kirito at SemEval-2026 Task 3: Dimensional Aspect-Based Sentiment Analysis via Sentence Structure Parsing Preprocessing and Prompt-Enhanced Instruction Tuning

Shuangjin Hu
Yunnan University, China
2163142837@qq.com

Abstract

Dimensional Aspect-Based Sentiment Analysis (DimABSA) integrates fine-grained aspect extraction with continuous Valence–Arousal (VA) regression, posing unique challenges for fine-grained opinion mining. This paper presents our system for SemEval-2026 Task 3 (Yu et al., 2026), with task-aligned strategies for three heterogeneous subtasks. For the DimASR task, we frame dimensional sentiment prediction as a supervised regression problem, paired with Low-Rank Adaptation (LoRA)-based parameter-efficient fine-tuning and a deep nonlinear regression head. For DimASTE and DimASQP tasks, we propose a lightweight sentence structure parsing preprocessing module, combined with prompt-enhanced instruction tuning for unified structured generation of aspect elements and VA scores. Experimental results on the official English test sets show that our system outperforms both official baselines across most settings, with syntax-guided prompting effectively improving aspect-opinion alignment and the dedicated regression head enhancing continuous sentiment modeling stability.

1 Introduction

Aspect-Based Sentiment Analysis (ABSA) has long been a core research direction in fine-grained opinion mining, with standardized benchmarks established by successive SemEval shared tasks (Pontiki et al., 2016). Most existing ABSA research relies on coarse-grained categorical sentiment labels, which inevitably lose nuanced differences in affective expression intensity. To address this limitation, SemEval-2026 Task 3 formalizes the DimABSA task, extending traditional ABSA with continuous VA sentiment dimensions and three heterogeneous subtasks with distinct input-output paradigms (Yu et al., 2026). The task defines three core subtasks. The first is DimASR, where the model predicts continuous VA scores for a pre-specified aspect term in a given sentence. The second is DimASTE, which requires end-to-end joint extraction of (Aspect, Opinion, VA) triplets from raw sentences. The third is DimASQP, which extends triplet extraction to (Aspect, Category, Opinion, VA) quadruples with additional aspect category constraints. Our work focuses on the English track of the task, covering two domains: restaurant user reviews (eng-res) and laptop user reviews (eng-lap). The dataset includes thousands of human-annotated review sentences, with all VA scores restricted to the range [1, 9] (Lee et al., 2026). Recent advances in Large Language Models (LLMs) have enabled flexible structured generation for ABSA tasks (Cai et al., 2021), while parameter-efficient tuning methods like QLoRA have made large model adaptation feasible with limited computing resources (Dettmers et al., 2024). Existing ABSA systems still have two key limitations. For continuous VA regression tasks, most works use simple linear projection heads that fail to capture the complex nonlinear relationship between text semantics and dimensional sentiment intensity. For structured extraction tasks, syntactic information is usually embedded as an internal model feature (Li et al., 2021), rather than being used as explicit upstream preprocessing to enhance prompt design and aspect-opinion alignment. To fill these gaps, we build a task-aligned dual-branch system for DimABSA, with two core contributions tailored to the heterogeneous characteristics of the three subtasks. We design a regression-based workflow with LoRA fine-tuning and a deep nonlinear regression head optimized for the given-aspect continuous value prediction paradigm of DimASR. We also propose a syntax-guided generative workflow with lightweight dependency parsing preprocessing and prompt-enhanced instruction tuning, which addresses the aspect-opinion alignment challenge in end-to-end structured extraction for DimASTE and DimASQP.

Proceedings of the The 20th International Workshop on Semantic Evaluation (2026), pages 913–918
July 3-4, 2026

2 Related Work

Our work is closely related to three core research lines in the field, with clear distinctions from existing approaches. Syntax-aware ABSA has been widely explored in prior work. Early studies used attention-enhanced recurrent architectures to incorporate syntactic information into target-dependent sentiment classification (Wang et al., 2016), while later works integrated pre-trained language models with graph convolutional networks to model dependency relations between aspect and opinion terms (Li et al., 2021). All these works treat syntactic information as an internal model feature, requiring additional structural encoders and increasing model complexity. In contrast, we convert syntactic relations into explicit textual signals injected into prompts, a lightweight approach that directly enhances the generation interpretability of LLMs without modifying the model architecture. Unified generative ABSA has been validated as an effective framework to avoid error accumulation in pipeline methods. Previous studies have shown that seq2seq frameworks and instruction tuning enable LLMs to complete joint structured extraction of aspect and opinion terms in a unified generation space (Cai et al., 2021; Lee et al., 2024). However, existing generative ABSA systems rarely leverage syntactic information to constrain the generation process, leading to frequent aspect-opinion alignment errors in complex sentences. Our work addresses this limitation by integrating syntactic preprocessing into prompt design, guiding the model to capture the correct dependency between aspect and opinion terms. Dimensional sentiment modeling originates from the circumplex model of affect proposed by Russell (1980), which characterizes emotion through continuous valence and arousal dimensions. Subsequent studies have built large-scale VA-based sentiment lexicons and benchmarks (Warriner et al., 2013; Buechel and Hahn, 2017), with the most relevant work being the official DimABSA benchmark released by Lee et al. (2026). Unlike previous works focusing on categorical sentiment classification, our system is optimized for the continuous VA regression paradigm, with a dedicated regression head to improve the accuracy of continuous value prediction.

3 Methods

We design a task-aligned dual-branch system workflow, split into two independent branches based on the paradigm of each subtask. We use Llama-3.1-8B-Instruct as the unified backbone for both branches, with LoRA fine-tuning to ensure consistent semantic representation and reduce computational overhead. The overall architecture of our system is shown in Figure 1.

3.1 Regression Workflow for DimASR

The DimASR task requires predicting continuous VA scores for a given aspect term, which is essentially a representation-to-continuous-value projection problem. We design a three-component workflow optimized for this regression paradigm. The first component is LoRA-based encoder fine-tuning. We use Llama-3.1-8B-Instruct as the backbone encoder, applying LoRA for parameter-efficient fine-tuning. We only update the attention projection layers of the model, freezing the base model weights to preserve its pre-trained contextual representation ability, while also significantly reducing the memory cost and overfitting risk of fine-tuning. The second component is a deep nonlinear regression head, built on top of the encoder's aspect-aware hidden representation. The mapping between high-dimensional text semantics and continuous VA scores is a complex nonlinear relationship that cannot be fully captured by a single linear layer, motivating our multi-layer design. This head maps contextual features to 2-dimensional VA scores through multi-layer nonlinear transformations with GELU activation, enabling hierarchical feature learning before projection into the VA space. We use Root Mean Square Error (RMSE) as the training loss, which measures the Euclidean distance between predicted and gold VA scores and is fully aligned with the official evaluation metric for this subtask (Yu et al., 2026).

3.2 Workflow for DimASTE and DimASQP

Both DimASTE and DimASQP require end-to-end extraction of structured sentiment elements with continuous VA scores, which we frame as instruction-guided structured generation tasks. The workflow consists of two core modules, with the detailed framework shown in Figure 2.

3.2.1 Sentence Structure Preprocessing

The core challenge of generative ABSA is ensuring correct alignment between aspect and opinion terms, especially in complex sentences with multiple pairs. To address this, we design a lightweight syntactic preprocessing module. We first apply

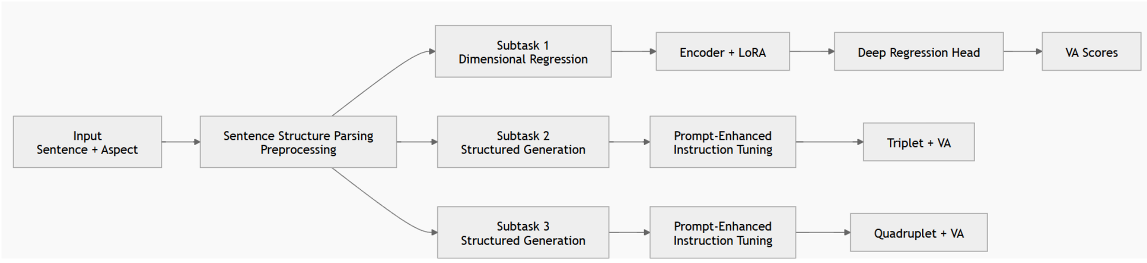

Figure 1: Overall dual-branch workflow of our DimABSA system. It is regression-based workflow for DimASR and syntax-guided generative workflow for DimASTE and DimASQP.

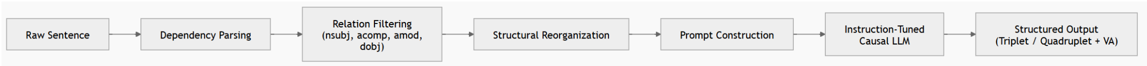

Figure 2: Detailed workflow of the syntax-guided prompt-enhanced instruction tuning framework for DimASTE and DimASQP.

dependency parsing to input sentences, retaining only four core syntactic relations most relevant to sentiment expression: nominal subject (nsubj), adjectival complement (acomp), adjectival modifier (amod), and direct object (dobj). These four relations cover the vast majority of interaction patterns between aspect entities and opinion descriptors in sentiment expressions, connecting target entities to their descriptive predicates and sentiment adjectives. We do not build additional graph neural networks based on these relations. Instead, we convert the filtered syntactic relations into structured textual pairs, which are injected into the prompt as explicit auxiliary structural signals to guide the generation process. This design avoids introducing additional model complexity, while directly constraining the model's attention to the correct aspect-opinion pairs.

3.2.2 Prompt-Enhanced Instruction Tuning

We use the same LoRA-adapted Llama-3.1-8B-Instruct as the backbone generation model. Each input to the model includes three parts: the original input text, the extracted syntactic relation pairs, and explicit output format instructions for triplet or quadruple generation. For DimASQP, we extend the output format instructions to include aspect category constraints, adding category definition prompts to improve the model's fine-grained category modeling ability, with all other settings consistent with DimASTE. The model generates structured outputs in strict JSON format, with all elements including continuous VA scores generated in a unified decoding space, avoiding error accumulation in pipeline methods.

4 Experiments

4.1 Experimental Setup

4.1.1 Tasks and Evaluation Metrics

We evaluate on the official English test sets (Lee et al., 2026), with VA predictions restricted to [1, 9] and rounded to two decimal places. DimASR is evaluated by RMSE in the VA space (lower = better), measuring the average Euclidean deviation between predicted and gold VA values. DimASTE and DimASQP are evaluated by the official continuous F1 (cF1) score (higher = better).

4.1.2 Baseline Models

We compare with two official baselines from the task leaderboard: the zero-/one-shot closed-source Kimi-K2 Thinking baseline, and the open-source Qwen-3 14B baseline with full supervised fine-tuning on the task training set. All baseline results are taken from the official leaderboard, with our experiments conducted on the same test sets under the identical evaluation protocol.

4.1.3 Implementation Details

All experiments use Hugging Face Transformers v4.45.0 and PEFT v0.12.0, with dependency parsing via spacy 3.7.2. For LoRA, we set rank=8, alpha=16, dropout=0.05, applying LoRA to all attention projection layers with base weights frozen. The regression head uses four fully connected layers (2048→512→128→2) . We use the AdamW optimizer (learning rate 2e-4, weight decay 1e-4), batch size 16, 50 training epochs with early stopping (patience=5), and random seed 42 for reproducibility. No external data is used beyond the official training set.

4.2 Main Results

We summarize main results in Table 1, with detailed analysis below.

For DimASR, our model achieves the lowest RMSE on both domains, outperforming the Kimi-K2 baseline by 34.9% (eng-res) and 31.4% (eng-lap), and the Qwen-3 baseline by 47.2% (eng-res) and 46.6% (eng-lap), demonstrating the effectiveness of our LoRA tuning and dedicated regression head. For DimASTE, our model outperforms both baselines on both domains, with 15.4% cF1 improvement over the Kimi-K2 baseline on eng-res and 7.0% on eng-lap, validating that our syntax-guided prompting improves aspect-opinion alignment and extraction consistency. For DimASQP, our model achieves the highest cF1 on eng-res (38.8% improvement over Kimi-K2), and outperforms the supervised Qwen-3 baseline by 62.2% on eng-lap. The smaller gain on eng-lap stems from the more fine-grained professional aspect categories in the laptop domain, which increase the difficulty of exact category matching required by the cF1 metric.

4.3 Ablation Study

We conduct ablation studies on both domains following the variable control method, with model variants created by removing one core component at a time from the full model.

4.3.1 Ablation for DimASR

We validate the contribution of LoRA fine-tuning and the deep regression head, with results in Table 2.

On both domains, LoRA-based tuning reduces RMSE compared to full fine-tuning, confirming that full fine-tuning damages pre-trained contextual representation and causes overfitting, while LoRA preserves general language ability during task adaptation. Removing the deep regression head increases RMSE on both domains, validating that the nonlinear head effectively captures the complex mapping between text semantics and VA scores.

4.3.2 Ablation for DimASTE and DimASQP

We validate the contribution of prompt enhancement, syntactic preprocessing, and core syntactic filtering, with results in Table 3.

Across both domains and tasks, removing prompt enhancement causes the largest performance drop, demonstrating that explicit task instructions and format constraints are critical for guiding structured generation. Removing syntactic preprocessing further reduces cF1, verifying that syntactic information enhances aspect-opinion alignment. Using full syntactic relations also degrades performance, confirming that redundant relations introduce prompt noise, while our filtered core relations retain critical aspect-opinion interaction information.

4.4 Error Analysis

We conduct quantitative error analysis on both test sets, organized by subtask. For DimASR, total RMSE is decomposed into two core subtypes: valence score deviation, the dominant error type where the model fails to capture implicit sentiment intensity; and arousal score deviation , occurring when the model misjudges emotional activation intensity in sentences with intensifiers or figurative language. For DimASTE, invalid predictions fall into three subtypes: aspect-opinion alignment errors, the most frequent error in sentences with multiple pairs; incomplete element extraction, mostly missing implicit opinion terms; and VA score mismatch, occurring even with correctly aligned elements. For DimASQP, invalid predictions fall into three subtypes: aspect category mismatch, the dominant error due to strict cF1 matching requirements; structured element extraction errors; and VA score mismatch.

4.5 Discussion

Across all three subtasks, we observe three consistent trends aligned with our method design and experimental results. Supervised parameter-efficient adaptation consistently outperforms zero-/one-shot prompting baselines on continuous VA regression tasks, as LoRA fine-tuning enables the model to learn the task-specific mapping between text semantics and VA scores, which cannot be fully captured by in-context learning alone. Syntax-guided prompt enhancement effectively improves structured extraction performance, especially on the eng-res domain with more straightforward sentiment expressions, while still bringing consistent performance gains on the more complex eng-lap domain, validating the generalizability of our preprocessing module. The strict categorical exact matching requirement in the cF1 metric amplifies the impact of category prediction errors, making quadruplet extraction in DimASQP significantly more challenging than triplet extraction in DimASTE, es-

Model	Setting	DimASR (RMSE ↓)		DimASTE (cF1 ↑)		DimASQP (cF1 ↑)	
		eng-res	eng-lap	eng-res	eng-lap	eng-res	eng-lap
Baseline (Qwen-3 14B)	Supervised Fine-Tuning	2.6427	2.8089	0.4483	0.3827	0.2673	0.1529
Baseline (Kimi-K2 Thinking)	Zero-/One-Shot	2.1461	2.1893	0.4920	0.4424	0.3746	**0.2795**
Ours (Llama-3.1-8B-Instruct)	LoRA Supervised Fine-Tuning	**1.3966**	**1.5010**	**0.5676**	**0.4733**	**0.5201**	0.2480

Table 1: Main results across all three subtasks on two English domains. Best results for each subtask and domain are marked in bold.

Model Variant	RMSE ↓	
	eng-res	eng-lap
w/o LoRA (Full fine-tuning + Linear head)	1.5602	1.6825
w/o Deep Regression Head (LoRA + Linear head)	1.5231	1.6312
Full Model	**1.3966**	**1.5010**

Table 2: Ablation results for DimASR on eng-res and eng-lap test sets.

Model variant	DimASTE (cF1 ↑)		DimASQP (cF1 ↑)	
	eng-res	eng-lap	eng-res	eng-lap
w/o Prompt Enhancement	0.4312	0.3528	0.2987	0.1721
w/o Syntactic Preprocessing	0.4528	0.3705	0.3295	0.1904
w/o Core Syntactic Filtering	0.5010	0.4186	0.3672	0.2215
Full Model	**0.5676**	**0.4733**	**0.5201**	**0.2480**

Table 3: Ablation results for DimASTE and DimASQP on eng-res and eng-lap test sets.

pecially in the eng-lap domain with professional fine-grained categories.

5 Conclusions

This work presents a task-aligned dual-branch system for SemEval-2026 Task 3 DimABSA, tailored for the heterogeneous characteristics of its three subtasks instead of forced architectural unification. For DimASR, we adopt LoRA-based parameter-efficient fine-tuning with a deep nonlinear regression head for continuous VA regression optimization. For DimASTE and DimASQP, we design a lightweight sentence structure parsing preprocessing module to extract sentiment-relevant syntactic pairs, which are integrated into prompt-enhanced instruction tuning for unified structured generation of aspect elements and VA scores. Experimental results on English restaurant and laptop domains show our system outperforms official baselines in most settings, and ablation studies verify the effectiveness and generalizability of all core modules. Our approach provides a lightweight and practical solution for structured dimensional sentiment analysis under limited computational resources.

Acknowledgments

We thank the organizers of SemEval-2026 Task 3 for providing the DimABSA benchmark, evaluation protocols and official baseline results. We also acknowledge the developers of open-source NLP tools and parameter-efficient fine-tuning libraries used in this work.

Limitations

Our approach has three key limitations consistent with error analysis results. First, the syntactic preprocessing module's performance relies on dependency parsing quality, with parsing errors propagating to prompt construction and generation consistency, especially in informal review texts with typos or non-standard syntax. Second, LoRA-based parameter-efficient fine-tuning limits the model's adaptation to professional domain terminology compared to full fine-tuning in high-resource scenarios. Third, our experiments are limited to the English domain, with no cross-lingual generalization exploration on the benchmark's multilingual subsets. Future work will focus on improving syntactic noise robustness, enhancing fine-grained category modeling, and exploring cross-lingual transfer learning for DimABSA.

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: Volume 2, Short Papers*, pages 578–585.

Hongjie Cai, Guanhao Huang, Qian Liu, and Xuanjing Huang. 2021. A unified generative framework for aspect-based sentiment analysis. 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 2416–2429.

Tim Dettmers, Artidoro Pagnoni, Ari Holtzman, and Luke Zettlemoyer. 2024. Qlora: Efficient finetuning of quantized llms. *Advances in Neural Information Processing Systems*, 37.

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.

Seonwoo Lee and 1 others. 2024. Instruction tuning for aspect-based sentiment analysis. In *Findings of the Association for Computational Linguistics: ACL 2024*.

Xin Li and 1 others. 2021. A dual graph convolutional network for aspect-based sentiment analysis. 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 6319–6329.

Maria Pontiki, Dimitrios Galanis, Harris Papageorgiou, Ion Androutsopoulos, Suresh Manandhar, Mohammad Al-Smadi, Mahmoud Al-Ayyoub, Yanyan Zhao, Bing Qin, Orphée 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.

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

Wenya Wang and 1 others. 2016. Attention-based lstm for target-dependent sentiment classification. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pages 606–615.

Amy Beth Warriner, Victor Kuperman, and Marc Brysbaert. 2013. Norms of valence, arousal, and dominance for 13,915 english lemmas. *Behavior Research Methods*, 45(4):1191–1207.

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.