electronics

MDPI

Article

A Hybrid Approach to Dimensional Aspect-Based Sentiment Analysis Using BERT and Large Language Models

Yice Zhang [1,2], Hongling Xu [1,2], Delong Zhang [1,2] and Ruifeng Xu [1,2,3,*]

[1] Harbin Institute of Technology, Shenzhen 518067, China; 20b951019@stu.hit.edu.cn (Y.Z.); xuhongling@stu.hit.edu.cn (H.X.); 22s151078@stu.hit.edu.cn (D.Z.)
[2] Guangdong Provincial Key Laboratory of Novel Security Intelligence Technologies, Shenzhen 518067, China
[3] Peng Cheng Laboratory, Shenzhen 518055, China
* Correspondence: xuruifeng@hit.edu.cn

Abstract: Dimensional aspect-based sentiment analysis (dimABSA) aims to recognize aspect-level quadruples from reviews, offering a fine-grained sentiment description for user opinions. A quadruple consists of aspect, category, opinion, and sentiment intensity, which is represented using continuous real-valued scores in the valence–arousal dimensions. To address this task, we propose a hybrid approach that integrates the BERT model with a large language model (LLM). Firstly, we develop both the BERT-based and LLM-based methods for dimABSA. The BERT-based method employs a pipeline approach, while the LLM-based method transforms the dimABSA task into a text generation task. Secondly, we evaluate their performance in entity extraction, relation classification, and intensity prediction to determine their advantages. Finally, we devise a hybrid approach to fully utilize their advantages across different scenarios. Experiments demonstrate that the hybrid approach outperforms BERT-based and LLM-based methods, achieving state-of-the-art performance with an F1-score of 41.7% on the quadruple extraction.

Keywords: aspect-based sentiment analysis; BERT; large language models; dimensional sentiment analysis

check for updates

Citation: Zhang, Y.; Xu, H.; Zhang, D.; Xu, R. A Hybrid Approach to Dimensional Aspect-Based Sentiment Analysis Using BERT and Large Language Models. *Electronics* **2024**, *13*, 3724. https://doi.org/10.3390/electronics13183724

Academic Editor: Arkaitz Zubiaga

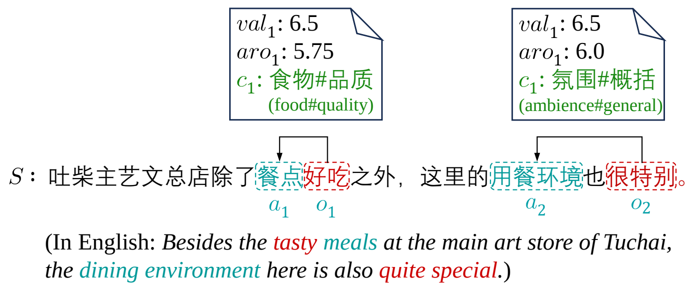
Received: 31 August 2024
Revised: 15 September 2024
Accepted: 16 September 2024
Published: 19 September 2024

1. Introduction

Sentiment analysis, a continually evolving and increasingly prominent subfield of natural language processing, has attracted widespread attention [1,2]. Within this domain, aspect-based sentiment analysis (ABSA) poses a critical and challenging problem, focusing on recognizing aspect-level sentiments and opinions of users [3]. ABSA research typically involves four sentiment elements, which are defined as follows: (1) aspect term (a), which refers to the entity mentioned in the sentence; (2) aspect category (c), a predefined category that represents specific facets and dimensions under evaluation; (3) opinion term (o), the sentiment word or phrase directed towards the entity; and (4) sentiment polarity, which classifies the sentiment into positive, neutral, or negative categories [4,5]. For example, given the review "*the crust is thin, the ingredients are fresh, and the staff is friendly*", the quadruples are (*crust*, food#quality, *thin*, positive), (*ingredients*, food#quality, *fresh*, positive) and (*staff*, services#general, *friendly*, positive).

Existing ABSA research has predominantly treated sentiment as coarse-grained polarities, thereby neglecting the complexities inherent in sentiment dimensions. Pioneering this field, Lee et al. [6] introduced dimensional aspect-based sentiment analysis (dimABSA), which quantifies sentiment states as continuous real-valued scores in valence–arousal dimensions. Valence measures the positivity or negativity, and arousal evaluates the degree of emotional activation [7]. As illustrated in Figure 1, dimABSA comprises three subtasks. Subtask 1 is intensity prediction, which focuses on predicting the valence–arousal scores, *val-aro*, of the given aspect. Subtask 2 is triplet extraction, which aims to extract the triplets composed of (a, o, *val-aro*) from the given sentence. Subtask 3 is quadruple extraction, extracting the quadruples composed of (a, c, o, *val-aro*) from the given sentence.

In Subtask 1, the valence–arousal scores are evaluated on a continuous scale, whereas in Subtasks 2 and 3, they are assessed at the integer level. Compared to traditional ABSA tasks, the main challenge of the dimABSA tasks lies in accurately identifying fine-grained sentiment intensity.



Subtask	Input	Output	Task Type
Intensity prediction	$S + a_1$	val_1-aro_1	Regression
	$S + a_2$	val_2-aro_2	
Triplet Extraction	S	$(a_1, o_1, val_1{-}aro_1)$	Extraction & Pairing & Regression
	S	$(a_2, o_2, val_2{-}aro_2)$	
Quadruple Extraction	S	$(a_1, c_1, o_1, val_1{-}aro_1)$	Extraction & Pairing & Classification & Regression
	S	$(a_2, c_2, o_2, val_2{-}aro_2)$	

Figure 1. Illustration of the three-dimensional aspect-based sentiment analysis (dimABSA) subtasks. In this visualization, aspect terms are highlighted in cyan, aspect categories in green, and opinion terms in red to facilitate clear identification. Additionally, the terms "*val*" and "*aro*" denote the valence and arousal intensities, respectively, which quantify the sentiment dimensions on a scale from 1 to 9.

Based on the definition of dimABSA tasks, Lee et al. [6] constructed a Chinese restaurant dimABSA dataset and organized a competition at the ACL SIGHAN 10 workshop. Our system [8] achieved first place among all participating teams. (This article is a revised and expanded version of a paper entitled "HITSZ-HLT at SIGHAN-2024 dimABSA Task: Integrating BERT and LLM for Chinese Dimensional Aspect-Based Sentiment Analysis", which was presented at the ACL SIGHAN 10 workshop in Bangkok, Thailand, on 16 August). Our participating system employs a hybrid approach. This approach synergizes the BERT model with a large language model (LLM), leveraging the strengths of both, which are two dominant paradigms for ABSA tasks. Specifically, we developed the BERT-based and LLM-based methods and explored their performance to underscore their advantages. The **BERT-based method** follows a pipeline paradigm, sequentially conducting aspect–opinion extraction, pairing and classification, and intensity prediction. To boost performance, we implemented three enhancements, i.e., domain-adaptive pre-training, negative pairs construction, and disabling BERT's internal dropout. The **LLM-based method** formulates dimABSA tasks into text generation tasks and fine-tunes a unified LLM through a multi-task learning strategy. Inspired by recent works [9,10], we designed code-style prompts to better align LLMs with extraction tasks. In addition, QLoRA [11] is utilized to reduce memory usage during model training.

We carried out preliminary experiments to investigate the advantages of the BERT-based and LLM-based methods. Our results yield two key observations. Firstly, for the task of extracting structures, specifically in aspect–opinion extraction and pairing, the BERT-based method significantly surpasses the LLM-based method. Secondly, in intensity prediction, the BERT-based method demonstrates superior performance with continuous values, whereas the LLM-based method is more effective in integer-level values. Consequently, we opted to utilize the BERT-based method for Subtask 1, i.e., intensity prediction. For Subtasks 2 and 3, we employed the BERT-based method to determine the aspects, categories, and opinions, which are subsequently input into an LLM to predict integer-level intensity.

Our contributions are summarized as follows:

- We introduce innovative solutions based on BERT and LLM for dimABSA tasks, along with a variety of strategies to optimize their effectiveness.
- We evaluate the advantages of the BERT-based and LLM-based methods across different tasks and devise a hybrid approach that leverages the advantages of both methods.
- We conduct comprehensive experiments on the dimABSA benchmark. Our results demonstrate that our hybrid approach achieves state-of-the-art performance. Further, ablation studies confirm the effectiveness of each component in our approach. We also provide detailed discussions to offer deeper insights into our findings.

2. Background

2.1. Aspect-Based Sentiment Analysis

Aspect-level Sentiment Classification (ASC) is the core task within ABSA, targeting the determination of the sentiment polarity towards the given aspect terms in a sentence [3]. Early methodologies primarily employed LSTM networks to learn contextual representations of sentences and incorporated attention mechanisms to model the interactions between aspect terms and their contexts [12–15]. Other approaches captured aspect-specific features from the contexts through relative position [16–18], memory networks [19,20], or gating mechanisms [21]. With the rise of pre-trained models [22,23], fine-tuning existing pre-trained models has emerged as the mainstream method for ASC tasks, with representative works including Sun et al. [24], Zhang et al. [25]. Subsequent research has explored post-training [26–28] or contrastive learning [29,30] to refine pre-trained language models for better representations. Additionally, some studies have incorporated syntactic information using graph neural networks [31,32]. Recent works have leveraged large language models (LLMs) to address ASC tasks, applying LLMs to ASC tasks using in-context learning [33,34], chain-of-thought prompting [35], or supervised fine-tuning [36,37]. Their results show LLMs can achieve performance comparable to current state-of-the-art (SOTA) methods without training. Additionally, some approaches generated sentiment explanations using LLMs and then utilized these explanations as supplementary features to enhance existing model training [38].

Aspect Sentiment Triplet Extraction. Aspect term extraction has been widely explored in previous research [39–43]. Opinion terms are crucial for extracting aspect terms and determining their associated sentiment polarities. This importance has spurred an increasing amount of research focused on the simultaneous extraction of both terms [44–47]. To explicitly delineate the relationship between aspect terms and opinion terms, Fan et al. [48] introduced the target-oriented opinion words extraction task, which focuses on extracting the opinion terms associated with a specific aspect term. Building on this, Peng et al. [49] proposed the aspect sentiment triplet extraction (ASTE) task, which aims to extract aspect terms along with their corresponding opinion terms and sentiment polarities. Subsequent research has approached this task through various methods, including formulating it into a reading comprehension problem [50–52], a span-relation extraction problem [53–55], a table-filling problem [56–58], and a sequence generation problem [59–61].

Aspect Sentiment Quad Prediction (ASQP) represents the most comprehensive task within ABSA. It is an extension of the ASTE task, seeking to predict all aspect-level quadruples from a review sentence [4,5], where each quadruple consists of an aspect term, aspect category, opinion term, and sentiment polarity. Existing ASQP methodologies can broadly be classified into two categories: discriminative methods and generative methods. Representative discriminative methods include Cai et al. [4] and Zhou et al. [62]. These methods utilize extract–classify techniques or table-based methods to extract aspect–category and opinion–sentiment pairs jointly. Generative methods cast the ASQP task to a text generation problem [5,63] or a tree generation problem [64,65]. Subsequent work augments the training of generative models by considering different permutations of quadruples [66,67], utilizing data augmentation techniques [68–71], or incorporating unlikelihood learning objectives [72]. In addition to these two methods, recent works attempted to apply LLMs to the ASQP task using in-context learning methods [73] or leveraged the rationales generated by the LLM as additional features to enhance existing model training [74].

Existing ABSA research treats sentiment as a three-class polarity rather than as a more granular sentiment intensity. This limitation inspired Lee et al. [6] to propose dimensional aspect-based sentiment analysis (dimABSA), which uses continuous real-valued scores in the valence–arousal dimensions to represent sentiment.

2.2. Dimensional Sentiment Analysis

Dimensional sentiment analysis offers a more nuanced approach to understanding sentiments expressed in text, moving beyond traditional positive, negative, or neutral classifications. Instead of categorizing sentiment into discrete classes, this method quantifies sentiment on continuous scales, typically across dimensions like valence (pleasantness) and arousal (intensity) [7]. Valence assesses whether an emotion is positive or negative, and arousal measures the level of excitement or calmness associated with an emotion. As shown in Figure 2, emotions can be mapped onto a two-dimensional plane using these two dimensions. The strength of this model lies in its ability to capture the continuity and interrelationships between different emotions rather than dividing them into distinct categories.

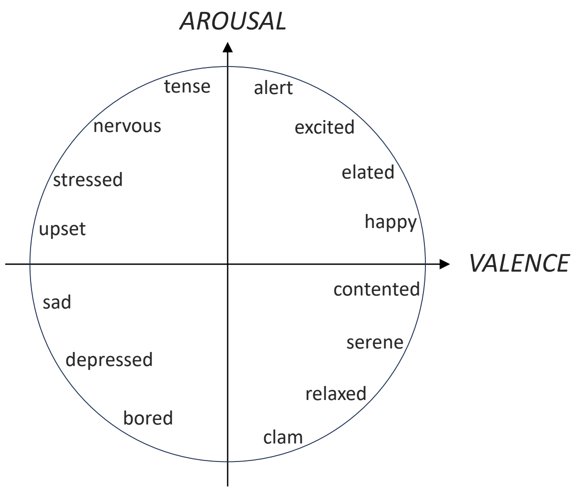


Figure 2. Russell's circumplex model of affect.

Prior research has developed many multi-dimensional affective resources, including lexicons [75] and sentence-level corpora [76,77]. Meanwhile, several studies produced multi-granularity Chinese dimensional sentiment resources, thereby addressing the gap in Chinese-language resources [78,79]. To effectively predict dimensional scores, early research predominantly utilized LSTM-based architectures. Notable implementations include a densely connected LSTM for phrase-level predictions [80], a relation interaction model for sentence-level predictions [81], and a regional CNN-LSTM model for text-level predictions [82,83]. With the rise of the Transformer architecture [84], researchers have increasingly adopted pre-trained language models to enhance performance. For example, Deng et al. [85] introduced a multi-granularity BERT fusion framework, and Wang et al. [86] proposed soft momentum contrastive learning for pre-training. Distinguishing our approach from these efforts, our work extends the use of LLMs for dimensional score prediction, enabling deeper exploration and discussion.

3. Task Definition

The dimABSA benchmark contains three subtasks: intensity prediction, triplet extraction, and quadruple extraction. They are formally defined as follows.

- **Subtask 1: Intensity prediction.** This task aims to predict sentiment intensities of given aspect terms in valence–arousal dimensions. The input includes a sentence $S = [w_1, w_2, \cdots, w_T]$ consisting of T words, along with a predefined aspect term a, which is a substring of the sentence. The output is the sentiment intensity, denoted as *val-aro*. As illustrated in Figure 1, given the sentence "吐柴主艺文总店除了餐点好吃之外，这里的用餐环境也很特别" (in English: "*Besides the tasty meals at the main art store of Tuchai, the dining environment here is also quite special*") and two aspect terms "餐点" (*meals*) and "用餐环境" (*dining environment*), this subtask requires systems to predict valence–arousal scores of 6.5#5.75 and 6.5#6.0, respectively.
- **Subtask 2: Triplet Extraction.** This task focuses on identifying aspect-level sentiments and opinions from given review sentences, outputting them as sets of triplets. The input is a sentence, and the corresponding output is a set containing all identified triplets. Each triplet consists of an aspect term a, an opinion term o, and sentiment intensity *val-aro*. For example, given the sentence "吐柴主艺文总店除了餐点好吃之外，这里的用餐环境也很特别" in Figure 1 (in English: "*Besides the tasty meals at the main art store of Tuchai, the dining environment here is also quite special*"), this subtask requires systems to produce the triplets {(餐点, 好吃, 6.5#5.75), (用餐环境, 很特别, 6.5#6.0)} (in English: {(*meals*, *tasty*, 6.5#5.75), (*dining environment*, *quite special*, 6.5#6.0)}).
- **Subtask 3: Quadruple Extraction.** This task builds on Subtask 2 by additionally requiring the identification of the aspect category, thus forming a quadruple. The aspect category falls within a predefined classification space, including 餐厅#概括 (restaurant#general), 餐厅#价格 (restaurant#prices), 餐厅#杂项 (restaurant#miscellaneous), 食物#价格 (food#prices), 食物#品质 (food#quality), 食物#份量与款式 (food#style&options), 饮料#价格 (drinks#prices), 饮料#品质 (drinks#quality), 饮料#份量与款式 (drinks#style&options), 氛围#概括 (ambience#general), 服务#概括 (services#general), and 地点#概括 (location#general). The specific meanings of each category can be found in the guideline [87]. For example, given the sentence in Figure 1, this subtask requires systems to produce the quadruples {(餐点, 食物#品质, 好吃, 6.5#5.75), (用餐环境, 氛围#概括, 很特别, 6.5#6.0)} (in English: {(*meals*, food#quality, *tasty*, 6.5#5.75), (*dining environment*, ambience#general, *quite special*, 6.5#6.0)}).

Valence and arousal scores are continuous values that range from 1 to 9. For these dimensions, a value of 1 indicates extremely negative and low-arousal sentiments, respectively. Conversely, a score of 9 signifies extremely positive and high-arousal sentiments. Meanwhile, a value of 5 represents a neutral sentiment and medium arousal. In Subtasks 2 and 3, the aspect term a can either be explicit substrings of the sentence S or be expressed implicitly, in which case they are denoted as "NULL". For instance, for the sentence "一口咬下去是真的很好吃" (in English: "*A bite is really delicious*"), the quadruple would be (NULL, 食物#品质, 真的很好吃, 7.33#7.0) (in English: (NULL, food#quality, *really delicious*, 7.33#7.0)).

4. Methods

This paper develops two solutions for the dimABSA task. The first solution is a pipeline approach leveraging the BERT model, while the second is an end-to-end solution that utilizes an LLM. We subsequently combine the strengths of both approaches to develop a hybrid method.

4.1. BERT-Based Method

We decompose the quadruple extraction of dimABSA into three subtasks: aspect–opinion extraction; aspect–opinion pairing and category classification; and intensity prediction. Additionally, we incorporate domain-adaptive pre-training to enhance the BERT model's domain awareness. Our framework is illustrated in Figure 3. A detailed description of each module follows.

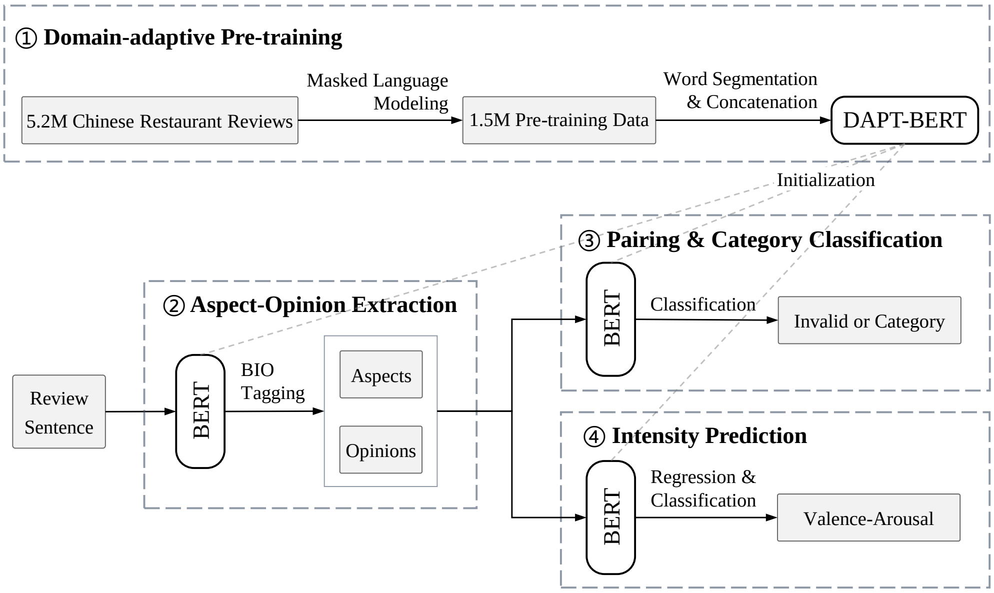


Figure 3. Overview of our BERT framework, consisting of four steps: (1) domain-adaptive pre-training, (2) aspect–opinion extraction, (3) aspect–opinion pairing and category classification, and (4) intensity prediction.

4.1.1. Domain-Adaptive Pre-Training

Previous research has demonstrated that pre-training language models on domain-specific sentiment-dense corpora can significantly improve their performance on downstream sentiment analysis tasks [26,28]. Consequently, we pre-train the BERT model on Chinese restaurant reviews before quadruple extraction, aiming to enhance its contextual understanding specific to the Chinese restaurant domain. Our data collection process contains three steps. Firstly, we collect 5.2 million open-source Chinese restaurant reviews. These reviews mainly consist of 4.4 million reviews from https://www.heywhale.com/mw/dataset/5e946de7e7ec38002d02d533/content accessed on 15 September 2024, 0.46 million reviews published on Li [88], and 0.33 million reviews from the AI Challenger 2018: sentiment analysis dataset (https://github.com/AIChallenger/AI_Challenger_2018/ accessed on 15 September 2024). Secondly, we cleanse these data to remove duplicates. Finally, we concatenate all the reviews and segment them based on the maximum sequence length, set at 512. This process yields a pre-training corpus containing 1.5 million samples.

After obtaining the pre-training corpus, we utilize the masked language modeling objective [22] to pre-train BERT. Unlike the static masking employed by Devlin et al. [22], we adopt the dynamic masking strategy proposed by Liu et al. [23], which randomly selects different tokens to mask in each training epoch. Additionally, we implement a whole-word masking strategy, which applies masking at the word level instead of the token level [89]. For Chinese word segmentation, we utilize the LTP tool [90]. We refer to our pre-trained BERT model as DAPT-BERT.

4.1.2. Aspect–Opinion Extraction

The objective of this step is to leverage our DAPT-BERT model to extract aspect terms and opinion terms from sentences. We employ the BIO tagging scheme to implement term extraction [91]. In this scheme, aspect terms and opinion terms in the sentences are represented by the token-level tags, with the tag space {`B-Aspect`, `I-Aspect`, `B-Opinion`, `I-Opinion`, `O`}. Subsequently, we superimpose a linear classifier on DAPT-BERT to identify these token-level tags. Furthermore, we augment the given sentence by prepending a special token [NULL] to identify implicit terms. This token is added to the vocabulary, and its embedding is initialized accordingly. The overall process can be formulated as follows:

$$\boldsymbol{h}_0, \boldsymbol{h}_1, \cdots, \boldsymbol{h}_T = \text{DAPT-BERT}(S'), \tag{1}$$

$$P(y_t) = \text{softmax}(\text{Linear}(\boldsymbol{h}_t)), \tag{2}$$

where $S' = [[\text{NULL}], w_1, \cdots, w_T]$ represents the augmented sentence and y_t denotes the tag for the t-th token in the sentence. We optimize the model along the linear classifier using a cross-entropy loss.

4.1.3. Aspect–Opinion Pairing and Category Classification

This step aims to match the aspect terms and opinion terms extracted in the previous step and to determine the aspect categories of each matched pair. We approach aspect–opinion pairing and category classification as a unified classification problem. The class space encompasses the aspect category space and includes an "unpaired' category. Specifically, we input the augmented sentence S' along with the aspect term a and opinion term o into DAPT-BERT to generate a discriminative representation, denoted as $\boldsymbol{h}_{[CLS]}$. Subsequently, $\boldsymbol{h}_{[CLS]}$ is fed into a linear classifier to determine the pairing and possible categories of the given aspect and opinion terms. We formulate this step as follows:

$$\boldsymbol{h}_{[CLS]} = \text{DAPT-BERT}(S', a, o), \tag{3}$$

$$P(c) = \text{softmax}(\text{Linear}(\boldsymbol{h}_{[\text{CLS}]})), \tag{4}$$

where $c \in$ {unpaired, 餐厅#概括, 餐厅#价格, 餐厅#杂项, 食物#价格, 食物#品质, 食物#份量与款式, 饮料#价格, 饮料#品质, 饮料#份量与款式, 氛围#概括, 服务#概括, 地点#概括} (in English: {restaurant#general, restaurant#prices, restaurant#miscellaneous, food#prices, food#quality, food#style& options, drinks#prices, drinks#quality, drinks#style&options, ambience#general, services#general, location#general}). We use the cross-entropy loss function to optimize the model and the linear classifier.

We introduce the strategy of **negative pairs construction** to mitigate error propagation, a common issue in pipeline methods. In the training phase, the input aspect and opinion terms are ground truths; however, during inference, they are predictions from the previous step, potentially containing errors. The discrepancy between training and inference can cause the classifier to be insensitive to minor boundary errors in the aspect and opinion terms, leading to further error propagation. To mitigate this, we construct negative pairs using incorrect aspect and opinion terms identified during the k-fold cross validation of the extraction model. These incorrect terms are then paired and labeled as "unpaired". We use these negative pairs, along with the correct terms, to train the relation model, thereby enhancing its ability to robustly handle errors.

4.1.4. Intensity Prediction

This step focuses on predicting the sentiment intensity of aspect–opinion pairs, namely, the valence–arousal scores. We develop two methods for this prediction. The first one approaches intensity prediction as a regression problem, employing a regression-based method. The second method transforms intensity prediction into a classification problem, subsequently utilizing a classification-based method. Detailed descriptions of these two methods are provided below.

The **regression-based method** inputs the sentence S' along with the aspect term a and opinion term o into DAPT-BERT. The resulting discriminative representation, $\boldsymbol{h}_{[CLS]}$, is then fed into two separate linear layers to predict the valence score s_{val} and arousal score s_{aro}, respectively. This process can be formulated as follows:

$$\boldsymbol{h}_{[CLS]} = \text{DAPT-BERT}(S', a, o), \tag{5}$$

$$\hat{s}_{val} = \text{Linear}(\boldsymbol{h}_{[\text{CLS}]}), \tag{6}$$

$$\hat{s}_{aro} = \text{Linear}(\boldsymbol{h}_{[\text{CLS}]}). \tag{7}$$

We calculate two losses using mean squared error (MSE) for each score and then compute the average to determine the overall regression loss. Furthermore, in the regression-based method, we employ the strategy of **disabling BERT's internal dropout**, a technique discussed in a Kaggle forum [92]. This approach is motivated by the concerns that BERT's internal dropout could introduce inconsistencies in the variance of neuron activations between training and inference phases, potentially compromising the numerical stability of the regression.

The **classification-based method** initially converts the continuous valence and arousal scores into categories through equidistant binning, with each bin having an interval of 0.25. Subsequently, we overlay two linear layers with softmax on DAPT-BERT to predict these categories, denoted as c_{val}, c_{aro}. This process can be formulated as follows:

$$\boldsymbol{h}_{[CLS]} = \text{DAPT-BERT}(S', a, o), \tag{8}$$

$$\hat{c}_{val} = \text{softmax}(\text{Linear}(\boldsymbol{h}_{[\text{CLS}]})), \tag{9}$$

$$\hat{c}_{aro} = \text{softmax}(\text{Linear}(\boldsymbol{h}_{[\text{CLS}]})). \tag{10}$$

We calculate two losses using the cross-entropy function for each category and then compute the average to determine the overall classification loss. During inference, we convert these categories back into scores. For example, if the predicted valence category is "[1–1.25]", the corresponding score after conversion is $\frac{1+1.25}{2} = 1.125$.

4.2. LLM-Based Method

Unlike the BERT-based method, the LLM-based method does not decompose quadruple extraction into multiple subtasks; instead, the LLM-based method transforms it into a text generation task for an end-to-end solution. We enhance the performance of the LLM through multi-tasking learning and code-style prompts [9]. Additionally, we employ the QLoRA [11] method to fine-tune the LLM under resource constraints. The overall framework is illustrated in Figure 4.

Multi-task learning: Wang et al. [93] highlighted that fine-tuning on similar multi-tasks can facilitate the capture of common structural information for information extraction tasks. Inspired by this insight, we develop a multi-task learning strategy for dimABSA. In addition to quadruple extraction, we incorporate five typical tasks: aspect term extraction, aspect intensity prediction, aspect–opinion intensity prediction, aspect–opinion–intensity triplet extraction, and aspect–category–opinion triplet extraction. The reason we chose these five tasks is that they are typical and complementary subtasks of quadruple extraction. Among them, aspect term extraction and aspect intensity prediction are fundamental subtasks that help the model enhance its understanding of basic aspect terms and sentiment intensity concepts. Aspect–opinion intensity prediction, which builds on aspect intensity prediction by incorporating opinion terms, enables the model to better grasp the relationship between sentiment intensity and opinion terms. The two triplet extraction tasks build upon the above subtasks, aiding the model in generating tuple-format outputs. We train the LLM by merging the data from these six tasks. We believe that such comprehensive training will enable the LLM to thoroughly acquire aspect-related knowledge.

Code-style prompt: LLMs are typically pre-trained on natural language. However, the output for quadruple extraction is a structured object, which deviates from the pre-training data. Li et al. [9] suggested that using code-style prompts can enhance the performance of

LLMs, as the structured nature of code more closely mirrors the requirements of information extraction tasks, and LLM pre-training corpora often include code snippets. Inspired by this, we apply code-style prompting in the dimABSA tasks. As shown in Figure 4, we convert the samples from dimABSA tasks into code-style instructions and outputs using the Python code format.

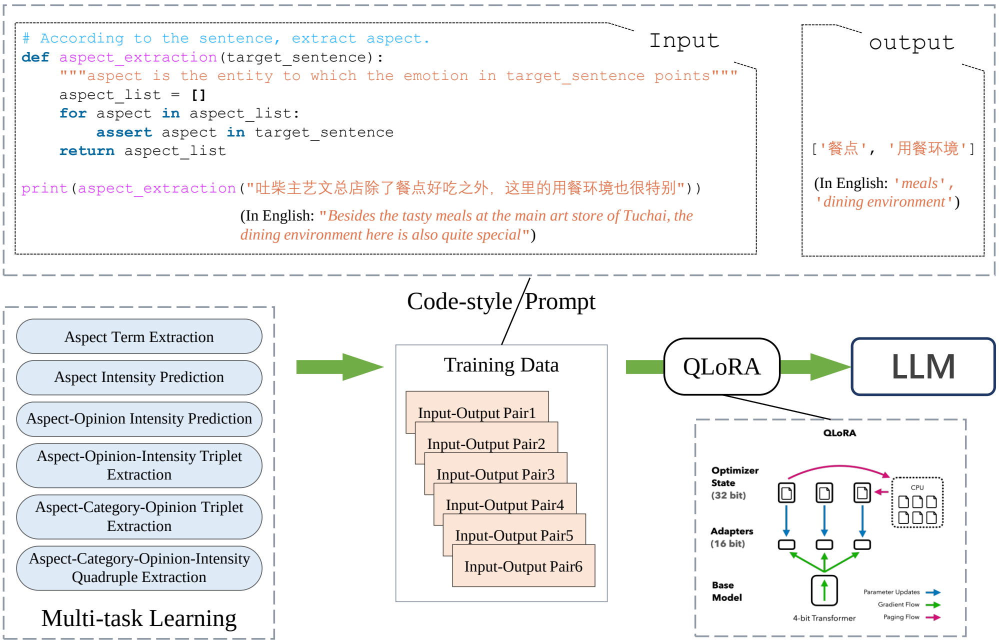


Figure 4. Overview of our LLM framework, which uses code-style prompts to build input–output pairs, trains the LLM on six tasks jointly, and optimizes the LLM via QLoRA methods.

Optimization with QLoRA: Fine-tuning LLMs is highly memory-intensive. To address this, we explore parameter-efficient fine-tuning methods. One such method, QLoRA [11], offers a novel and efficient fine-tuning approach that significantly reduces memory usage during the fine-tuning phase. QLoRA is an extension of the LoRA technique [94], which incorporates a small set of learnable low-rank adapters and then optimizes these adapters while keeping the original model weights unchanged. Building on LoRA, QLoRA introduces 4-bit NormalFloat formatting and double quantization techniques to quantize the model to 4 bits and introduces paged optimizers to prevent GPU memory overflow. Using QLoRA, we effectively optimize a 7b sized LLM on a 40G A100 GPU.

During inference, we set the temperature to 1 and use beam search for decoding, with the number of beams set to 2.

4.3. Ensemble Strategy

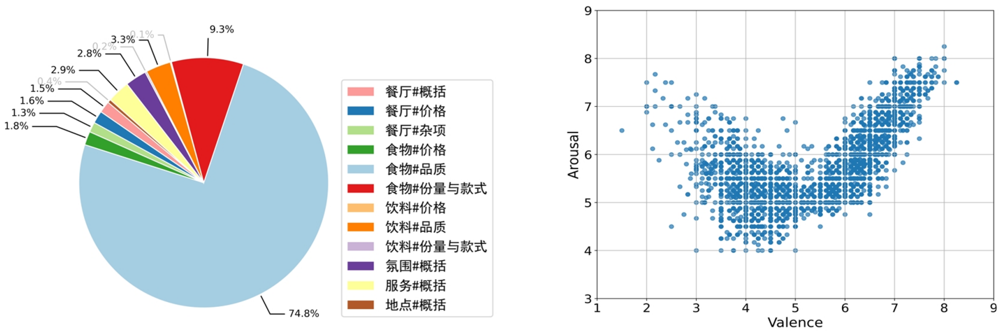
We carry out preliminary experiments to explore the comparative advantages of the BERT-based and LLM-based methods. Our results reveal that the BERT-based method excels in continuous intensity prediction and aspect–opinion extraction and pairing. We believe the reasons are as follows: (1) LLMs lack task-specific structures, preventing the model from learning task-specific representations; and (2) generating dimABSA labels in LLMs using an autoregressive approach is neither natural nor straightforward. In addition, the LLM-based method outperforms the BERT-based method in integer-level intensity prediction. We hypothesize that these differences stem from the inherent nature of LLMs, whose natural language generation output format may hinder precise comprehension and extraction of continuous values. However, LLMs tend to deliver superior results in coarse-grained predictions due to their larger parameter sizes.

We develop a hybrid approach that capitalizes on the advantages of both BERT-based and LLM-based methods. For Subtask 1, we combine the predictions from the regression and classification models in the BERT-based method by averaging them. For Subtasks 2 and 3, we employ the BERT-based method to extract aspect–category–opinion tuples. Subsequently, we feed all valid aspect–opinion pairs into the LLM, utilizing the aspect–opinion intensity prediction prompt to generate integer-level predictions for valence and arousal.

5. Experiments

5.1. Experimental Setup

The dataset used for the experiments originates from Lee et al. [6]. Initially, they collected restaurant reviews from Google reviews and the online bulletin board system PTT. Subsequently, annotators were organized to label aspects, categories, opinions, and intensities. To address inconsistencies in the annotations, for aspect–category–opinion, a majority vote method was applied, while sentiment intensity values were averaged after excluding samples with significant discrepancies. The dataset is available in two versions: one using Traditional Chinese characters and the other in Simplified Chinese characters. We select the Simplified Chinese version for our experiments. Table 1 presents the statistical information of this dataset. This dataset encompasses 12 predefined aspect categories, with their distribution depicted in Figure 5a. Notably, the "food#quality" category constitutes nearly three-quarters of the entire dataset. The distribution of valence–arousal scores across these categories is illustrated in Figure 5b.



(a) Category distribution. (b) Intensity distribution.

Figure 5. Data distribution: (**a**) category distribution, and (**b**) intensity distribution. The English translation of the non-English characters in subfigure (**a**) are restaurant#general, restaurant#prices, restaurant#miscellaneous, food#prices, food#quality, food#style& options, drinks#prices, drinks#quality, drinks#style&options, ambience#general, services#general, and location#general.

Table 1. Data statistics. The terms #Sent, #Char, and #Tuple denote the number of sentences, characters, and tuples in the dataset, respectively. Additionally, #Unique and #Repeat indicate the number of aspects or opinions that occur only once or more than once, respectively.

Subtask	**Dataset**	**#Sent**	**#Char**	**#Tuple**	**Aspect**			**Opinion**	
					#NULL	**#Unique**	**#Repeat**	**#Unique**	**#Repeat**
ST1	train	6050	85,769	8523	169	6430	1924	-	-
	dev	100	1.109	115	0	115	0	-	-
	test	2000	34,002	2658	0	2658	0	-	-
ST2 and ST3	train	6050	85,769	8523	169	6430	1924	7986	537
	dev	100	1280	150	0	78	72	143	7
	test	2000	39,014	3566	52	1693	1821	3263	303

Evaluation metrics: The performance of Subtask 1, i.e., intensity prediction, is assessed using the mean absolute error (MAE) and the Pearson correlation coefficient (PCC). These metrics measure the discrepancy between the model's predictions and the human-annotated scores. They are defined as follows:

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|, \tag{11}$$

$$\text{PCC} = \frac{\sum_{i=1}^{n}(y_i - \bar{y})(\hat{y}_i - \bar{\hat{y}})}{\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}\sqrt{\sum_{i=1}^{n}(\hat{y}_i - \bar{\hat{y}})^2}}, \tag{12}$$

where y_i and $\hat{y}_i$ represents the gold truth and prediction, n denotes the number of samples, and $\bar{y}$ and $\bar{\hat{y}}$ are the means of gold truths and predictions, respectively. The minimum value for MAE is 0, and the PCC ranges from -1 to 1. A lower MAE and higher PCC signify more accurate predictions.

The performance of Subtasks 2 and 3, namely, triplet and quadruple extraction, is evaluated using the F1-score. A tuple is considered correct only if all elements match the gold truth, where valence and arousal values are rounded to the nearest integer. The F1-score is calculated using the following equation:

$$\text{F1} = \frac{2 \times \text{P} \times \text{R}}{\text{P} + \text{R}}, \tag{13}$$

where P represents the precision, defined as the number of correct tuples divided by the total number of extracted tuples, and R represents the recall, defined as the number of correct tuples divided by the total number of gold tuples. Higher F1 values indicate better performance. Additionally, each metric is calculated independently for the valence and arousal dimensions or in combination.

Implementation details: For the BERT-based method, we utilize *ernie-3.0-xbase-zh* [95] as the backbone, which contains 296M parameters. The pre-training setup includes a batch size of 32, gradient accumulation steps of 12, bf16 mixed precision, a total of five training epochs, an initial learning rate of 1×10^{-4}, and a maximum sequence length of 512. During the fine-tuning phase, the learning rate is adjusted to 2×10^{-5}, with the batch size maintained at 32. The fine-tuning duration is set to seven epochs for the aspect–opinion extraction, pairing and classification, and the classification model for intensity prediction, and six for the regression model for intensity prediction. All models are fine-tuned using five different random seeds to ensure robustness, with the results aggregated through a voting mechanism.

For the LLM-based method, we employ *deepseek-7b-instruct-v1.5* [96] as the backbone. The fine-tuning setup includes the learning rate of 1×10^{-4}, five epochs, batch size of four, bf16 mixed precision, and maximum sequence length of 2048. Additionally, we set the rank of QLoRA fine-tuning to eight and the scaler factor to 16. LLM fine-tuning is implemented using the PyTorch framework on an NVIDIA A100 GPU.

Comparison models: We compare our hybrid approach with other participating systems, including yangnan, DS-Group [97], YNU-HPCC [98], TMAK-Plus [99], USTC-IAT, SUDA-NLP, BIT-NLP, JN-NLP [100], ZZU-NLP [101], and CCIIPLab [102]. Furthermore, we compare our hybrid approach with individual BERT-based and LLM-based methods, including the following: (1) $BERT_{REG}$, which utilizes the regression model for intensity prediction; (2) $BERT_{CLS}$, which employs the classification model for intensity prediction; (3) LLM_{INT}, which trains the LLM with integer-level intensity; and (4) LLM_{DEC}, which represents intensity using one decimal place.

5.2. Experimental Results

Comparison with other participating systems: Table 2 displays the performance of different participating systems in the ACL SIGHAN 10 shared task on dimABSA, as reported by Lee et al. [6]. Our method achieves the best results across all three subtasks, significantly outperforming other systems. Notably, in Subtasks 2 and 3, our VA-F1 scores are approximately three points higher than the second-best system. These results indicate that the proposed hybrid approach achieves state-of-the-art (SOTA) performance, demonstrating its effectiveness.

Table 2. Comparison results with other participating systems across three subtasks. V for valence, A for arousal, T for triplet, and Q for quadruple. We highlight the best results in bold. ↑ indicates that a higher value is better, while ↓ indicates that a lower value is better.

Methods	Subtask 1				Subtask 2			Subtask 3		
	V-MAE ↓	V-PCC ↑	A-MAE ↓	A-PCC ↑	V-T-F1 ↑	A-T-F1 ↑	VA-T-F1 ↑	V-Q-F1↑	A-Q-F1 ↑	VA-Q-F1 ↑
yangnan	1.032	0.877	1.095	0.097	-	-	-	-	-	-
DS-Group	0.460	0.858	0.501	0.490	-	-	-	-	-	-
YNU-HPCC	0.294	0.917	0.318	0.771	-	-	-	-	-	-
TMAK-Plus	-	-	-	-	0.269	0.307	0.157	-	-	-
USTC-IAT	-	-	-	-	-	-	-	0.438	0.437	0.312
SUDA-NLP	-	-	-	-	0.475	0.448	0.326	0.487	0.444	0.336
BIT-NLP	-	-	-	-	0.490	0.450	0.342	0.470	0.434	0.329
JN-NLP	-	-	-	-	-	-	-	0.482	0.439	0.331
ZZU-NLP	-	-	-	-	0.542	0.507	0.389	0.522	0.489	0.376
CCIIPLab	0.294	0.916	0.309	0.766	0.573	0.522	0.403	0.555	0.507	0.389
Ours	**0.279**	**0.933**	**0.309**	**0.777**	**0.589**	**0.545**	**0.433**	**0.567**	**0.526**	**0.417**

Comparison with BERT-based and LLM-based methods. Our hybrid approach integrates BERT-based and LLM-based methods. We compare this hybrid approach against individual methods and their variants. The results, as shown in Table 3, lead to the following four observations:

- Firstly, the hybrid approach outperforms the individual approaches on the majority of metrics, indicating that it effectively leverages the strengths of both BERT-based and LLM-based methods to achieve enhanced performance. Note that the A-Q-F1 metrics for the hybrid approach are slightly lower than those for BERT_{CLS}, indicating that the advantage of large model methods in arousal scores is relatively weak, as also reflected in the A-T-F1.
- Secondly, despite having significantly fewer parameters (296M) compared to the LLM-based method (7B), the BERT-based method exhibits superior performance across all metrics. We attribute this advantage to two main limitations of LLMs: (1) LLMs lack specific structures or designs to model the interactions among sentiment elements or between sentiment elements and context. This deficiency hinders the model's ability to learn task-specific representations. (2) The mapping from representations to dimABSA labels in LLMs is unnatural. Specifically, representing continuous valence–arousal scores as text reduces the semantic information inherent in the numerical values.
- Thirdly, within the BERT-based approaches, the regression model performs better in Subtask 1, while the classification model excels in Subtasks 2 and 3. This suggests that the regression model is more advantageous for fine-grained intensity assessments, whereas the classification model is more effective for coarse-grained intensity assessments.
- Finally, in the LLM-based methods, representing scores as decimals (LLM_{DEC}) yields better results in Subtask 1, while integer representations (LLM_{INT}) are more effective in Subtasks 2 and 3. This mirrors the conclusions drawn from the BERT-based methods.

Table 3. Comparison results between the hybrid approach and separate methods across three subtasks. V for valence, A for arousal, T for triplet, and Q for quadruple. We highlight the best results in bold. ↑ indicates that a higher value is better, while ↓ indicates that a lower value is better.

Methods	Subtask 1				Subtask 2			Subtask 3		
	V-MAE ↓	**V-PCC ↑**	**A-MAE ↓**	**A-PCC ↑**	**V-T-F1 ↑**	**A-T-F1 ↑**	**VA-T-F1 ↑**	**V-Q-F1 ↑**	**A-Q-F1 ↑**	**VA-Q-F1 ↑**
BERT_{REG}	0.287	0.930	0.311	0.773	0.574	0.526	0.405	0.555	0.511	0.393
BERT_{CLS}	**0.279**	0.930	0.316	0.766	0.583	0.543	0.425	0.564	**0.527**	0.411
LLM_{INT}	0.367	0.884	0.394	0.683	0.530	0.498	0.392	0.512	0.482	0.379
LLM_{DEC}	0.294	0.919	0.331	0.738	0.457	0.437	0.312	0.443	0.426	0.302
Hybrid approach	**0.279**	**0.933**	**0.309**	**0.777**	**0.589**	**0.545**	**0.433**	**0.567**	0.526	**0.417**

6. Discussion

We conduct further experiments and analyses to discuss our proposed approach in more depth, providing technical insights that will inform subsequent work.

6.1. Analysis of Ensemble Strategy

We conduct experiments to compare different ensemble strategies. We establish four ensemble strategies: (1) voting1, where predicted intensities from the regression model and classification model are averaged; (2) voting2, where predicted intensities from the regression model, classification model, and LLM-based method are averaged if the aspect–category–opinion tuples are consistent; (3) replace, where predicted intensities from the BERT-based methods are replaced with those from the LLM-based methods if the aspect–category–opinion tuples are consistent; (4) pipeline, where the BERT-based method outputs aspect–category–opinion tuples and these tuples are fed to the LLM to obtain the integer-level intensity. As shown in Table 4, we observe that the pipeline ensemble strategy achieves the best performance, meaning that the pipeline strategy can more effectively leverage the advantages of both BERT-based and LLM-based methods.

Table 4. Comparison results of different ensemble strategies on Subtask 3. We highlight the best results in bold.

Methods	Type	V-Q-F1	A-Q-F1	VA-Q-F1
Voting1	BERT	0.557	0.509	0.393
Voting2	BERT&LLM	0.563	0.526	0.413
Replace	BERT&LLM	0.565	**0.526**	0.416
Pipeline	BERT&LLM	**0.567**	**0.526**	**0.417**

6.2. Ablation Study

Ablation of the BERT-based method: In the BERT-based method, we implement three enhancements: domain-adaptive pre-training, negative pair construction, and disabling BERT's internal dropout. Therefore, we conduct experiments to investigate the effectiveness of these components, with the results presented in Table 5. Firstly, removing domain-adaptive pre-training results in a decrease across all metrics, confirming its effectiveness in enhancing the performance of the BERT model. Secondly, enabling BERT's internal dropout has a significant negative impact on MAE, suggesting that dropout indeed introduces instability in numerical values for the regression model. We also find that enabling BERT's internal dropout has a smaller effect on PCC, indicating that dropout's impact on correlation metrics is relatively minor. Lastly, the removal of negative aspect–opinion pairs also leads to performance declines in Subtasks 2 and 3, underscoring their necessity.

Table 5. Ablation results of the BERT-based method. *w/o* pre-training denotes removing domain-adaptive pre-training, *w/o* disabling-dropout represents enabling BERT's internal dropout, and *w/o* negative-pair means removing negative pair construction in pairing and classification. We highlight the best results in bold. ↑ indicates that a higher value is better, while ↓ indicates that a lower value is better.

Methods	Subtask 1				Subtask 2			Subtask 3		
	V-MAE ↓	V-PCC ↑	A-MAE ↓	A-PCC ↑	V-T-F1 ↑	A-T-F1 ↑	VA-T-F1 ↑	V-Q-F1 ↑	A-Q-F1 ↑	VA-Q-F1 ↑
$BERT_{REG}$	**0.287**	**0.930**	**0.311**	0.773	**0.574**	**0.526**	**0.405**	**0.555**	**0.511**	**0.393**
w/o pre-training	0.294	0.924	0.313	0.771	0.565	0.520	0.401	0.544	0.502	0.386
w/o disabling-dropout	0.337	0.933	0.348	**0.779**	0.537	0.503	0.365	0.521	0.487	0.354
w/o negative-pair	-	-	-	-	0.567	0.518	0.399	0.549	0.502	0.387

Ablation of the LLM-based method: In our LLM-based method, we transform dimABSA tasks into text generation tasks and employ two additional strategies to enhance performance: multi-task learning and code-style prompting. As shown in Table 6, eliminating multi-task learning results in a general performance decline, indicating that LLMs can benefit from improved generalization through multi-task learning. Furthermore, replacing code-style prompts with natural language prompts leads to significant performance reductions across all tasks, underscoring the importance of code-style prompts. During the inference phase, we use the beam search strategy for decoding. We observe that replacing this with greedy decoding also leads to a slight performance drop, confirming its necessity. We also observe that removing certain components sometimes results in slight improvements in a few metrics, which we attribute to the randomness in model training.

Table 6. Ablation results of the LLM-based method. *w/o* multi-task denotes removing multi-task learning, *w/o* code prompt represents replacing code-style prompts with natural language prompts, and *w/o* beam search means using greedy decoding. We highlight the best results in bold. ↑ indicates that a higher value is better, while ↓ indicates that a lower value is better.

Methods	Subtask 1				Subtask 2			Subtask 3		
	V-MAE ↓	V-PCC ↑	A-MAE ↓	A-PCC ↑	V-T-F1 ↑	A-T-F1 ↑	VA-T-F1 ↑	V-Q-F1 ↑	A-Q-F1 ↑	VA-Q-F1 ↑
LLM_{INT}	**0.367**	**0.884**	0.394	**0.683**	0.530	**0.498**	**0.392**	0.512	**0.482**	**0.379**
w/o multi-task	0.381	0.876	0.406	0.632	**0.535**	0.481	0.381	**0.514**	0.464	0.367
w/o code prompt	0.367	0.882	0.394	0.672	0.515	0.472	0.373	0.495	0.454	0.358
w/o beam search	0.377	0.880	**0.391**	0.670	0.531	0.489	0.388	0.511	0.472	0.374

6.3. Effect of Pre-Trained Language Models

The choice of pre-trained language models (PLMs) is a critical factor affecting performance. We select five representative Chinese PLMs for experimentation in Subtask 1, including *chinese-roberta-wwm-ext* and *chinese-roberta-wwm-ext-large* [89], *ernie-3.0-base-zh* and *ernie-3.0-xbase-zh* [95], and *erlangshen-deberta-v2-320m-chinese* [103]. The experimental results, shown in Table 7, indicate that models with larger numbers of parameters tend to perform better. Among these, *ernie-3.0-base-zh*, with a moderate parameter size, demonstrates superior performance, balancing training efficiency with excellent results for our system.

Table 7. Comparison of different pre-trained language models on Subtask 1. We highlight the best results in bold. ↑ indicates that a higher value is better, while ↓ indicates that a lower value is better.

Model	Params	Valence		Arousal	
		MAE ↓	PCC ↑	MAE ↓	PCC ↑
chinese-roberta-wwm-ext [89]	102M	0.300	0.918	0.310	0.766
ernie-3.0-base-zh [95]	118M	0.300	0.915	0.313	0.762
ernie-3.0-xbase-zh [95]	296M	0.286	0.926	**0.309**	**0.776**
erlangshen-deberta-v2-320m-chinese [103]	320M	**0.284**	**0.930**	0.310	0.774
chinese-roberta-ext-large [89]	326M	0.289	0.923	0.314	0.769

6.4. Error Analysis

We conduct an error analysis to understand the limitations of the proposed approach. As shown in Table 8, we find that errors related to aspect–opinion pairing and category classification are minimal, accounting for less than 7% of the total errors. However, errors in aspect and opinion terms are more substantial, comprising 40.46% of the errors. The most significant issue is sentiment intensity prediction errors, which contributed to 52.72% of the total errors. These findings suggest that future work should focus on improving term extraction and intensity prediction.

Table 8. The proportion of different errors in wrong predictions.

	Aspect	Opinion	Pairing	Category	Valence	Arousal
Error proportion	18.68%	21.78%	2.34%	4.48%	25.90%	26.82%

7. Conclusions and Future Works

This paper introduces a hybrid approach for dimensional aspect-based sentiment analysis (dimABSA), combining the strengths of BERT and LLM across various scenarios. Our method participated in the ACL SIGHAN 10 shared task, achieving the highest scores in multiple subtasks, thus validating the effectiveness of our proposed approach. Additionally, we conducted extensive experiments and provided technical discussions that contributed valuable insights and established a robust foundation for future inquiries. Although we conducted experiments solely on Chinese restaurant reviews, we believe that our approach can achieve promising results for dimABSA tasks in other languages and domains as well.

Despite its innovative integration of BERT and LLM for the dimABSA task and its notable performance, our study has several limitations. Firstly, it primarily explores ensemble methods such as voting and pipeline strategies, leaving more sophisticated integration techniques like knowledge distillation and the development of hybrid architectures unexplored. These approaches could potentially enhance performance by leveraging a wider array of benefits from both models. Secondly, our research is constrained by limited computational resources, which hampers our ability to deploy more advanced LLMs that may offer improved accuracy and generalization. Lastly, this study does not utilize existing dimensional sentiment resources, such as sentiment lexicons and annotated datasets, which could further refine sentiment dimension predictions. Future work should aim to incorporate these resources to augment the robustness and accuracy of sentiment analysis.

Author Contributions: Conceptualization, Y.Z. and H.X.; methodology, Y.Z., H.X., and D.Z.; software, H.X. and D.Z.; validation, H.X. and D.Z.; formal analysis, H.X.; investigation, Y.Z. and H.X.; resources, H.X.; data curation, H.X.; writing—original draft preparation, Y.Z. and H.X.; writing—review and editing, Y.Z.; visualization, Y.Z. and H.X.; supervision, R.X.; project administration, Y.Z.; funding acquisition, R.X. All authors have read and agreed to the published version of the manuscript.

Funding: This work was partially supported by the National Natural Science Foundation of China 62176076, Natural Science Foundation of Guangdong 2023A1515012922, the Shenzhen Foundational Research Funding JCYJ20220818102415032, the Major Key Project of PCL2021A06, and Guangdong Provincial Key Laboratory of Novel Security Intelligence Technologies 2022B1212010005.

Data Availability Statement: The dimABSA dataset used in this study is public at https://github.com/NYCU-NLP/SIGHAN2024-dimABSA (accessed on 15 September 2024).

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Medhat, W.; Hassan, A.; Korashy, H. Sentiment analysis algorithms and applications: A survey. *Ain Shams Eng. J.* **2014**, *5*, 1093–1113. [CrossRef]
2. Liu, B. Sentiment analysis and subjectivity. In *Handbook of Natural Language Processing*; Routledge: Abingdon-on-Thames, UK, 2010; Volume 2, pp. 627–666.
3. Pontiki, M.; Galanis, D.; Papageorgiou, H.; Androutsopoulos, I.; Manandhar, S.; AL-Smadi, M.; Al-Ayyoub, M.; Zhao, Y.; Qin, B.; De Clercq, O.; et al. SemEval-2016 Task 5: Aspect Based Sentiment Analysis. In Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016), San Diego, CA, USA, 16–17 June 2016; pp. 19–30. [CrossRef]
4. Cai, H.; Xia, R.; Yu, J. Aspect-Category-Opinion-Sentiment Quadruple Extraction with Implicit Aspects and Opinions. 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), Online, 1–6 August 2021; pp. 340–350. [CrossRef]
5. Zhang, W.; Deng, Y.; Li, X.; Yuan, Y.; Bing, L.; Lam, W. Aspect Sentiment Quad Prediction as Paraphrase Generation. In Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing, Online, 7–11 November 2021; pp. 9209–9219. [CrossRef]
6. Lee, L.H.; Yu, L.C.; Wang, S.; Liao, J. 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 (SIGHAN-10), Bangkok, Thailand, 11–16 August 2024; pp. 165–174.
7. Russell, J.A. A circumplex model of affect. *J. Personal. Soc. Psychol.* **1980**, *39*, 1161. [CrossRef]
8. Xu, H.; Zhang, D.; Zhang, Y.; Xu, R. HITSZ-HLT at SIGHAN-2024 dimABSA Task: Integrating BERT and LLM for Chinese Dimensional Aspect-Based Sentiment Analysis. In Proceedings of the 10th SIGHAN Workshop on Chinese Language Processing (SIGHAN-10), Bangkok, Thailand, 11–16 August 2024; pp. 175–185.
9. Li, P.; Sun, T.; Tang, Q.; Yan, H.; Wu, Y.; Huang, X.; Qiu, X. CodeIE: Large Code Generation Models are Better Few-Shot Information Extractors. In Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), Toronto, ON, Canada, 9–14 July 2023; pp. 15339–15353. [CrossRef]
10. Li, Z.; Zeng, Y.; Zuo, Y.; Ren, W.; Liu, W.; Su, M.; Guo, Y.; Liu, Y.; Li, X.; Hu, Z.; et al. KnowCoder: Coding Structured Knowledge into LLMs for Universal Information Extraction. In Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), Bangkok, Thailand, 11–16 August 2024; pp. 8758–8779.
11. Dettmers, T.; Pagnoni, A.; Holtzman, A.; Zettlemoyer, L. QLoRA: Efficient Finetuning of Quantized LLMs. In Proceedings of the Advances in Neural Information Processing Systems, New Orleans, LA, USA, 10–16 December 2023; Oh, A., Naumann, T., Globerson, A., Saenko, K., Hardt, M., Levine, S., Eds.; Curran Associates, Inc.: Red Hook, NY, USA, 2023; Volume 36, pp. 10088–10115.
12. Wang, Y.; Huang, M.; Zhu, X.; Zhao, L. Attention-based LSTM for Aspect-level Sentiment Classification. In Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, Austin, TX, USA, 1–5 November 2016; pp. 606–615. [CrossRef]
13. Ma, D.; Li, S.; Zhang, X.; Wang, H. Interactive attention networks for aspect-level sentiment classification. In Proceedings of the 26th International Joint Conference on Artificial Intelligence, Melbourne, Australia, 19–25 August 2017; pp. 4068–4074.
14. Liu, J.; Zhang, Y. Attention Modeling for Targeted Sentiment. In Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 2, Short Papers, Valencia, Spain, 3–7 April 2017; pp. 572–577.
15. Ma, Y.; Peng, H.; Cambria, E. Targeted Aspect-Based Sentiment Analysis via Embedding Commonsense Knowledge into an Attentive LSTM. *Proc. AAAI Conf. Artif. Intell.* **2018**, *32*, 5876–5883. [CrossRef]
16. Tang, D.; Qin, B.; Feng, X.; Liu, T. Effective LSTMs for Target-Dependent Sentiment Classification. In Proceedings of the COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers, Osaka, Japan, 11–16 December 2016; pp. 3298–3307.
17. Vo, D.T.; Zhang, Y. Target-dependent twitter sentiment classification with rich automatic features. In Proceedings of the 24th International Conference on Artificial Intelligence, Buenos Aires, Argentina, 25–31 July 2015; pp. 1347–1353.
18. Zhang, M.; Zhang, Y.; Vo, D.T. Gated Neural Networks for Targeted Sentiment Analysis. *Proc. AAAI Conf. Artif. Intell.* **2016**, *30*, 3087–3093. [CrossRef]
19. Tang, D.; Qin, B.; Liu, T. Aspect Level Sentiment Classification with Deep Memory Network. In Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, Austin, TX, USA, 1–5 November 2016; pp. 214–224. [CrossRef]

20. Fan, C.; Gao, Q.; Du, J.; Gui, L.; Xu, R.; Wong, K.F. Convolution-based Memory Network for Aspect-based Sentiment Analysis. In Proceedings of the SIGIR'18: 41st International ACM SIGIR Conference on Research & Development in Information Retrieval, New York, NY, USA, 8–12 July 2018; pp. 1161–1164. [CrossRef]
21. Xue, W.; Li, T. Aspect Based Sentiment Analysis with Gated Convolutional Networks. In Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), Melbourne, Australia, 15–20 July 2018; pp. 2514–2523. [CrossRef]
22. Devlin, J.; Chang, M.W.; Lee, K.; Toutanova, K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers), Minneapolis, MN, USA, 2–7 June 2019; pp. 4171–4186. [CrossRef]
23. Liu, Y.; Ott, M.; Goyal, N.; Du, J.; Joshi, M.; Chen, D.; Levy, O.; Lewis, M.; Zettlemoyer, L.; Stoyanov, V. RoBERTa: A Robustly Optimized BERT Pretraining Approach. *arxiv* **2019**, arXiv:1907.11692.
24. Sun, C.; Huang, L.; Qiu, X. Utilizing BERT for Aspect-Based Sentiment Analysis via Constructing Auxiliary Sentence. In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers), Minneapolis, MN, USA, 2–7 June 2019; pp. 380–385. [CrossRef]
25. Zhang, K.; Zhang, K.; Zhang, M.; Zhao, H.; Liu, Q.; Wu, W.; Chen, E. Incorporating Dynamic Semantics into Pre-Trained Language Model for Aspect-based Sentiment Analysis. In Proceedings of the Findings of Annual Meeting of the Association for Computational Linguistics—ACL, Dublin, Ireland, 22–27 May 2022; pp. 3599–3610.
26. Xu, H.; Liu, B.; Shu, L.; Philip, S.Y. BERT Post-Training for Review Reading Comprehension and Aspect-based Sentiment Analysis. In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers), Minneapolis, MN, USA, 2–7 June 2019; pp. 2324–2335.
27. Li, Z.; Zou, Y.; Zhang, C.; Zhang, Q.; Wei, Z. Learning Implicit Sentiment in Aspect-based Sentiment Analysis with Supervised Contrastive Pre-Training. In Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing, Online, 7–11 November 2021; pp. 246–256.
28. Zhang, Y.; Yang, Y.; Liang, B.; Chen, S.; Qin, B.; Xu, R. An Empirical Study of Sentiment-Enhanced Pre-Training for Aspect-Based Sentiment Analysis. In Proceedings of the Findings of the Association for Computational Linguistics: ACL 2023, Toronto, ON, Canada, 9–14 July 2023; pp. 9633–9651. [CrossRef]
29. Liang, B.; Luo, W.; Li, X.; Gui, L.; Yang, M.; Yu, X.; Xu, R. Enhancing aspect-based sentiment analysis with supervised contrastive learning. In Proceedings of the 30th ACM International Conference on Information & Knowledge Management, New York, NY, USA, 1–5 November 2021; pp. 3242–3247.
30. Cao, J.; Liu, R.; Peng, H.; Jiang, L.; Bai, X. Aspect is not you need: No-aspect differential sentiment framework for aspect-based sentiment analysis. In Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Seattle, WA, USA, 10–15 July 2022; pp. 1599–1609.
31. Wang, K.; Shen, W.; Yang, Y.; Quan, X.; Wang, R. Relational Graph Attention Network for Aspect-based Sentiment Analysis. In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics—ACL, Online, 5–10 July 2020; pp. 3229–3238.
32. Chen, C.; Teng, Z.; Wang, Z.; Zhang, Y. Discrete Opinion Tree Induction for Aspect-based Sentiment Analysis. In Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), Dublin, Ireland, 22–27 May 2022; pp. 2051–2064. [CrossRef]
33. Wang, Z.; Xie, Q.; Feng, Y.; Ding, Z.; Yang, Z.; Xia, R. Is ChatGPT a good sentiment analyzer? A preliminary study. *arXiv* **2023**, arXiv:2304.04339.
34. Xu, H.; Wang, Q.; Zhang, Y.; Yang, M.; Zeng, X.; Qin, B.; Xu, R. Improving In-Context Learning with Prediction Feedback for Sentiment Analysis. *arXiv* **2024**, arXiv:2406.02911.
35. Fei, H.; Li, B.; Liu, Q.; Bing, L.; Li, F.; Chua, T.S. Reasoning Implicit Sentiment with Chain-of-Thought Prompting. In Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers); Rogers, A., Toronto, ON, Canada, 9–14 July 2023; pp. 1171–1182. [CrossRef]
36. Simmering, P.F.; Huoviala, P. Large language models for aspect-based sentiment analysis. *arXiv* **2023**, arXiv:2310.18025.
37. Šmíd, J.; Priban, P.; Kral, P. LLaMA-Based Models for Aspect-Based Sentiment Analysis. In Proceedings of the 14th Workshop on Computational Approaches to Subjectivity, Sentiment, & Social Media Analysis, Bangkok, Thailand, 15 August 2024; pp. 63–70.
38. Wang, Q.; Ding, K.; Liang, B.; Yang, M.; Xu, R. Reducing Spurious Correlations in Aspect-based Sentiment Analysis with Explanation from Large Language Models. In Proceedings of the Findings of the Association for Computational Linguistics: EMNLP 2023, Singapore, 6–10 December 2023; pp. 2930–2941. [CrossRef]
39. Yin, Y.; Wei, F.; Dong, L.; Xu, K.; Zhang, M.; Zhou, M. Unsupervised word and dependency path embeddings for aspect term extraction. In Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence, New York, NY, USA, 9–15 July 2016; pp. 2979–2985.
40. Xu, H.; Liu, B.; Shu, L.; Yu, P.S. 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), Melbourne, Australia, 15–20 July 2018; pp. 592–598. [CrossRef]

41. Hu, M.; Peng, Y.; Huang, Z.; Li, D.; Lv, Y. Open-Domain Targeted Sentiment Analysis via Span-Based Extraction and Classification. In Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, Florence, Italy, 28 July–2 August 2019; pp. 537–546. [CrossRef]
42. Wei, Z.; Hong, Y.; Zou, B.; Cheng, M.; Yao, J. Don't Eclipse Your Arts Due to Small Discrepancies: Boundary Repositioning with a Pointer Network for Aspect Extraction. In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, Online, 5–10 July 2020; pp. 3678–3684. [CrossRef]
43. Wang, Q.; Wen, Z.; Zhao, Q.; Yang, M.; Xu, R. Progressive Self-Training with Discriminator for Aspect Term Extraction. In Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing, Online and Punta Cana, Dominican Republic, 7–11 November 2021; pp. 257–268. [CrossRef]
44. Wang, W.; Pan, S.J.; Dahlmeier, D.; Xiao, X. Recursive Neural Conditional Random Fields for Aspect-based Sentiment Analysis. In Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, Austin, TX, USA, 1–5 November 2016; pp. 616–626. [CrossRef]
45. Wang, W.; Pan, S.J.; Dahlmeier, D.; Xiao, X. Coupled Multi-Layer Attentions for Co-Extraction of Aspect and Opinion Terms. *Proc. AAAI Conf. Artif. Intell.* **2017**, *31*, 3316–3322. [CrossRef]
46. Li, X.; Lam, W. Deep Multi-Task Learning for Aspect Term Extraction with Memory Interaction. In Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, Copenhagen, Denmark, 9–11 September 2017; pp. 2886–2892. [CrossRef]
47. Li, X.; Bing, L.; Li, P.; Lam, W.; Yang, Z. Aspect term extraction with history attention and selective transformation. In Proceedings of the 27th International Joint Conference on Artificial Intelligence, Stockholm, Sweden, 13–19 July 2018; pp. 4194–4200.
48. Fan, Z.; Wu, Z.; Dai, X.Y.; Huang, S.; Chen, J. Target-oriented Opinion Words Extraction with Target-fused Neural Sequence Labeling. In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers), Minneapolis, MN, USA, 2–7 June 2019; pp. 2509–2518. [CrossRef]
49. Peng, H.; Xu, L.; Bing, L.; Huang, F.; Lu, W.; Si, L. Knowing What, How and Why: A Near Complete Solution for Aspect-Based Sentiment Analysis. *Proc. AAAI Conf. Artif. Intell.* **2020**, *34*, 8600–8607.
50. Chen, S.; Wang, Y.; Liu, J.; Wang, Y. Bidirectional Machine Reading Comprehension for Aspect Sentiment Triplet Extraction. *Proc. AAAI Conf. Artif. Intell.* **2021**, *35*, 12666–12674. [CrossRef]
51. Mao, Y.; Shen, Y.; Yu, C.; Cai, L. A Joint Training Dual-MRC Framework for Aspect Based Sentiment Analysis. *Proc. AAAI Conf. Artif. Intell.* **2021**, *35*, 13543–13551. [CrossRef]
52. Zhai, Z.; Chen, H.; Feng, F.; Li, R.; Wang, X. COM-MRC: A COntext-Masked Machine Reading Comprehension Framework for Aspect Sentiment Triplet Extraction. In Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing, Abu Dhabi, United Arab Emirates, 7–11 December 2022; pp. 3230–3241. [CrossRef]
53. Li, Y.; Lin, Y.; Lin, Y.; Chang, L.; Zhang, H. A span-sharing joint extraction framework for harvesting aspect sentiment triplets. *Knowl.-Based Syst.* **2022**, *242*, 108366. .: 10.1016/j.knosys.2022.108366 [CrossRef]
54. Chen, Y.; Keming, C.; Sun, X.; Zhang, Z. A Span-level Bidirectional Network for Aspect Sentiment Triplet Extraction. In Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing, Abu Dhabi, United Arab Emirates, 7–11 December 2022; pp. 4300–4309. [CrossRef]
55. Xu, L.; Chia, Y.K.; Bing, L. 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), Online, 1–6 August 2021; pp. 4755–4766. [CrossRef]
56. Wu, Z.; Ying, C.; Zhao, F.; Fan, Z.; Dai, X.; Xia, R. Grid Tagging Scheme for Aspect-oriented Fine-grained Opinion Extraction. In Proceedings of the Findings of the Association for Computational Linguistics: EMNLP 2020, Online, 16–20 November 2020; pp. 2576–2585. [CrossRef]
57. Chen, H.; Zhai, Z.; Feng, F.; Li, R.; Wang, X. 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), Dublin, Ireland, 22–27 May 2022; pp. 2974–2985. [CrossRef]
58. Zhang, Y.; Yang, Y.; Li, Y.; Liang, B.; Chen, S.; Dang, Y.; Yang, M.; Xu, R. Boundary-Driven Table-Filling for Aspect Sentiment Triplet Extraction. In Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing, Abu Dhabi, United Arab Emirates, 7–11 December 2022; pp. 6485–6498. [CrossRef]
59. Yan, H.; Dai, J.; Ji, T.; Qiu, X.; Zhang, Z. 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), Online, 1–6 August 2021; pp. 2416–2429. [CrossRef]
60. Lu, Y.; Liu, Q.; Dai, D.; Xiao, X.; Lin, H.; Han, X.; Sun, L.; Wu, H. Unified Structure Generation for Universal Information Extraction. In Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), Dublin, Ireland, 22–27 May 2022; pp. 5755–5772. [CrossRef]
61. Zhang, W.; Li, X.; Deng, Y.; Bing, L.; Lam, W. Towards Generative 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 2: Short Papers), Online, 1–6 August 2021; pp. 504–510. [CrossRef]

62. Zhou, J.; Yang, H.; He, Y.; Mou, H.; Yang, J. A Unified One-Step Solution for Aspect Sentiment Quad Prediction. In Proceedings of the Findings of the Association for Computational Linguistics: ACL 2023, Toronto, ON, Canada, 9–14 July 2023; pp. 12249–12265. [CrossRef]
63. Qin, Y.; Lv, S. Generative Aspect Sentiment Quad Prediction with Self-Inference Template. *Appl. Sci.* **2024**, *14*, 6017. [CrossRef]
64. Bao, X.; Wang, Z.; Jiang, X.; Xiao, R.; Li, S. Aspect-based Sentiment Analysis with Opinion Tree Generation. In Proceedings of the 31st International Joint Conference on Artificial Intelligence—IJCAI, Vienna, Austria, 23–29 July 2022; Volume 2022; pp. 4044–4050.
65. Mao, Y.; Shen, Y.; Yang, J.; Zhu, X.; Cai, L. Seq2Path: Generating Sentiment Tuples as Paths of a Tree. In Proceedings of the Findings of the Association for Computational Linguistics: ACL 2022, Dublin, Ireland, 22–27 May 2022; pp. 2215–2225. [CrossRef]
66. Hu, M.; Wu, Y.; Gao, H.; Bai, Y.; Zhao, S. Improving Aspect Sentiment Quad Prediction via Template-Order Data Augmentation. In Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing, Abu Dhabi, United Arab Emirates, 7–11 December 2022; pp. 7889–7900. [CrossRef]
67. Gou, Z.; Guo, Q.; Yang, Y. MvP: Multi-view Prompting Improves Aspect Sentiment Tuple Prediction. In Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), Toronto, ON, Canada, 9–14 July 2023; pp. 4380–4397. [CrossRef]
68. Zhang, W.; Zhang, X.; Cui, S.; Huang, K.; Wang, X.; Liu, T. Adaptive Data Augmentation for Aspect Sentiment Quad Prediction. In Proceedings of the ICASSP 2024—2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Seoul, Republic of Korea, 14–19 April 2024; pp. 11176–11180. [CrossRef]
69. Yu, Y.; Zhao, M.; Zhou, S. Boosting Aspect Sentiment Quad Prediction by Data Augmentation and Self-Training. In Proceedings of the 2023 International Joint Conference on Neural Networks (IJCNN), Gold Coast, Australia, 18–23 June 2023; pp. 1–8. [CrossRef]
70. Wang, A.; Jiang, J.; Ma, Y.; Liu, A.; Okazaki, N. Generative Data Augmentation for Aspect Sentiment Quad Prediction. In Proceedings of the 12th Joint Conference on Lexical and Computational Semantics (*SEM 2023), Toronto, ON, Canada, 13–14 July 2023; pp. 128–140. [CrossRef]
71. Zhang, Y.; Zeng, J.; Hu, W.; Wang, Z.; Chen, S.; Xu, R. Self-Training with Pseudo-Label Scorer for Aspect Sentiment Quad Prediction. In Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), Bangkok, Thailand, 11–16 August 2024; pp. 11862–11875.
72. Hu, M.; Bai, Y.; Wu, Y.; Zhang, Z.; Zhang, L.; Gao, H.; Zhao, S.; Huang, M. Uncertainty-Aware Unlikelihood Learning Improves Generative Aspect Sentiment Quad Prediction. In Proceedings of the Findings of the Association for Computational Linguistics: ACL 2023, Toronto, ON, Canada, 9–14 July 2023; pp. 13481–13494. [CrossRef]
73. Xu, X.; Zhang, J.D.; Xiao, R.; Xiong, L. The Limits of ChatGPT in Extracting Aspect-Category-Opinion-Sentiment Quadruples: A Comparative Analysis. *arxiv* **2023**, arXiv:2310.06502.
74. Kim, J.; Heo, R.; Seo, Y.; Kang, S.; Yeo, J.; Lee, D. Self-Consistent Reasoning-based Aspect-Sentiment Quad Prediction with Extract-Then-Assign Strategy. *arXiv* **2024**, arXiv:2403.00354.
75. Warriner, A.B.; Kuperman, V.; Brysbaert, M. Norms of valence, arousal, and dominance for 13,915 English lemmas. *Behav. Res. Methods* **2013**, *45*, 1191–1207. [CrossRef]
76. Preoţiuc-Pietro, D.; Schwartz, H.A.; Park, G.; Eichstaedt, J.; Kern, M.; Ungar, L.; Shulman, E. Modelling valence and arousal in facebook posts. In Proceedings of the 7th workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis, San Diego, CA, USA, 16 June 2016; pp. 9–15.
77. Buechel, S.; Hahn, U. 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, Valencia, Spain, 3–7 April 2017; pp. 578–585.
78. Yu, L.C.; Lee, L.H.; Hao, S.; Wang, J.; He, Y.; Hu, J.; Lai, K.R.; Zhang, X. Building Chinese affective resources in valence-arousal dimensions. In Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, San Diego, CA, USA, 12–17 June 2016; pp. 540–545.
79. Lee, L.H.; Li, J.H.; Yu, L.C. Chinese EmoBank: Building valence-arousal resources for dimensional sentiment analysis. *ACM Trans. Asian-Low-Resour. Lang. Inf. Process.* **2022**, *21*, 1–18. [CrossRef]
80. Wu, C.; Wu, F.; Huang, Y.; Wu, S.; Yuan, Z. Thu_ngn at ijcnlp-2017 task 2: Dimensional sentiment analysis for chinese phrases with deep lstm. In Proceedings of the IJCNLP 2017, Shared Tasks, Taipei, Taiwan, 27 November–1 December 2017; pp. 47–52.
81. Xie, H.; Lin, W.; Lin, S.; Wang, J.; Yu, L.C. A multi-dimensional relation model for dimensional sentiment analysis. *Inf. Sci.* **2021**, *579*, 832–844. [CrossRef]
82. Wang, J.; Yu, L.C.; Lai, K.R.; Zhang, X. Dimensional sentiment analysis using a regional CNN-LSTM model. In Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers) Berlin, Germany, 7–12 August 2016; pp. 225–230.
83. Wang, J.; Yu, L.C.; Lai, K.R.; Zhang, X. Tree-structured regional CNN-LSTM model for dimensional sentiment analysis. *IEEE/ACM Trans. Audio Speech Lang. Process.* **2019**, *28*, 581–591. [CrossRef]
84. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, Ł.; Polosukhin, I. Attention is all you need. *Adv. Neural Inf. Process. Syst.* **2017**, *30*, 6000–6010.
85. Deng, Y.C.; Wang, Y.R.; Chen, S.H.; Lee, L.H. Towards Transformer Fusions for Chínese Sentiment Intensity Prediction in Valence-Arousal Dimensions. *IEEE Access* **2023**, *11*, 109974–109982. [CrossRef]

86. Wang, J.; Yu, L.C.; Zhang, X. SoftMCL: Soft Momentum Contrastive Learning for Fine-grained Sentiment-aware Pre-training. In Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024), Torino, Italy, 20–25 May 2024; pp. 15012–15023.
87. Pontiki, M.; Galanis, D.; Papageorgiou, H.; Androutsopoulos, I.; Manandhar, S.; Al-Smadi, M.; Al-Ayyoub, M.; Zhao, Y.; Qin, B.; De Clercq, O.; et al. SemEval 2016 Task 5 Aspect Based Sentiment Analysis (ABSA-16) Annotation Guidelines. 2016. Available online: https://alt.qcri.org/semeval2016/task5/data/uploads/absa2016_annotationguidelines.pdf (accessed on 14 September 2024).
88. Li, T. Restaurant Review Data on Dianping.com. 2018. Available online: https://opendata.pku.edu.cn/dataset.xhtml?persistentId=doi:10.18170/DVN/GCIUN4 (accessed on 15 September 2024).
89. Cui, Y.; Che, W.; Liu, T.; Qin, B.; Yang, Z. Pre-Training With Whole Word Masking for Chinese BERT. *IEEE/ACM Trans. Audio Speech, Lang. Process.* **2021**, *29*, 3504–3514. [CrossRef]
90. Che, W.; Li, Z.; Liu, T. LTP: A Chinese language technology platform. In *Coling 2010: Demonstrations*; Coling 2010 Organizing Committee: Beijing, China, 2010; pp. 13–16.
91. Ramshaw, L.A.; Marcus, M.P. Text Chunking Using Transformation-Based Learning. In *Natural Language Processing Using Very Large Corpora*; Armstrong, S., Church, K., Isabelle, P., Manzi, S., Tzoukermann, E., Yarowsky, D., Eds.; Springer: Dordrecht, The Netherlands, 1999; pp. 157–176. [CrossRef]
92. Deotte, C. The Magic of No Dropout. 2021. Available online: https://www.kaggle.com/competitions/commonlitreadabilityprize/discussion/260729 (accessed on 1 August 2024).
93. Wang, X.; Zhou, W.; Zu, C.; Xia, H.; Chen, T.; Zhang, Y.; Zheng, R.; Ye, J.; Zhang, Q.; Gui, T.; et al. InstructUIE: Multi-task Instruction Tuning for Unified Information Extraction. *arxiv* 2023, arXiv:2304.08085.
94. Hu, E.J.; Shen, Y.; Wallis, P.; Allen-Zhu, Z.; Li, Y.; Wang, S.; Wang, L.; Chen, W. LoRA: Low-Rank Adaptation of Large Language Models. In Proceedings of the International Conference on Learning Representations, Online, 25–29 April 2022.
95. Sun, Y.; Wang, S.; Feng, S.; Ding, S.; Pang, C.; Shang, J.; Liu, J.; Chen, X.; Zhao, Y.; Lu, Y.; et al. Ernie 3.0: Large-scale knowledge enhanced pre-training for language understanding and generation. *arXiv* **2021**, arXiv:2107.02137.
96. Guo, D.; Zhu, Q.; Yang, D.; Xie, Z.; Dong, K.; Zhang, W.; Chen, G.; Bi, X.; Wu, Y.; Li, Y.; et al. DeepSeek-Coder: When the Large Language Model Meets Programming–The Rise of Code Intelligence. *arXiv* **2024**, arXiv:2401.14196.
97. Meng, L.a.; Zhao, T.; Song, D. DS-Group at SIGHAN-2024 dimABSA Task: Constructing In-context Learning Structure for Dimensional Aspect-Based Sentiment Analysis. In Proceedings of the 10th SIGHAN Workshop on Chinese Language Processing (SIGHAN-10), Bangkok, Thailand, 11–16 August 2024; pp. 127–132.
98. Wang, Z.; Zhang, Y.; Wang, J.; Xu, D.; Zhang, X. YNU-HPCC at SIGHAN-2024 dimABSA Task: Using PLMs with a Joint Learning Strategy for Dimensional Intensity Prediction. In Proceedings of the 10th SIGHAN Workshop on Chinese Language Processing (SIGHAN-10), Bangkok, Thailand, 11–16 August 2024; pp. 96–101.
99. Kang, X.; Zhang, Z.; Zhou, J.; Wu, Y.; Shi, X.; Matsumoto, K. TMAK-Plus at SIGHAN-2024 dimABSA Task: Multi-Agent Collaboration for Transparent and Rational Sentiment Analysis. In Proceedings of the 10th SIGHAN Workshop on Chinese Language Processing (SIGHAN-10), Bangkok, Thailand, 11–16 August 2024; pp. 88–95.
100. Jiang, Y.; Lu, H.Y. JN-NLP at SIGHAN-2024 dimABSA Task: Extraction of Sentiment Intensity Quadruples Based on Paraphrase Generation. In Proceedings of the 10th SIGHAN Workshop on Chinese Language Processing (SIGHAN-10), Bangkok, Thailand, 11–16 August 2024; pp. 121–126.
101. Zhu, S.; Zhao, H.; Wxr, W.; Jia, Y.; Zan, H. ZZU-NLP at SIGHAN-2024 dimABSA Task: Aspect-Based Sentiment Analysis with Coarse-to-Fine In-context Learning. In Proceedings of the 10th SIGHAN Workshop on Chinese Language Processing (SIGHAN-10), Bangkok, Thailand, 11–16 August 2024; pp. 112–120.
102. Tong, Z.; Wei, W. CCIIPLab at SIGHAN-2024 dimABSA Task: Contrastive Learning-Enhanced Span-based Framework for Chinese Dimensional Aspect-Based Sentiment Analysis. In Proceedings of the 10th SIGHAN Workshop on Chinese Language Processing (SIGHAN-10), Bangkok, Thailand, 11–16 August 2024; pp. 102–111.
103. Zhang, J.; Gan, R.; Wang, J.; Zhang, Y.; Zhang, L.; Yang, P.; Gao, X.; Wu, Z.; Dong, X.; He, J.; et al. Fengshenbang 1.0: Being the Foundation of Chinese Cognitive Intelligence. *arXiv* **2022**, arXiv:2209.02970.