on relatively small datasets for continuous regression introduces significant risk of overfitting and high variance. To combat this, we optimize using AdamW with differential learning rates:  $1e-6$  for the encoder to preserve foundational weights, and  $1e-4$  for the randomly initialized regression head. To safely increase lexical diversity, we utilize WordNet via NLTK to perform stochastic synonym replacement ( $p = 0.3$ ) on non-aspect tokens, explicitly excluding adjectives and adverbs to ensure sentiment-bearing modifiers are not inadvertently altered.

Crucially, we optimize the model using a composite objective function integrated with R-Drop (Wu et al., 2021). By passing each batch through the network twice with different dropout masks, we yield two distinct prediction distributions,  $P_1$  and  $P_2$ . The total loss is defined as:

$$\mathcal{L}_{total} = \mathcal{L}_{MSE} + 0.1 \cdot \mathcal{L}_{VCC} + 0.5 \cdot \mathcal{D}_{KL}(P_1 \parallel P_2) \quad (5)$$

This composite loss balances absolute error reduction ( $\mathcal{L}_{MSE}$ ) and relative ranking accuracy ( $\mathcal{L}_{VCC}$ ). For scalar regression, we assume the model outputs represent the means of Gaussian distributions with fixed variance; under this assumption, the Kullback-Leibler divergence term  $\mathcal{D}_{KL}$  practically reduces to the L2 distance between the pre-activation outputs of the two passes, heavily regularizing against prediction variance.

## 4 Experimental Setup

### 4.1 Datasets and Evaluation Metrics

We evaluated our system utilizing the official data splits provided for SemEval-2026 Task 3, Track A, Subtask 1 (English Track) (Lee et al., 2026). To strictly isolate domain-specific linguistic features and prevent negative transfer, we partitioned the

data into two independent training regimes: Laptop and Restaurant. To ensure robust ensembling, we bypassed static train-validation splits and dynamically partitioned the training data using a 5-fold cross-validation strategy.

The primary metric for overall system ranking is the Root Mean Square Error (RMSE) computed between the predicted and gold Valence-Arousal (VA) values. Secondary metrics evaluate the Pearson Correlation Coefficient for Valence (PCC-V) and Arousal (PCC-A).

### 4.2 Implementation Details

To ensure full reproducibility, our code is publicly available at <https://github.com/omkardj2/DimASR>. We utilized the roberta-large pre-trained checkpoint with a maximum sequence length of 128 tokens. We maintained an effective batch size of 16, achieved via a micro-batch size of 4 accumulated over 4 forward passes, utilizing gradient checkpointing to save memory.

The models were optimized using AdamW (weight decay 0.01) for a maximum of 10 epochs. Stochastic Weight Averaging (SWA) (Izmailov et al., 2018) was engaged starting at epoch 5 with a dedicated static learning rate of  $1e-6$ . Training a single domain pipeline (5-fold CV) takes approximately 2.5 hours on a single NVIDIA GPU utilizing 18GB of VRAM. While domain-isolation increases total compute compared to multi-task learning, the complete avoidance of negative transfer justifies this trade-off.

## 5 Results and Analysis

System	RMSE <sub>VA</sub> ↓	PCC <sub>V</sub> ↑	PCC <sub>A</sub> ↑
<i>Laptop Domain</i>			
Org. Baseline (Qwen-3 14B)	2.8089	–	–
Org. Baseline (Kimi-K2)	2.1893	–	–
PICT System (9th)	<b>1.3261</b>	<b>0.8504</b>	<b>0.5519</b>
<i>Restaurant Domain</i>			
Org. Baseline (Qwen-3 14B)	2.6427	–	–
Org. Baseline (Kimi-K2)	2.1461	–	–
PICT System (3rd)	<b>1.1958</b>	<b>0.8938</b>	<b>0.6132</b>

Table 1: SemEval-2026 Task 3 (Track A) official result vs baselines.

Our domain-isolated approach achieved highly competitive performance in the shared task, yielding a combined average RMSE of 1.2610. As detailed in Table 1, our model achieved an RMSE of 1.1958 in the Restaurant domain and 1.3261 in the Laptop domain, significantly outperforming

the organizer’s zero-shot LLM baselines (Kimi-K2 Thinking and Qwen-3 14B). Our architectural focus on stabilization proved empirically effective: the standard deviation of the RMSE across the 5 independent validation folds was exceptionally low ( $\pm 0.012$  for Restaurant,  $\pm 0.018$  for Laptop), confirming reliable variance reduction.

### 5.1 Ablation Study

Architecture	RMSE <sub>VA</sub>
Full PICT System	<b>1.195</b>
– Deep Residual Head (Use Linear)	1.254
– Weighted Layer Pooling	1.238
– R-Drop Regularization	1.221
– SWA Smoothing	1.215
– Synonym Augmentation	1.206

Table 2: Ablation study on the Restaurant validation set.

To isolate component contributions, Table 2 details an ablation study. Replacing the deep residual head with a standard linear projection caused the most significant degradation ( $\Delta\text{RMSE} = +0.059$ ), confirming the necessity of non-linear mapping for continuous variables. Removing weighted layer pooling ( $\Delta\text{RMSE} = +0.043$ ) and R-Drop ( $\Delta\text{RMSE} = +0.026$ ) further validated our core stabilization techniques.

### 5.2 Error Analysis and Limitations

Error analysis reveals two key performance gaps. First, the gap between the Restaurant and Laptop domains stems from distinct lexical characteristics. Laptop reviews often feature highly technical, implicit sentiments (e.g., “battery lasts 2 hours”) lacking explicit emotional anchors, making continuous regression inherently more difficult than evaluating explicit restaurant feedback.

Second, Arousal correlation (PCC-A  $\sim 0.58$ ) consistently lagged behind Valence (PCC-V  $\sim 0.87$ ). We attribute this to lexical scarcity: Valence is heavily signaled by explicit adjectives, whereas Arousal relies on subtle intensifiers and complex syntactic structures. These subtle nuances are substantially harder for the attention heads to consistently isolate. Furthermore, our strict domain-isolation strategy limits zero-shot generalizability to unseen domains without dedicated retraining.

## 6 Conclusion

The PICT team’s submission demonstrates the potential of targeted variance reduction in continuous sentiment regression. By combining weighted layer pooling, aspect-aware attention, and a regularized residual head, we effectively addressed the instability inherent in dimensional ABSA. Future work will explore integrating Parameter-Efficient Fine-Tuning (PEFT), such as LoRA, to allow a single universal base model to leverage domain-specific adapters, thereby bridging the gap between our isolated approach and a compute-efficient general system.

## References

Sven Buechel and Udo Hahn. 2017. Emobank: Studying the impact of annotation perspective and representation format on dimensional emotion analysis. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics*, pages 578–585.

Hongjie Cai, Rui Xia, and Jie Yu. 2021. Aspect-category-opinion-sentiment quadruple extraction with implicit aspects and opinions. In *Findings of the Association for Computational Linguistics: EMNLP 2021*, pages 2909–2920.

Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2016. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778.

Pavel Izmailov, Dmitrii Podoprikin, Timur Garipov, Dmitry Vetrov, and Andrew Gordon Wilson. 2018. Averaging weights leads to wider optima and better generalization. In *Proceedings of the 34th Conference on Uncertainty in Artificial Intelligence*.

Ganesh Jawahar, Benoît Sagot, and Djamel Seddah. 2019. What does BERT learn about the structure of language? In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics (ACL)*, pages 3651–3657.

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 Loukashovich, 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 sighthan 2024 shared task for chinese dimensional aspect-based sentiment analysis. In *Proceedings of the 10th SIGHAN Workshop on Chinese Language Processing*, pages 165–174.
- Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2019. Roberta: A robustly optimized bert pretraining approach. *arXiv preprint arXiv:1907.11692*.
- Saif M Mohammad and Felipe Bravo-Marquez. 2017. Wasssa-2017 shared task on emotion intensity. In *Proceedings of the 8th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, pages 34–49.
- Saif M Mohammad, Felipe Bravo-Marquez, Mohammad Salameh, and Svetlana Kiritchenko. 2018. Semeval-2018 task 1: Affect in tweets. In *Proceedings of the 12th International Workshop on Semantic Evaluation*, pages 1–17.
- 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*, pages 8600–8607.
- Maria Pontiki, Dimitrios Galanis, Haris Papageorgiou, Ion Androutsopoulos, Suresh Manandhar, Mohammad AL-Smadi, Mahmoud Al-Ayyoub, Yanyan Zhao, Bing Qin, Orphee De Clercq, and 1 others. 2016. Semeval-2016 task 5: Aspect based sentiment analysis. In *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016)*, pages 19–30.
- Maria Pontiki, Dimitrios 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.
- Maria Pontiki, Dimitrios Galanis, John Pavlopoulos, Haris 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.
- 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.
- Ian Tenney, Dipanjan Das, and Ellie Pavlick. 2019. BERT rediscovers the classical NLP pipeline. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics (ACL)*.
- Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. In *Advances in Neural Information Processing Systems*, volume 30.
- Lijun Wu, Juntao Li, Yue Wang, Qi Meng, Tao Qin, Wei Chen, Zheng Min, Tie-Yan Liu, and 1 others. 2021. R-drop: Regularized dropout for neural networks. In *Advances in Neural Information Processing Systems*, volume 34, pages 10890–10905.
- 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)*. 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*, pages 540–545.
- Chen Zhang, Qiuchi Li, Dawei Song, and Linqi Song. 2021. Aspect sentiment quad prediction as paraphrase generation. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 9209–9219.
- 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.