CCIIPLab at SIGHAN-2024 dimABSA Task: Contrastive Learning-Enhanced Span-based Framework for Chinese Dimensional Aspect-Based Sentiment Analysis

Zeliang Tong, Wei Wei[✉]
Cognitive Computing and Intelligent Information Processing (CCIIP) Laboratory,
School of Computer Science and Technology, Huazhong University of Science and Technology
{tongzeliang, weiw}@hust.edu.cn

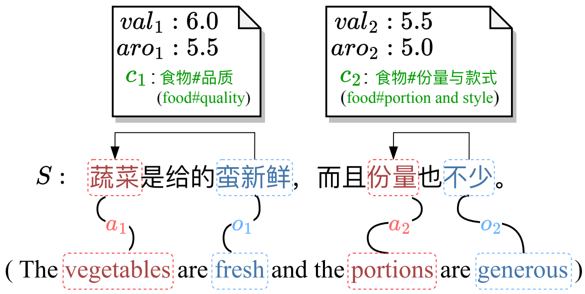
Abstract

This paper describes our system and findings for SIGHAN-2024 Shared Task Chinese Dimensional Aspect-Based Sentiment Analysis (dimABSA). Our team **CCIIPLab** proposes an Contrastive Learning-Enhanced **Span**-based (**CL-Span**) framework to boost the performance of extracting triplets/quadruples and predicting sentiment intensity. We first employ a span-based framework that integrates contextual representations and incorporates rotary position embedding. This approach fully considers the relational information of entire aspect and opinion terms, and enhancing the model's understanding of the associations between tokens. Additionally, we utilize contrastive learning to predict sentiment intensities in the valence-arousal dimensions with greater precision. To improve the generalization ability of the model, additional datasets are used to assist training. Experiments have validated the effectiveness of our approach. In the official test results, our system ranked **2nd** among the three subtasks. Our code is publicly available at https://github.com/tongzeliang/SIGHAN2024.

1 Introduction

Aspect-Based Sentiment Analysis (ABSA) is an important task in Natural Language Processing (NLP), and is beneficial for many downstream tasks, such as emotional conversation generation (Wei et al., 2019; Liu et al., 2022) and recommendation system (Zhao et al., 2023; Wang et al., 2023). However, previous work has focused primarily on discrete sentiment polarity, with little attention given to the Valence-Arousal (VA) space. This dimensional approach represents affective states as continuous numerical values across multiple dimensions, providing more fine-grained sentiment information.

To address this issue, the SIGHAN-2024 shared task formulates three subtasks that challenge participants to develop ABSA systems based on dimensional sentiment information (Lee et al., 2024). As Figure 1 shows, the three subtask can be illustrated as follows:



Subtask	Input	Output
Intensity Prediction	$S + a_1$	(a_1, val_1, aro_1)
	$S + a_2$	(a_2, val_2, aro_2)
Triplet Extraction	S	(a_1, o_1, val_1, aro_1)
	S	(a_2, o_2, val_2, aro_2)
Quadruple Extraction	S	$(a_1, o_1, c_1, val_1, aro_1)$
	S	$(a_2, o_2, c_2, val_2, aro_2)$

Figure 1: Illustration of three dimABSA subtasks. Aspect terms, opinion terms and categories are highlighted in red, blue and green, respectively. The terms "val" and "aro" represent the valence and arousal intensity of affective states, respectively, both ranging from 1 to 9.

- **Subtask 1: Intensity Prediction.** Predict the *valence-arousal ratings* for a given sentence and its specific aspect.
- **Subtask 2: Triplet Extraction.** Extract all sentiment triplets (*aspect*, *opinion*, *intensity*) from a given sentence.
- **Subtask 3: Quadruple Extraction.** Extract all sentiment quadruples (*aspect*, *category*, *opinion*, *intensity*) from a given sentence.

As an extension of the ABSA task, dimABSA becomes notably challenging due to the following two difficulties: 1) **Multiple Aspect-Opinion Pairing**. In sentences with multiple aspects and opinions, determining which opinion corresponds

✉ Corresponding Author

Proceedings of the 10th SIGHAN Workshop on Chinese Language Processing (SIGHAN-10), pages 102–111
August 16, 2024 ©2024 Association for Computational Linguistics

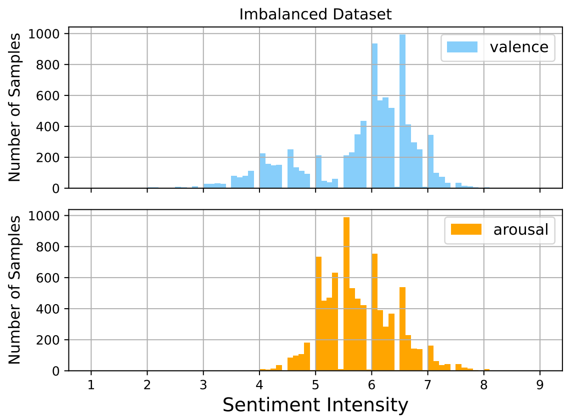


Figure 2: The distribution of valence-arousal ratings, where intensity ratings ranging from 1 to 9 are segmented into equal interval groups.

to which aspect becomes particularly challenging. To mitigate this problem, several efforts (Wu et al., 2020; Chen et al., 2022b; Liang et al., 2023) have been made, but they are not comprehensive for modeling the relationship between tokens, which is significant in the aspect-opinion pairing process (Chen et al., 2022a). 2) **Imbalanced Dataset**. As illustrated by Figure 2, the number of samples with neutral sentiment intensity is much greater than those with extreme sentiment intensity for both the valence and arousal dimensions, which leads to biased predictions from the model. Although there are some methods for addressing data imbalance like Re-Sampling (RS) (Zhou et al., 2020; Zhang and Pfister, 2021), most of them improve the performance of sparsely labeled samples at the expense of densely labeled samples (Zhang et al., 2023), leading to suboptimal results and and cannot be seamlessly migrated to dimABSA tasks.

In this paper, we develop a Contrastive Learning-Enhanced Span-based Framework for the dimABSA task to address the aforementioned challenges. **Firstly**, given the excellent performance of span-based methods in various NLP tasks (Xu et al., 2021), we explicitly generate span representations for all possible aspect and opinion spans. To comprehensively capture the relational information between spans, we integrate the contextual representations and incorporate Rotary Position Embedding (RoPE) (Su et al., 2024), which facilitates improved semantic understanding. **Secondly**, as self-supervised learning can improve robustness to data imbalance (Li et al., 2022), we employ contrastive learning to optimize feature representations in regression tasks. This approach adjusts the distance between samples in the embedding space according to their target values and subsequently leverages this feature to predict sentiment intensity.

Extensive experiments show that our method performs well across all three subtasks. On the official leaderboard, the Mean Absolute Error (MAE) for valence and arousal in subtask 1 ranks **2nd** and **1st**, respectively. The Pearson Correlation Coefficient (PCC) for valence and arousal in subtask 1 both rank **3rd**. The F1 scores for triplet and quadruple extraction in Subtasks 2 and 3 both rank **2nd**.

The paper is structured as follows: Section 2 provides a concise review of related work. In Section 3, we outline our proposed system. Section 4 covers the experimental details, including the dataset, setup, results, and discussions. Section 5 analyzes the effectiveness of contrastive learning and further examines the performance of our methods in low-resource settings. Finally, Section 6 presents a brief conclusion.

2 Related Work

2.1 ABSA Tasks

ABSA tasks, which aim to analyze sentiment from a fine-grained perspective, include three fundamental subtasks: Aspect Term Extraction (ATE) (Xu et al., 2018; Ma et al., 2019; Yang et al., 2020), Opinion Term Extraction (OTE) (Wan et al., 2020; Veyseh et al., 2020), and Aspect Sentiment Classification (ASC) (Tian et al., 2021; Wang et al., 2021a; Zhou et al., 2021). In recent years, research has increasingly focused on composite ABSA tasks, which integrate multiple basic tasks. Peng et al. (2020) introduced the Aspect Sentiment Triplet Extraction (ASTE) task, and they proposed a two-stage pipeline model to independently extract aspect-opinion-sentiment triplets. Subsequently, some end-to-end methods (Fei et al., 2021; Liang et al., 2023) were also applied to this task. Following this advancement, Zhang et al. (2021) introduced the Aspect-Sentiment Quad Prediction (ASQP) task, addressing it through the Seq2Seq generative modeling paradigm. However, these works primarily focus on discrete sentiment polarity, making it challenging to perceive subtle sentiment differences when predicting continuous sentiment intensity.

2.2 Contrastive Learning

Contrastive Learning methods learn feature representations by contrasting positive pairs against negative pairs, and have widely used in many downstream tasks, such as recommendation systems (Zou et al., 2022; Wang et al., 2022), knowledge graphs (Fang et al., 2022; Xu et al., 2023), etc.. Recent research has started to utilize contrastive learning to address the long-tail distribution problem in image classification (Wang et al., 2021b; Xuan and Zhang, 2024), aiming to obtain improved feature representations. This prompts us to utilize contrastive learning in the dimABSA task to tackle the challenge of imbalanced datasets.

3 Methodology

3.1 Overview

Problem Statement. Let $s = \{w_i\}_n$ and $\mathbf{A} = \{a_j\}_m$ be a sentence and a predefined set of aspects, where n and m represents the length of s and the number of aspects contained in s ($\mathbf{A}$ is only provided in subtask 1). The goal of subtask 1 is to predict the sentiment intensity $val_j, aro_j \in [1, 9]$ for each aspect $a_j \in \mathbf{A}$. The object of subtask 2 and 3 is to extract a set of sentiment triplets $\mathcal{T} = \{(a, o, val\text{-}aro)_m\}_{m=1}^{|\mathcal{T}|}$ and quadruples $\mathcal{Q} = \{(a, o, c, val\text{-}aro)_m\}_{m=1}^{|\mathcal{Q}|}$, where a, o and c represent aspect term, opinion term and category.

Architecture. As Figure 3 demonstrates, our system contains three components: 1) the **Aspect-Opinion Pairing Module** identifies aspect terms and opinion terms from the original sentence, and establishes their relationships to form valid aspect-opinion pairs, 2) the **Sentiment Scoring Module** assesses the sentiment intensity based on the original sentence and the extracted aspect-opinion pairs, 3) the **Category Prediction Module** conducts category classification utilizing the original sentence and the extracted aspect-opinion pairs. Each module is trained independently, and each subtask is accomplished through the collaboration of different modules. This pipeline structure enhances the flexibility and scalability of the system, allowing different processing steps to be optimized and adjusted independently.

3.2 Aspect-Opinion Pairing Module

This module identifies relevant aspects and their corresponding opinions within the sentence and accurately pairs them. As Figure 4 shows, this foundational step is crucial for subsequent analysis and prediction, ensuring that each aspect is matched with its opinion, forming the basis for further inference.

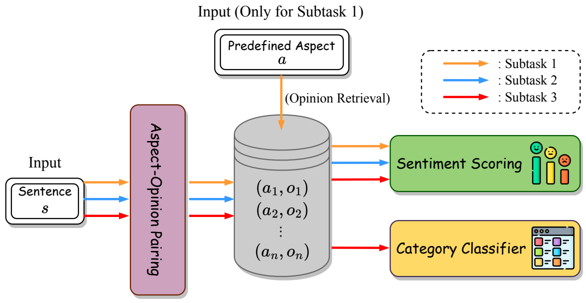


Figure 3: Architecture of our system. Arrows of different colors indicate the computational processes specific to each corresponding subtask.

Sentence Encoder. We use MacBERT (Cui et al., 2021) to generate contextual word representations by,

$$\hat{\mathbf{H}} = \hat{\mathbf{h}}_{cls}, \{\hat{\mathbf{h}}_j\}_n, \hat{\mathbf{h}}_{sep} = \text{MacBERT}(\{w_i\}_n) \tag{1}$$

where $\hat{\mathbf{h}}_j$ is the contextual embedding of word w_j. We then integrate RoPE into the token representation via an additional multi-head attention layer,

$$\mathbf{H} = \text{MultiHead}\left(\mathbf{Q}, \mathbf{K}, \mathbf{V}\right) \tag{2}$$

$$= ||_{z=1}^{Z} \text{Attention}\left(R_\theta^i \mathbf{W}_q^z \hat{\mathbf{H}}, R_\theta^j \mathbf{W}_k^z \hat{\mathbf{H}}, \mathbf{W}_v^z \hat{\mathbf{H}}\right) \tag{3}$$

Z is the number of attention heads, W_q^z, W_k^z and W_v^z are trainable parameter of the zth head of attention. Note that the rotational position encoding matrix should vary for different positions in the sequence, here we use R_θ^i and R_θ^j for simplicity.

Aspect and Opinion Extraction. We use $SP = \{\mathbf{sp}_{i,j} \mid 0 \leq j - i \leq l\}$ to represent all possible spans in s, where i and j represent the start and end positions in s respectively, and the maximum length of span $\mathbf{sp}_{i,j}$ is l. We define the representation of span $\mathbf{sp}_{i,j}$ as,

$$\mathbf{sp}_{i,j} = [\mathbf{h}_i; \mathbf{h}_j] \tag{4}$$

where the semicolon represents concatenation.

Next, we employ a fully connected layer to evaluate the validity of each span $\mathbf{sp}_{i,j}$, assigning a label distribution $y_e \in \{Aspect, Opinion, Invalid\}$,

$$p_{i,j}^A, p_{i,j}^O, p_{i,j}^{IV} = \text{softmax}\left(\mathbf{W}_e \mathbf{sp}_{i,j} + \mathbf{b}_e\right) \tag{5}$$

where W_e, b_e are trainable parameters.

Inspred by Xu et al. (2021), to mitigate the complexity inherent in the subsequent calculation process, we retain a specified proportion of spans for

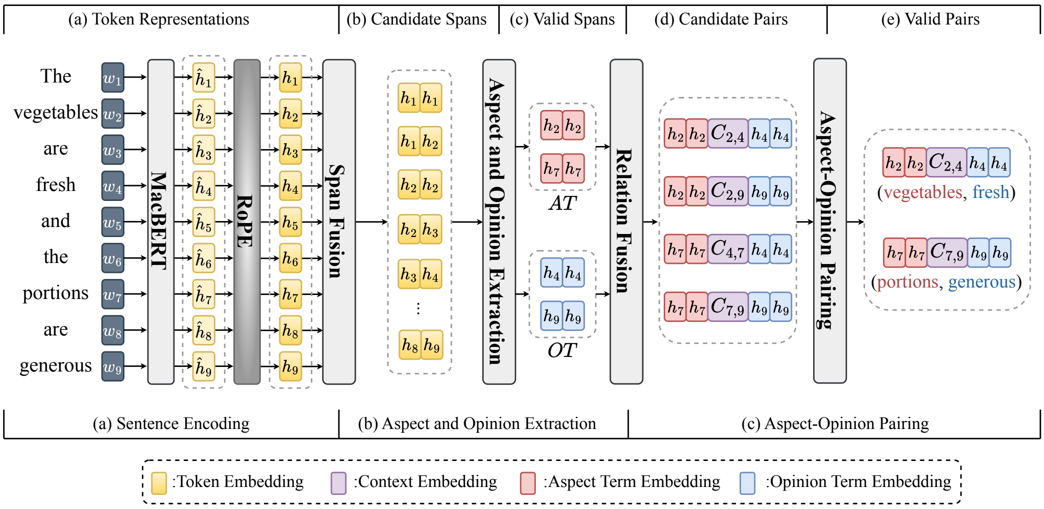


Figure 4: The overall framework of our Aspect-Opinion Pairing Module. Initially, the encoder derives base contextual representations for the input sentence. Subsequently, we integrate Rotary Position Embedding (RoPE) into the token representations to facilitate enhanced discourse comprehension. Following this, aspect terms and opinion terms are extracted based on the RoPE-enhanced representations. Finally, we identify valid aspect-opinion pairs from the extracted aspect and opinion terms.

both the aspect and the opinion candidate set, selecting those with the highest scores as determined by Equation 5. The refined sets of aspects and opinions can be denoted as $AT = \{\ldots, \mathbf{sp}^A_{i,j}, \ldots\}$ and $OT = \{\ldots, \mathbf{sp}^O_{i,j}, \ldots\}$, respectively, each comprising nr elements, where $r \in [0, 1]$ indicates the proportion of retained elements.

Aspect-Opinion Pairing. After acquiring the aspect and opinion candidate sets from the previous stage, we proceed by pairing them in all possible combinations, resulting in the following representation,

$$\mathbf{f}_{a,b,c,d} = \left[\mathbf{sp}^A_{a,b}; \mathbf{sp}^O_{c,d}; \mathbf{C}_{b,c}\right] \quad (6)$$

$$\mathbf{C}_{b,c} = \text{Max-Pooling}\left([\mathbf{h}_{b+1} : \mathbf{h}_{c-1}]\right) \quad (7)$$

$\mathbf{C}_{b,c}$ represents the contextual information of $\mathbf{sp}_{a,b}$ and $\mathbf{sp}_{c,d}$. Subsequently, we employ a fully connected layer to process the representation of each $\mathbf{f}_{a,b,c,d}$. This layer evaluates the validity of each aspect-opinion pair, assigning a label distribution $y_g \in \{Valid, Invalid\}$.

$$p^V_{a,b,c,d}, p^{IV}_{a,b,c,d} = \text{softmax}\left(\mathbf{W}_g \mathbf{f}_{a,b,c,d} + \mathbf{b}_g\right) \quad (8)$$

where W_g, b_g are trainable parameters.

Training. The training target is to minimize the cross-entropy loss of the extraction and pairing tasks.

$$\mathcal{L} = \alpha \mathcal{L}_e + \beta \mathcal{L}_g \quad (9)$$

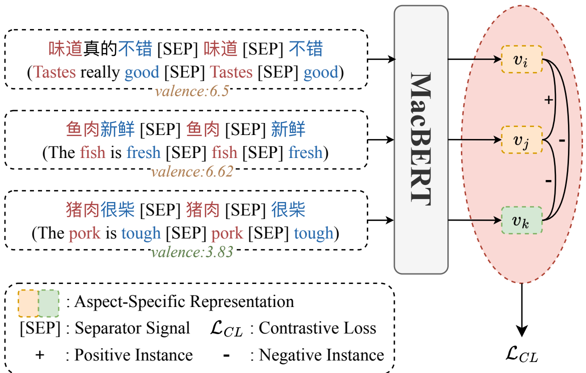


Figure 5: The overall framework of the Sentiment Scoring Module employs a contrastive loss, which ensures that samples with similar regression labels share similar features in the embedding space, while samples with differing labels are positioned further apar.

$$\mathcal{L}_e = -\sum_{sp_{i,j} \in SP} \log P\left(y^*_e \mid p^e_{i,j}\right) \quad (10)$$

$$\mathcal{L}_g = -\sum_{sp_{a,b} \in AT, sp_{c,d} \in OT} \log P\left(y^*_g \mid p^g_{a,b,c,d}\right) \quad (11)$$

Here, y^*_e and y^*_g represents the ground-truth label of the extraction and pairing tasks for $\mathbf{sp}_{i,j}$ and $\mathbf{f}_{a,b,c,d}$, respectively.

3.3 Sentiment Scoring Module

In this section, we employ contrastive learning to enhance aspect-specific representations and predict sentiment intensity for each aspect in the valence-arousal space, as illustrated in Figure 5.

Aspect Specific Representation. In this part, we first utilize MacBERT as encoder to generate aspect-specific representation for each aspect:

$$u_j = < \{w_i\}_n, \text{[SEP]}, \{a_j\}_{\hat{t}}, \text{[SEP]}, \{o_j\}_{\tilde{t}} > \tag{12}$$

$$\mathbf{H}_j = \text{MacBERT}(u_j) \tag{13}$$

where $\hat{t}$ and $\tilde{t}$ are lengths of the aspect a_j and its corresponding opinion o_j. Note that in the case of multi-aspect sentences, this module is employed multiple times, with each iteration focusing on one aspect. The aspect-specific feature representations is then obtained by max pooling,

$$\mathbf{v}_j = \text{Max-Pooling}\left(\mathbf{H}_j\right) \tag{14}$$

Contrastive Learning. After generating aspect-specific representations, most prior studies directly employ these representations for downstream tasks. Nonetheless, the performance is constrained by imbalanced datasets, resulting in suboptimal outcomes. To address this limitation, we incorporate contrastive learning to enhance feature optimization. Let $\{\mathbf{v}_i\}_G$ be defined as the set of all representations within a batch, and G denote the number of these representations, we first translate them for contrastive loss through a MLP combined with ℓ_2-normalization,

$$\mathbf{u}_i = \ell_2\text{-norm}\left(\text{MLP}\left(\mathbf{v}_i\right)\right) \tag{15}$$

In the absence of category labels, we establish two thresholds, δ_1 and δ_2, to facilitate the selection of positive and negative sample pairs respectively,

$$< i, j >= \begin{cases} + & \text{if } |y_i^* - y_j^*| \leq \delta_1 \\ - & \text{if } |y_i^* - y_j^*| \geq \delta_2 \end{cases} \tag{16}$$

where y_i^*, y_j^* represent the ground-truth of the sentiment intensity. Therefore, through the above rules, we can construct a positive set $\mathcal{P}_i$ and a negative set $\mathcal{N}_i$ for each representation u_i. The contrastive loss is calculated as follows,

$$\mathcal{L}_{CL} = -\frac{1}{G}\sum_{i=1}^{G}\sum_{\mathbf{u}_j \in \mathcal{P}_i} \log \frac{e^{\text{sim}(\mathbf{u}_i, \mathbf{u}_j^+)/\tau}}{\sum_{\mathbf{u}_k \in \mathcal{N}_i} e^{\text{sim}(\mathbf{u}_i, \mathbf{u}_k^-)/\tau}} \tag{17}$$

Training. The sentiment intensity was calculate by the aspect-specific representations $\{v_i\}_G$ through a single linear layer, and the total loss can be calculated as follow,

$$\mathcal{L} = \alpha\mathcal{L}_R + (1-\alpha)\mathcal{L}_{CL} \tag{18}$$

$$\mathcal{L}_R = \frac{1}{G}\sum_{i=1}^{G} ||y_i^* - f_\theta\left(v_i\right)|| \tag{19}$$

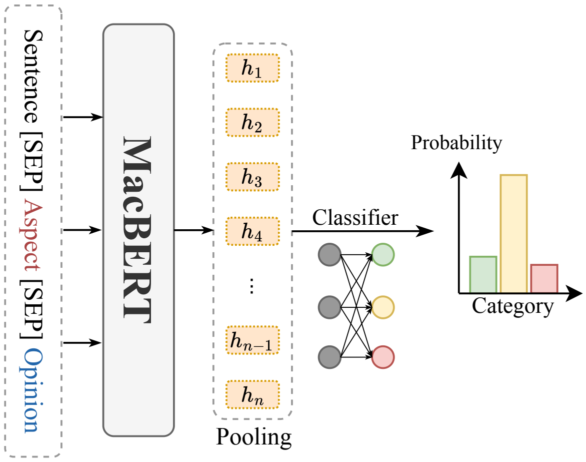


Figure 6: Architecture of the Category Prediction Module.

where f_θ denotes the linear projection. Note that the Sentiment Scoring Module is deployed twice within the system, with two identical components operating in parallel to independently extract valence and arousal features for regression prediction. This design allows each encoder to specialize in a specific emotional dimension, optimizing for the unique characteristics of each dimension and reducing feature interference during the contrastive learning process.

3.4 Category Prediction Module

This part employs the same method as the Sentiment Scoring Module to obtain aspect-specific representations v_i. As Figure 6 shows, these representations are subsequently passed through a fully connected layer with a softmax activation function, producing probability distributions across all categories,

$$p_i = \text{softmax}\left(\mathbf{W}_p\mathbf{v}_i + \mathbf{b}_p\right) \tag{20}$$

The training loss is formulated as the cross-entropy loss between the ground-truth and the predicted label distributions for all aspects,

$$\mathcal{L}_{CE} = -\frac{1}{N}\sum_{i=1}^{N}\log\left(y_i^* \mid p_i\right) \tag{21}$$

where y^* represent the ground-truth label.

3.5 Deployment Order

Table 1 illustrates the computational sequence of each component in the model across the three subtasks. All three subtasks necessitate an initial phase

	Task 1	Task 2	Task 3
Aspect-Opinion Pairing	✓	✓	✓
Opinion Retrieval	✓	✗	✗
Category Prediction	✗	✗	✓
Sentiment Scoring	✓	✓	✓

Table 1: The computational sequence of each component within the model across the three subtasks.

Dataset	Sentence-Level			Aspect-Level	
	Sgl-Senti	Mul-Senti	All	Null	All
train	4165	1885	6000	169	8354
test_1	1460	540	2000	-	2658
$\text{test}_{2,3}$	-	-	2000	-	-

Table 2: Dataset statistics. "Sgl-Senti" and "Mul-Senti" indicate the number of sentences expressing sentiment toward single or multiple aspects, respectively. "NULL" signifies that the aspect entity is omitted in sentence.

of aspect-opinion extraction and pairing. In Subtask 1, the Opinion Retrieval (OR) Module is employed, meaning that during sentiment intensity regression, we retrieve the corresponding opinion extracted in the aspect-opinion pair module for each predefined aspect, as this is a critical feature for both valence and arousal predictions. In cases where extraction or pairing fails, "NULL" is used to fill the missing opinion term.

4 Experiment

4.1 Dataset and Setup

We evaluate our model on the official dataset of the SIGHAN-2024 shared task (Lee et al., 2024), which uses Simplified Chinese characters. Dataset statistics are shown in Table 2. To enhance the model's ability to discern subtle sentiment nuances when predicting continuous sentiment intensity, we incorporated the Chinese EmoBank (EB) (Lee et al., 2022) as an auxiliary training resource. We fine-tuned the Sentiment Scoring Module on this supplementary dataset using the methodology outlined in Section 3.3, subsequently employing the fine-tuned parameters to initialize the model for the ensuing task training.

The Aspect-Opinion Pairing Module is trained for 30 epochs with a batch size of 16, and the other modules are trained for 10 epochs with a batch size of 128. AdamW optimizer (Loshchilov and Hutter, 2018) is adopted with a learning rate 2e-5 and weight decay 1e-2 for model training. The two thresholds δ_1 and δ_2 used in contrastive learning are set to 0.5 and 2 respectively. The maximum span length l is set as 10. We select the best model weights for testing based on performance on the validation set. MAE and PCC are evaluation metrics for subtask 1, while the F1 score is used as the evaluation metric for subtasks 2 and 3.

4.2 Baseline

Since no existing method is specifically designed for dimABSA, we re-implemented Span-ASTE (Xu et al., 2021) and STAGE (Liang et al., 2023), which are high-performing span-based systems closely related to the task, and used them as our baseline.

4.3 Main Results

Table 3 presents the results of out method in the final test set. Observations are: 1) Our purposed model outperforms the baseline, and achieves relatively good results in the final rankings, with one metric ranking **1st**, seven metrics ranking **2nd**, and two metrics ranking **3rd**. The performance improvement of our model primarily stems from a more powerful pre-training model, richer relational information for aspect-opinion pairing, and more robust feature representation for sentiment scoring. 2) Predicting sentiment intensities in the arousal dimension is significantly more challenging than in the valence dimension. In subtask 1, all models exhibit higher MAE in the arousal dimension compared to the valence dimension. In subtask 2 and 3, the F1 score based on arousal is about 5% lower than the F1 score based on valence. We infer that this complexity arises because predicting the level of arousal requires a comprehensive assessment of the overall context, tone, and other nuanced factors, which introduces corresponding challenges in the data annotation and training process.

4.4 Ablation Study

We also conduct an ablation study to verify the effectiveness of our proposed method. The results are shown in Table 3. Observations are: 1) For the Aspect-Opinion Pairing Module, **w/o CR** and **w/o RoPE** mean that we remove the contextual representation and rotational position embedding during the computation. Without the enhancement of relational features between spans and spans, the model's performance slightly degrades. 2) For the Sentiment Scoring Module, **w/o OR** indicates that the opinion term has been removed from the input, and **w/o CL** indicates that the contrastive loss has been omitted during the training process. As

Models	Subtask 1				Subtask 2			Subtask 3		
	V-MAE	V-PCC	A-MAE	A-PCC	V-F1	A-F1	VA-F1	V-F1	A-F1	VA-F1
Span-ASTE$^\natural$	-	-	-	-	0.473	0.458	0.310	-	-	-
STAGE$^\natural$	-	-	-	-	0.491	0.468	0.324	-	-	-
CL-Span$^\natural$	0.320	0.900	0.321	**0.767**	0.562	0.517	0.385	0.540	0.500	0.375
CL-Span†	0.302	0.910	0.314	**0.767**	0.565	0.519	0.391	0.547	0.505	0.379
CL-Span$^\circ$	**0.294**$_{(2)}$	**0.916**$_{(3)}$	**0.309**$_{(1)}$	0.766$_{(3)}$	**0.573**$_{(2)}$	**0.522**$_{(2)}$	**0.403**$_{(2)}$	**0.555**$_{(2)}$	**0.507**$_{(2)}$	**0.389**$_{(2)}$
CL-Span$^\circ_{\text{w/o-OR}}$	0.327	0.913	0.354	0.761	0.523	0.484	0.371	0.511	0.470	0.359
CL-Span$^\circ_{\text{w/o-CL}}$	0.311	0.912	0.331	0.764	0.548	0.511	0.380	0.542	0.503	0.377
CL-Span$^\circ_{\text{w/o-EB}}$	0.319	0.912	0.340	**0.767**	0.539	0.501	0.374	0.535	0.487	0.364
CL-Span$^\circ_{\text{w/o-RoPE}}$	-	-	-	-	0.565	0.514	0.391	0.547	0.510	0.379
CL-Span$^\circ_{\text{w/o-CR}}$	-	-	-	-	0.564	0.518	0.391	0.545	0.504	0.378

Table 3: Main results and ablation results on the test set. "∘", "†" and "♮" indicates that the context encoder is MacBERT-base, RoBERTa-base (Cui et al., 2020) and BERT-base (Kenton and Toutanova, 2019) in Chinese version respectively. Note that "w/o" indicates the removal of the corresponding component from the model. The numbers in brackets represent the ranking of the metric in the official leaderboard.

a result, the model's performance drops dramatically, indicating that the opinion term is crucial for predicting sentiment intensity and that contrastive loss guides the model to obtain a more appropriate feature distribution when the dataset is imbalanced. 3) **w/o EB** indicates that the additional data from Chinese EmoBank was not used during training, resulting in deteriorated model performance. This verifies that Chinese EmoBank provides valuable supplementary information when the training data is insufficient. In summary, each module of our method significantly contributes to the overall performance on the dimABSA task.

5 Analysis

5.1 Effect of Contrastive Learning

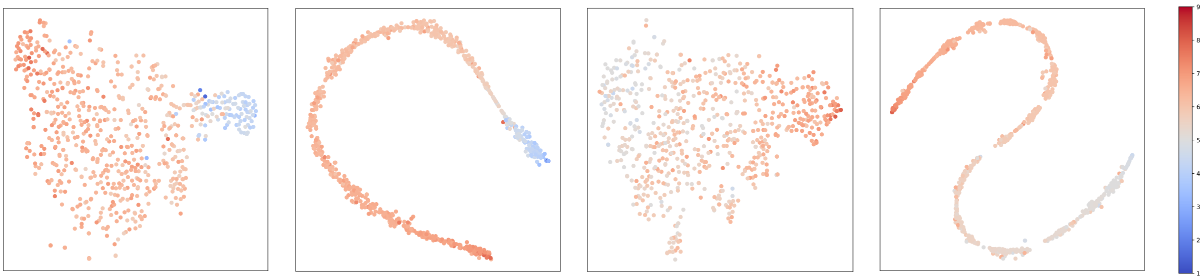
To further verify the effectiveness of contrastive learning, we visualize the sample features with and without it, as shown in Figure 7. Models without contrastive loss struggle to capture the underlying continuous information in the data, resulting in fragmented and disordered representations. Conversely, features derived through contrastive learning preserve a coherent semantic structure, ensuring that semantically similar target values remain proximate in the feature space. Therefore, we infer that the improvement in effect comes from the neat and sequential feature representation brought by contrastive learning, which makes the feature space more discriminative and has stronger generalization ability in unknown data. At the same time, through contrastive learning, even if there are fewer samples with labels in certain intervals, the model will still learn the feature representation of these samples because they are frequently used for comparison during training. This method helps to balance the model's attention to different labels, thereby alleviating the problem of imbalanced datasets.

5.2 Low-Resource Scenario

As a challenging task, dimABSA faces significant issues related to data scarcity. To address this, we investigated the impact of contrastive learning under various training data conditions. As depicted in Figure 8, the model utilizing contrastive learning consistently achieves lower MAE values, especially as the dataset size diminishes. Furthermore, the slower increase in MAE for the contrastive learning model indicates that contrastive learning enhances the model's robustness and generalization capabilities, allowing it to maintain performance even under low-resource conditions.

5.3 Case Study

Figure 9 presents some case studies of this system, where aspect terms are highlighted in red and opinion terms in blue. Observations are: 1) In cases (a) and (b), the complete system achieved optimal results in the majority of sentiment intensity predictions. Notably, even for test data with sparse training data distribution, such as values like "7.62" and "2.17", CL-Span consistently outperformed other methods, underscoring its robustness in accurately predicting less frequent valence and arousal values. 2) In case (c), our proposed CL-Span successfully pairs all aspect terms with their corresponding opinion terms. In contrast, Span-ASTE fails to recognize the pair ("*onion*", "*caught*



(a) Valence - w/o CL (b) Valence - CL (c) Arousal - w/o CL (d) Arousal - CL

Figure 7: Visualization of learned aspect-specific representations of different methods on the validation set of dimABSA. The features are reduced to two dimensions using TSNE (Van der Maaten and Hinton, 2008), with the sentiment intensity ranging from 1 to 9. The color gradient from blue to red represents the increasing intensity of sentiment, where blue indicates the lowest intensity (1) and red indicates the highest intensity (9).

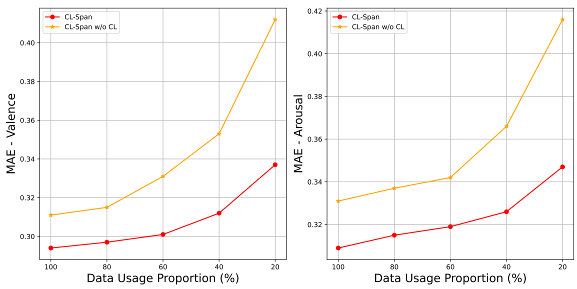


Figure 8: Comparison of the MAE for valence and arousal predictions by models with contrastive learning (red) and without contrastive learning (yellow) at different data usage ratios.

(a)	Sentence:	猪排油腻感很高，吃太多会想吐。 Pork chops are very greasy and eating too much will make you feel like vomiting.	
	Golden (val-aro):	3.50#5.00(Pork chops, greasy)	2.17#7.00(Pork chops, feel like vomiting)
	CL-Span:	**3.42#5.13**	**2.26#6.97**
	w/o CL:	3.65#5.37	2.52#6.85
	w/o EB:	3.73#5.21	3.01#6.88
(b)	Sentence:	多样且超级美味的蔬菜料理。 Various and super delicious vegetarian dishes.	
	Golden (val-aro):	6.75#6.12(vegetarian dishes, various)	7.62#7.25(vegetarian dishes, super delicious)
	CL-Span:	6.80#**6.14**	**7.55#7.21**
	w/o CL:	**6.72**#6.01	7.01#6.96
	w/o EB:	6.54#6.05	6.93#6.80

(c)	Sentence:	让我眼前一亮的是洋葱，吃起来很甜并且完全都不呛。 What really caught my eye was the onion; it tasted sweet and wasn't pungent at all .		
	Golden (Pair):	(onion, caught my eye)	(onion, sweet)	(onion, wasn't pungent at all)
	CL-Span:	✓	✓	✓
	Span-ASTE:	✗	✓	✓
	STAGE:	✓	✓	✗

Figure 9: Example cases with golden standard labels alongside the predictions from our model compared with other baseline models. The bold numbers indicate the relatively optimal results.

my eye"), and the STAGE model overlooks the pair ("*onion*", "*wasn't pungent at all*"). We attribute the superior performance of our model to the integration of contextual representations and RoPE, which enhances the semantic understanding and connectivity between aspect and opinion terms.

6 Conclusion

This paper describes our system for the dimABSA task. We develop a Contrastive Learning-Enhanced Span-based Framework, which integrates contextual representations and RoPE into feature representation to enhance semantic understanding. Additionally, we employ contrastive learning to optimize feature representations. Our system demonstrates significant effectiveness, achieving a 2nd place ranking across three subtasks.

Limitations

This section discusses some improvements that can be made in future work. 1) The pipeline model structure used in this study divides the processing steps into independent modules, allowing each module to be developed, tested, and optimized separately. However, it also introduces the issue of error propagation, where errors in earlier stages can affect subsequent modules. In future work, we will focus on minimizing the impact of error propagation or consider testing an end-to-end model paradigm. 2) In the Sentiment Scoring Module, our system employs two MacBERT encoders to separately extract valence and arousal features for independent regression prediction. This approach reduces feature interference during the contrastive learning process and better captures the unique characteristics of each dimension. However, this results in the parameters of this module doubling to 204M. We will consider other encoding strategies instead of simply deploying two MacBERT separately.

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China under Grant No. 62276110, No. 62172039 and in part by the fund of Joint Laboratory of HUST and Pingan Property Casualty Research (HPL). The authors would also like to thank the anonymous reviewers for their comments on improving the quality of this paper.

References

Hao Chen, Zepeng Zhai, Fangxiang Feng, Ruifan Li, and Xiaojie Wang. 2022a. Enhanced multi-channel graph convolutional network for aspect sentiment triplet extraction. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 2974–2985.

Yuqi Chen, Chen Keming, Xian Sun, and Zequn Zhang. 2022b. 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.

Yiming Cui, Wanxiang Che, Ting Liu, Bing Qin, Shijin Wang, and Guoping Hu. 2020. Revisiting pre-trained models for Chinese natural language processing. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: Findings*, pages 657–668, Online. 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.

Yin Fang, Qiang Zhang, Haihong Yang, Xiang Zhuang, Shumin Deng, Wen Zhang, Ming Qin, Zhuo Chen, Xiaohui Fan, and Huajun Chen. 2022. Molecular contrastive learning with chemical element knowledge graph. In *Proceedings of the AAAI conference on artificial intelligence*, volume 36, pages 3968–3976.

Hao Fei, Yafeng Ren, Yue Zhang, and Donghong Ji. 2021. Nonautoregressive encoder–decoder neural framework for end-to-end aspect-based sentiment triplet extraction. *IEEE Transactions on Neural Networks and Learning Systems*, 34(9):5544–5556.

Jacob Devlin Ming-Wei Chang Kenton and Lee Kristina Toutanova. 2019. Bert: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of NAACL-HLT*, pages 4171–4186.

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.*

Tianhong Li, Peng Cao, Yuan Yuan, Lijie Fan, Yuzhe Yang, Rogerio S Feris, Piotr Indyk, and Dina Katabi. 2022. Targeted supervised contrastive learning for long-tailed recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6918–6928.

Shuo Liang, Wei Wei, Xian-Ling Mao, Yuanyuan Fu, Rui Fang, and Dangyang Chen. 2023. Stage: span tagging and greedy inference scheme for aspect sentiment triplet extraction. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, pages 13174–13182.

Yifan Liu, Wei Wei, Jiayi Liu, Xianling Mao, Rui Fang, and Dangyang Chen. 2022. Improving personality consistency in conversation by persona extending. In *Proceedings of the 31st ACM International Conference on Information & Knowledge Management*, pages 1350–1359.

Ilya Loshchilov and Frank Hutter. 2018. Decoupled weight decay regularization. In *International Conference on Learning Representations*.

Dehong Ma, Sujian Li, Fangzhao Wu, Xing Xie, and Houfeng Wang. 2019. Exploring sequence-to-sequence learning in aspect term extraction. In *Proceedings of the 57th annual meeting of the association for computational linguistics*, pages 3538–3547.

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.

Jianlin Su, Murtadha Ahmed, Yu Lu, Shengfeng Pan, Wen Bo, and Yunfeng Liu. 2024. Roformer: Enhanced transformer with rotary position embedding. *Neurocomputing*, 568:127063.

Yuanhe Tian, Guimin Chen, and Yan Song. 2021. Aspect-based sentiment analysis with type-aware graph convolutional networks and layer ensemble. In *Proceedings of the 2021 conference of the North American chapter of the association for computational linguistics: human language technologies*, pages 2910–2922.

Laurens Van der Maaten and Geoffrey Hinton. 2008. Visualizing data using t-sne. *Journal of machine learning research*, 9(11).

Amir Pouran Ben Veyseh, Nasim Nouri, Franck Dernoncourt, Dejing Dou, and Thien Huu Nguyen. 2020. Introducing syntactic structures into target opinion word extraction with deep learning. In *Proceedings of the 2020 Conference on Empirical Methods in*

Natural Language Processing (EMNLP), pages 8947–8956.

Hai Wan, Yufei Yang, Jianfeng Du, Yanan Liu, Kunxun Qi, and Jeff Z Pan. 2020. Target-aspect-sentiment joint detection for aspect-based sentiment analysis. In *Proceedings of the AAAI conference on artificial intelligence*, volume 34, pages 9122–9129.

Bo Wang, Tao Shen, Guodong Long, Tianyi Zhou, and Yi Chang. 2021a. Eliminating sentiment bias for aspect-level sentiment classification with unsupervised opinion extraction. In *Findings of the Association for Computational Linguistics: EMNLP 2021*, pages 3002–3012.

Peng Wang, Kai Han, Xiu-Shen Wei, Lei Zhang, and Lei Wang. 2021b. Contrastive learning based hybrid networks for long-tailed image classification. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 943–952.

Ziyang Wang, Huoyu Liu, Wei Wei, Yue Hu, Xian-Ling Mao, Shaojian He, Rui Fang, and Dangyang Chen. 2022. Multi-level contrastive learning framework for sequential recommendation. In *Proceedings of the 31st ACM International Conference on Information & Knowledge Management*, pages 2098–2107.

Ziyang Wang, Wei Wei, Shanshan Feng, Xian-Ling Mao, Minghui Qiu, Dangyang Chen, and Rui Fang. 2023. Exploiting group-level behavior pattern for session-based recommendation. *IEEE Transactions on Knowledge and Data Engineering*.

Wei Wei, Jiayi Liu, Xianling Mao, Guibing Guo, Feida Zhu, Pan Zhou, and Yuchong Hu. 2019. Emotion-aware chat machine: Automatic emotional response generation for human-like emotional interaction. In *Proceedings of the 28th ACM international conference on information and knowledge management*, pages 1401–1410.

Zhen Wu, Chengcan Ying, Fei Zhao, Zhifang Fan, Xinyu Dai, and Rui Xia. 2020. Grid tagging scheme for aspect-oriented fine-grained opinion extraction. In *Findings of the Association for Computational Linguistics: EMNLP 2020*, pages 2576–2585.

Hu Xu, Bing Liu, Lei Shu, and S Yu Philip. 2018. Double embeddings and cnn-based sequence labeling for aspect extraction. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 592–598.

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.

Yi Xu, Junjie Ou, Hui Xu, and Luoyi Fu. 2023. Temporal knowledge graph reasoning with historical contrastive learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, pages 4765–4773.

Shiyu Xuan and Shiliang Zhang. 2024. Decoupled contrastive learning for long-tailed recognition. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pages 6396–6403.

Yunyi Yang, Kun Li, Xiaojun Quan, Weizhou Shen, and Qinliang Su. 2020. Constituency lattice encoding for aspect term extraction. In *Proceedings of the 28th international conference on computational linguistics*, pages 844–855.

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.

Yifan Zhang, Bingyi Kang, Bryan Hooi, Shuicheng Yan, and Jiashi Feng. 2023. Deep long-tailed learning: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.

Zizhao Zhang and Tomas Pfister. 2021. Learning fast sample re-weighting without reward data. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 725–734.

Sen Zhao, Wei Wei, Xian-Ling Mao, Shuai Zhu, Minghui Yang, Zujie Wen, Dangyang Chen, and Feida Zhu. 2023. Multi-view hypergraph contrastive policy learning for conversational recommendation. In *Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 654–664.

Boyan Zhou, Quan Cui, Xiu-Shen Wei, and Zhao-Min Chen. 2020. Bbn: Bilateral-branch network with cumulative learning for long-tailed visual recognition. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 9719–9728.

Yuxiang Zhou, Lejian Liao, Yang Gao, Zhanming Jie, and Wei Lu. 2021. To be closer: Learning to link up aspects with opinions. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 3899–3909.

Ding Zou, Wei Wei, Xian-Ling Mao, Ziyang Wang, Minghui Qiu, Feida Zhu, and Xin Cao. 2022. Multi-level cross-view contrastive learning for knowledge-aware recommender system. In *Proceedings of the 45th international ACM SIGIR conference on research and development in information retrieval*, pages 1358–1368.