Modeling Adaptive Inter-Task Feature Interactions via Sentiment-Aware Contrastive Learning for Joint Aspect-Sentiment Prediction

Wei Chen[1], **Yuxuan Liu**[2*], **Zhao Zhang**[3], **Fuzhen Zhuang**[1,4†], **Jiang Zhong**[2]

[1]Institute of Artificial Intelligence, Beihang University, Beijing 100191, China
[2]College of Computer Science, Chongqing University, Chongqing 400044, China
[3]Institute of Computing Technology, Chinese Academy of Sciences, Beijing, China
[4]Zhongguancun Laboratory, Beijing, China
{chenwei23,zhuangfuzhen}@buaa.edu.cn, {liuyuxuan,zhongjiang}@cqu.edu.cn, zhangzhao2021@ict.ac.cn

Abstract

Aspect prediction (AP) and sentiment prediction (SP) are representative applications in fine-grained sentiment analysis. They can be considered as sequential tasks, where AP identifies mentioned aspects in a sentence, and SP infers fine-grained sentiments for these aspects. Recent models perform the aspect-sentiment prediction in a joint manner, but heavily rely on the feature interactions of aspect and sentiment. One drawback is that they ignore correlation strength varies between aspect features and sentiment features across different sentences, and employ a fixed feature interaction strategy may limit effective knowledge transfer across tasks. To tackle this issue, in this paper, we propose an Adaptive Inter-task Feature Interaction framework, **AIFI**, for joint aspect-sentiment prediction. Specifically, we introduce a novel contrast-based alignment method based on contrastive learning. Our approach considers the AP-specific and SP-specific representations of a given sentence as a positive pair, while representation of another random sentence serves as a negative example. Moreover, we propose an inter-task feature correlation network to predict the contrast strength, which is determined by the temperature coefficient in the InfoNCE loss. This dynamic correlation adjustment enhances model's ability to capture proper feature interactions more efficiently. Experimental results on three datasets validate the effectiveness of our approach.

	Sentences	Sub-Task AP	Sub-Task SP	Correlation Strength
(1)	The Windows 7 seems to be quite expensive.	Windows 7	Negative	★★
(2)	I asked for seltzer with lime, no ice.	seltzer with lime	Neutral	★
(3)	The food is delicious – from the specials to regular menu-fare, the dishes are never a disappointment.	food	Positive	★★★
		specials	Positive	
		regular menu-fare	Positive	
		dishes	Positive	

Figure 1: Examples of ALSA tasks from real-world datasets. The aspects are framed by rectangles and opinion words are marked in blue. And the greater the number of five-pointed stars, the stronger the correlation.

Introduction

In comparison to traditional sentiment analysis tasks, aspect-level sentiment analysis (ALSA) requires a more in-depth examination of the context and the extraction of more detailed information (Liu 2012; Zhang et al. 2022b). ALSA involves several sub-tasks: aspect prediction (AP), opinion prediction (OP), and aspect term sentiment prediction (SP). AP identifies aspects or features in the context (Lv et al. 2023), OP extracts expressions conveying sentiment about the aspects (Liu et al. 2023), and SP determines the sentiments expressed towards each aspect (Liang et al. 2023). Sometimes, opinion terms are not explicitly provided, requiring the ALSA system to predict aspects and sentiments directly from the context without predefined opinion terms. In this study, we concentrate on jointly predicting aspects and sentiments from the sentences, as shown in Figure 1.

The initial approach to ALSA is employing a two-stage approach (Fan et al. 2019; Hu et al. 2019). Firstly, it accomplishes the AP and is subsequently followed by another model to perform SP. Nevertheless, this strategy could potentially disrupt the interaction between aspects and sentiments and may also lead to error propagation problems.

Recent efforts aim to achieve the two sub-tasks cohesively, predicting aspects and corresponding sentiments jointly or interactively through a unified methodology (Lin and Yang 2020; Lv et al. 2021; Chen et al. 2022b; Zhang et al. 2022c; Liang et al. 2023; Cui et al. 2023). These studies have revealed a strong mutual implication between these two sub-tasks, and have devised some interaction strategies to facilitate the knowledge exchange. One of the extensively researched and proven effective techniques in this field is regularization alignment approach (Chen et al. 2022b,a). It involves two key steps: first, encoding aspect- and sentiment-specific features for each task, and second, aligning these task-specific features (i.e., maximize the similarity) within the latent space. In this way, the aspect words simultaneously encode both aspect- and sentiment-related features,

[*]This work is done during his internship at vivo.
[†]Corresponding author.

and vice versa. However, this approach, while effective, may not always distinguish aspect- and sentiment-specific features well, potentially leading to negative transfer problem. For instance, in the movie scenario, suppose that the majority of the reviews regarding Spielberg's films tend to be positive. When predicting the aspect and sentiment of the review "I watched a Spielberg's film.", the regularization approach might amplify the associations between aspect and sentiment features in the semantic space due to its rigid alignment strategy. This may lead to aspect "Spielberg's film" being consistently associated with a positive sentiment, despite the absence of explicit sentimental words in this sentence, which could inevitably affect the prediction.

To address this issue, we present a contrast-based alignment framework based on contrastive learning (Oord, Li, and Vinyals 2018), which has proven successful in other multi-task learning tasks (Lin et al. 2022; Bai et al. 2022). It aims to encourage learning sufficient inter-task interactive features while mitigating negative transfer issues by pulling closer the aspect and sentiment features in the semantic space belonging to the same sentence, pushing away features with different sentences. In the proposed framework, the temperature parameter τ regulates the contrast strength between aspect and sentiment features (Wang and Liu 2021; Lin et al. 2022). When there is a strong correlation between aspect and sentiment features, we assign a smaller value to τ. Conversely, when the correlation is weak (e.g., "I watched a Spielberg film."), we assign it a larger value. By adeptly fine-tuning this parameter τ, the model can alleviate the aforementioned pseudo-correlation pattern to some extent, generating proper interaction features, and thereby effectively reducing the negative knowledge transfer issue.

Generally, the temperature parameter τ is assumed to be fixed in contrast-based alignment method. While manually adjusting τ can yield satisfactory results across datasets, the correlation strength between aspect and sentiment features still manifests differently across various sentences. Intuitively, the more sentimental words there are in a sentence, the stronger the mutual implications between aspect and sentiment features, as sentimental words are often associated with specific domains or types of aspects. Take Figure 1 for example, **(1)** in the 1st sentence, the appearance of word "expensive" in the laptop dataset would likely be related to some electronic product, helping the AP task in predicting aspect "windows 7". **(2)** However, in the 2nd sentence with neutral sentiment, due to there is an absence of obvious sentimental words, each sub-task may rely more on contextual clues. **(3)** Additionally, in the 3rd sentence, where several sentiment words are directly related to aspects, these mutually indicative cues can assist in predicting the corresponding aspects and sentiments more quickly. These examples illustrate it is necessary to design distinct correlation intensities between task-specific features for different sentences.

Towards this end, we further propose a sentiment-aware contrastive learning framework to explicitly model adaptive feature interactions, named **AIFI**. More specifically, we have observed the significant role of sentiment information in indicating the aspect-sentiment correlations, thus, we establish a feature correlation network to predict the correlation intensity between task-specific features. This enables us to adaptively set the temperature for contrastive learning in a fine-grained manner, improving the capture of inter-task corrections. The experimental results on three datasets show our model's effectiveness. Overall, this work contributes to three major aspects: **(1)**. We propose a novel contrast-based alignment method based on contrastive learning, which is able to learn sufficient inter-task interactive features while avoiding negative transfer issues. To the best of our knowledge, this is the first work that applies contrastive learning to address the problem of joint aspect-sentiment prediction. **(2)**. We establish a feature correlation network to predict the contrast strength. This allows us to dynamically adjust the correlation intensity in a highly detailed manner, thereby enhancing the ability to capture inter-task correlations. Besides, our approach can be generally extended to other multi-task learning (MTL) problems as well. **(3)** Experimental results on three real-world datasets show that the proposed framework achieves state-of-the-art performances.

Related Work

In this section, we provide a brief overview of the relevant literature, focusing on two areas: aspect-level sentiment analysis and contrastive learning.

Aspect-Level Sentiment Analysis. ALSA requires extracting the aspect terms with their corresponding sentiment polarities in the sentence, which is an active research topic in recent years. Early studies treat them as two separate tasks and use some prevalent algorithms such as Conditional Random Fields (CRF) (Li and Lu 2017), Long Short-Term Memory (LSTM) (Chen et al. 2020) and Bidirectional Encoder Representations from Transformers (BERT) (Yang et al. 2020) to complete the task. Meanwhile, some researchers have explored end-to-end solutions for aspects and sentiments extraction using the multi-task learning framework (Luo et al. 2019; Lin and Yang 2020; Chen et al. 2022b; Liang et al. 2023). They emphasize the close relationship between aspect and sentiment and propose various interaction approaches, such as shared-private feature interaction. (Lin and Yang 2020), feature combination interaction (Lv et al. 2021), and task-regularization interaction (Chen et al. 2022b). Although successful, most of these methods overlook the varying degree of feature correlations in different sentences, making it challenging to transfer knowledge effectively across tasks. Recently, large language models have had a profound impact on the entire NLP community due to their impressive representation abilities (Zhou et al. 2023). Large models like ChatGPT are trained using general-purpose data and possess strong generalization abilities. However, in low-resource ALSA task, while they may outperform fine-tuned BERT in certain cases, they still lag behind state-of-the-art (SOTA) models (Wang et al. 2023).

Contrastive Learning. Recently, numerous studies have been applying contrastive learning techniques to NLP tasks, yielding promising results (Giorgi et al. 2021; Zhang et al. 2022a,d,a; Rethmeier and Augenstein 2023; Lingling et al. 2023). In the task of aspect-level sentiment classification, several insightful researchers have explored incorporating contrastive learning into the model training process.

Liang (Liang et al. 2021) employ contrastive learning to capture aspect-invariant and aspect-dependent features, thereby distinguishing the roles of valuable sentiment features and achieving sentiment classification. Xiong (Xiong et al. 2022) propose a triplet contrastive learning network that effectively coordinates syntactic and semantic information. They construct an aspect-oriented sub-tree to replace the syntactic adjacency matrix, and then introduce a sentence-level contrastive learning scheme to emphasize the features of sentiment words. To the best of our knowledge, there is no existing work that integrates contrastive learning into aspect and sentiment extraction. Thus, in this work, we propose an innovative adaptive contrastive learning framework aimed at jointly predicting aspects and sentiments.

Methodology

In this section, we introduce our proposed approach AIFI for joint aspect-sentiment prediction in Figure 2. First, we introduce the task formulation and the backbone architecture. Then, we present the adaptive learning method. Finally, we show the optimization details.

Problem Formulation

In this work, we aim to tackle the task of jointly predicting aspects and their corresponding sentiments. As shown in prior studies (Chen et al. 2022b; Liang et al. 2023; Liu et al. 2023), AP and SP are highly related when applied to the same sentence, and thus a MTL solution can potentially benefit for both sub-task. It is important to note that our approach is not restricted to these two sub-tasks alone, but can be generally extended to other MTL problems as well.

Formally, given an input sentence $\boldsymbol{x} = \{w_1, w_2, \ldots, w_n\}$, where w_i represents the i^{th} word in the sentence, and n is the sentence's length. Our goal is to identify and extract aspects mentioned in the sentence, along with their corresponding sentiment labels (i.e., positive, negative, or neutral), denoted as $\mathrm{O} = \{(a_1, s_1), (a_2, s_2), \ldots, (a_m, \mathrm{s}_m)\}$, where a_i represents the i-th aspects, and s_i represents the sentiment label associated with a_i, m is the aspect number of a sentence.

Base Structure

Following (Chen et al. 2022b; Lv et al. 2023), we adopt pre-trained BERT language model (Devlin et al. 2019) as the underlying encoder to generate contextualized clause representations. Given a sentence $\boldsymbol{x} = \{w_1, w_2, \ldots, w_n\}$, the output of sentence representation is $\boldsymbol{h} = \{\boldsymbol{e}_1, \boldsymbol{e}_2, \ldots, \boldsymbol{e}_n\} \in \mathbb{R}^{n \times d}$, where d represents the embedding size.

After obtaining the embedding representation $\boldsymbol{h}$, we leverage an expert network (Ma et al. 2018) to capture task-specific features for each branch $\boldsymbol{h}^{ap}$ and $\boldsymbol{h}^{sp}$. This allows our model to possess increased flexibility in capturing intricate patterns of relatedness across tasks. Specifically, each standalone task is optimization via:

$$\begin{aligned} \hat{p}_s, \hat{p}_e &= f_{ap}(g(\cdot)), \\ \hat{y} &= f_{sp}(g(\cdot)), \end{aligned} \tag{1}$$

where f_{ap}, f_{sp} and $g(\cdot)$ indicate the AP-specific layer, SP-specific layer and expert network respectively. In this equation, $\hat{p}_s$ and $\hat{p}_e$ represent the probability of being an aspect boundary (i.e., start and end position) in AP task. Once obtaining the aspect boundary probabilities, we can employ a heuristic extraction algorithm (Lin and Yang 2020; Chen et al. 2022b) to extract the corresponding aspects. $\hat{y}$ represents the predicted sentiment probabilities in SP task.

Learning Objective. In training phase, each task has its own learning objective. For AP, we need to minimize the cross-entropy loss of aspect boundary:

$$\mathcal{L}_{ap} = \text{CrossEntropy}(p_s, \hat{p}_s) + \text{CrossEntropy}(p_e, \hat{p}_e), \tag{2}$$

where p_s and p_e are the ground truths of the start and end boundaries (i.e., 0-1 vectors), $\hat{p}_s$ and $\hat{p}_e$ represent the predicted boundary scores. Besides, for SP, the parameters are learned based on:

$$\mathcal{L}_{sp} = -\sum_{i=1}^{N}\sum_{j=1}^{m} y_{i,j} \log \hat{y}_{i,j}, \tag{3}$$

where $\hat{y}_{i,j}$ represents the predicted sentiment class and $y_{i,j}$ represents the corresponding ground truth. N is the training samples and m is the number of sentiment polarities.

Actually, capturing task correlations between AP and SP can be challenging, and merely optimizing the loss functions described in Eq. (2) and Eq. (3) is insufficient. It's essential to find the right balance between sharing information across tasks and preserving task-specific elements. Over-sharing information might lead to interference or negative impact between the sub-tasks, while sharing too little information could limit the potential benefits derived from the inter-task relationships. Based on the base structure, we next present our solution by further modeling the inter-task interactions in an adaptive and fine-grained manner.

Modeling Adaptive Inter-Task Interactions

Inter-task Feature Alignment In this section, we aim to enhance the inter-task interactions. As mentioned above, the expert network is a component of a multi-task learning model that aims to leverage shared knowledge across different tasks while allowing each task to have its own specific features. Besides the expert network, we further learn the alignment relation between aspect-specific and sentiment-specific features for knowledge transfer across tasks.

A straightforward way to implement this is to directly align aspect- and sentiment-related features in latent space (Chen et al. 2022b) through a regularization strategy, such as KL divergence or mean squared error (MSE). Although this regularization alignment strategy is commonly used in existing multi-task learning framework (Lee, Yang, and Hwang 2016; Huang et al. 2023), it doesn't ensure proper discrimination between aspect and sentiment features, potentially leading to the negative transfer problem.

Thus, to address this issue, we propose to model the inter-task feature interactions using a more precise alignment approach based on contrastive learning. In comparison to regularization-based strategy, contrastive learning proves to be more potent and flexible in establishing associations across distinct representation spaces (Wang and Isola 2020). Specifically, when presented with a sentence representation $\boldsymbol{h}_{\boldsymbol{x}}$, we exchange its output representations with those from

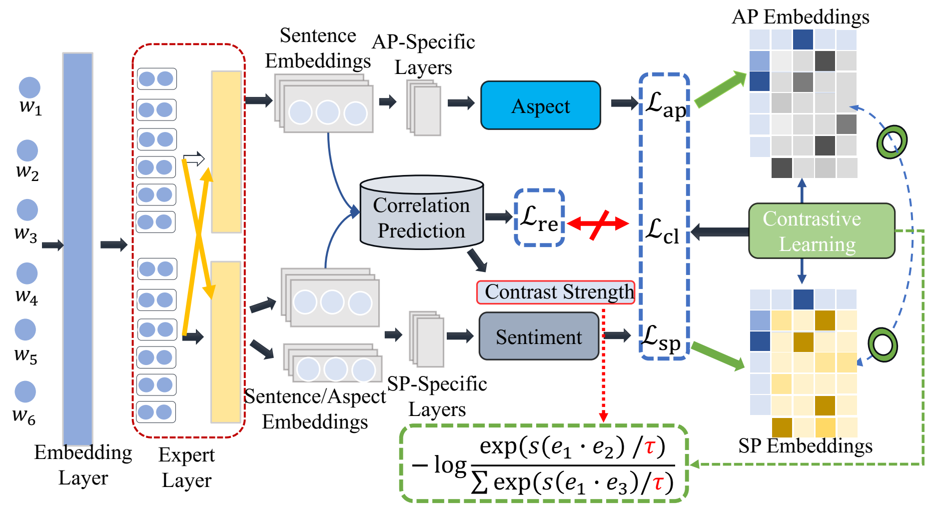

Figure 2: The architecture of our proposed framework AIFI. The red double arrow indicates that the gradient is cut-off.

another task as positive pairs. Meanwhile, we consider the output representation from a different sentence representation $\boldsymbol{h}_{\boldsymbol{x}'}$ as a negative pair. To implement this, we utilize in-batch negatives for contrastive learning:

$$\mathcal{L}_{cl} = -\sum_{\boldsymbol{x}\in\mathcal{D}} \log \frac{\exp\left(\text{sim}(\boldsymbol{h}_{\boldsymbol{x}}^{ap} \cdot \boldsymbol{h}_{\boldsymbol{x}}^{sp})/\tau\right)}{\sum_{\boldsymbol{x}'\in\mathcal{B}} \exp\left(\text{sim}(\boldsymbol{h}_{\boldsymbol{x}}^{ap} \cdot \boldsymbol{h}_{\boldsymbol{x}'}^{sp})/\tau\right)}, \tag{4}$$

where $\text{sim}(\cdot)$ measures the similarity between two vectors, which is set as cosine similarity function, $\mathcal{D}$ denotes the dataset, and $\mathcal{B}$ symbolizes a dataset batch. τ is a temperture coefficient that regulates the contrast strength between two representations.

In this equation, τ is usually set as a fixed value. In fact, it is important to note that the strength of feature correlations between AP and SP varies across different sentences. As a result, our subsequent objective involves exploring a more adaptive approach to establish the temperature coefficient.

Adaptive Inter-Task Feature Alignment Essentially, two representations of data in contrastive learning are compared based on a similarity metric, and the temperature parameter plays a crucial role in this process. The smaller the temperature used in this process, the more strict the constraint becomes, and vice versa (Wang and Liu 2021; Zhang et al. 2021). Consequently, our goal is to find an optimal balance in the temperature parameter in an adaptive manner. We want the model to be sensitive enough to capture similarities between task-specific features, while also being robust enough to avoid overgeneralization and negative knowledge transfer. Hence, we propose to construct a feature correlation network and utilize the estimated relatedness to adaptively adjust the temperature coefficient.

Particularly, we first obtain the task-specific representations ($\boldsymbol{h}^{ap}$ and $\boldsymbol{h}^{sp}$) for AP and SP respectively. The $\boldsymbol{h}^{ap}$ aims to summarize the aspect-relevant features from the sentence, while $\boldsymbol{h}^{sp}$ captures the sentiment-related information in the embedding space. Then, the feature combinations will be subsequently transformed by MLP layers and a softmax function to generate the predicted correlation strength:

$$\hat{r} = \text{softmax}\left(\text{MLP}\left(\boldsymbol{h}^{ap} \otimes \boldsymbol{h}^{sp}\right)\right), \tag{5}$$

where $\otimes$ represents element-wise product. $\hat{r}$ is the predicted relatedness score.

Intuitively, the greater the number of emotional words in a sentence, the stronger the mutual implications between aspect and sentiment features, as such words are commonly linked to distinct domains or types of aspects. For this reason, we adopt the sentiment polarity of each sentence as the ground truth for correlation model. More precisely, we gauge the correlation of task-specific features by assessing the quantity of positive and negative sentiment polarity within the sentence. Formally, the ground truth is defined as:

$$r^{la} = \sum_{i=1}^{m} \left(\mathbb{I}\left[s_{\text{i}} = pos\right] + \mathbb{I}\left[s_{\text{i}} = neg\right]\right), \tag{6}$$

where m is the number of aspect in each sentence. $\mathbb{I}$ is an indicator function, which returns 1 if the value is true, and 0 otherwise. $s_i = pos$ indicates the sentiment of aspect i is positive. However, the label range covered by r^{la} may quite extensive (i.e., some sentences contain more than a dozen aspect words related to non-neutral sentiments), which may impede network optimization. To expedite this process, we handle the predicted scores greater than two separately:

$$r = \begin{cases} 0, & \text{if } r^{la} = 0 \\ 1, & \text{if } r^{la} = 1 \\ 2, & \text{if } r^{la} \geq 2. \end{cases} \tag{7}$$

Based on the predicted score $\hat{r}$ and ground truth r, we leverage the CrossEntropy loss to optimize the softmax classifier:

$$\mathcal{L}_{re} = -\sum_{i=1}^{N} \sum_{j=1}^{j=3} r_{i,j} \log \hat{r}_{i,j}, \tag{8}$$

Algorithm 1: Training Process

Input: Training sample $\mathcal{D}$, learning rate α, inter-task feature relatedness parameters Φ, task-specific parameters Θ

while *Converged == False* **do**
  for $\mathcal{B}$ *in Dataloader*($\mathcal{D}$) **do**
    Obtaining model prediction:
    Obtain $\hat{p}_s$, $\hat{p}_e$ and $\hat{y}$ via Eq. (1);
    Obtain $\hat{r}$ via Eq. (5);
    Select task representation $\boldsymbol{h}_{ap}$ and $\boldsymbol{h}_{sp}$;
    Optimization parameter Φ:
    Constructe r based on Eq. (7);
    Calcuate $\mathcal{L}_{re}$ based on Eq. (8);
    Update $\Phi = \Phi - \alpha \nabla_{\Phi} \mathcal{L}_{re}$;
    Optimization parameter Θ:
    Calcuate $\mathcal{L}_{ap}$ and $\mathcal{L}_{sp}$ via Eq. (2) and Eq. (3);
    Calcuate L_{cl} via Eq. (10) ;
    $\mathcal{L} = \mathcal{L}_{ap} + \mathcal{L}_{\text{sp}} + \mathcal{L}_{cl}$;
    Update $\Theta = \Theta - \alpha \nabla_{\Theta} \mathcal{L}$;
  end
end

where N is the number of sentence.

However, the correlation strength is positively correlated with $\hat{r}$, which contradicts the definition of the temperature parameter in Eq. (4). Furthermore, $\hat{r}$ is a discrete value and cannot be directly applied to the temperature parameter in contrastive learning. Accordingly, after obtaining the predicted relatedness score $\hat{r}$, we utilize linear interpolation (Blu, Thévenaz, and Unser 2004; Lin et al. 2022) to adjust the temperature parameter within a reasonable range:

$$\tau_r = \tau^{min} + (\tau^{max} - \tau^{min}) \times (1 - \frac{\mathbb{E}(\hat{r})}{2}), \tag{9}$$

where τ^{min} and τ^{max} represent the minimum and maximum pre-set values for the temperature parameter, respectively. $\mathbb{E}(\hat{r})$ is the expectation of predicted relatedness score.

With the learned τ_r, Eq. (4) can be rewritten as:

$$\mathcal{L}_{cl} = -\sum_{\boldsymbol{x} \in \mathcal{D}} \log \frac{\exp\left(\text{sim}(\boldsymbol{h}_{\boldsymbol{x}}^{ap} \cdot \boldsymbol{h}_{\boldsymbol{x}}^{sp})/\tau_r\right)}{\sum_{\boldsymbol{x}' \in \mathcal{B}} \exp\left(\text{sim}(\boldsymbol{h}_{\boldsymbol{x}}^{ap} \cdot \boldsymbol{h}_{\boldsymbol{x}'}^{sp})/\tau_r\right)}. \tag{10}$$

Overall Optimization

We train the proposed framework by jointly minimizing the sum of the aforementioned four losses:

$$\mathcal{L} = \mathcal{L}_{ap} + \lambda_1 \mathcal{L}_{\text{sp}} + \lambda_2 \mathcal{L}_{re} + \lambda_3 \mathcal{L}_{cl}, \tag{11}$$

where λ_1, λ_2 and λ_3 are hyperparameters to control corresponding strengths. In the training phase, it's worth noting that the gradients produced by the contrastive loss solely update task-specific parameters, while the optimization of the feature correlation model is governed by $\mathcal{L}_{re}$. Algorithm 1 depicts the entire optimization process.

Dataset	#Sentences	#Pair	#POS	#NEG	#NEU
$\mathbb{D}_{\text{R}}$	3900	6603	4134	1538	931
$\mathbb{D}_{\text{L}}$	1869	2936	1326	900	620
$\mathbb{D}_{\text{T}}$	2350	3243	703	274	2266

Table 1: Statistics of three datasets. "Pair" denotes the number of aspect-sentiment pair.

Experiment

In this section, we conduct experiments to address the following five research questions:

RQ1. How do the proposed method AIFI perform when compared to baseline methods? **RQ2.** Do AIFI model really help in the effective prediction of aspect and corresponding sentiments? **RQ3.** Which part of the AIFI model majorly contributed to the effective prediction of optimal ALSA? **RQ4.** Is it possible to achieve improved performance by manually fine-tuning τ in real-world situations? **RQ5.** Do adaptive interactions between task-specific facilitate settling the challenging that are mentioned above?

Datasets

We conduct experiments on three publicly available datasets, i.e., $\mathbb{D}_{\text{R}}$, $\mathbb{D}_{\text{L}}$, and $\mathbb{D}_{\text{T}}$. The $\mathbb{D}_{\text{R}}$ is a combination of restaurant review sets from SemEval2014, SemEval2015, and SemEval2016 (Pontiki et al. 2014, 2015, 2016). The $\mathbb{D}_{\text{L}}$ dataset comprises customer reviews in the electronic product domain, which were collected from the SemEval Challenge 2014 (Pontiki et al. 2014). The $\mathbb{D}_{\text{L}}$ consists of twitter posts from different users (Mitchell et al. 2013). For $\mathbb{D}_{\text{R}}$ and $\mathbb{D}_{\text{L}}$ datasets, we maintain the official data division for the training, validation, and testing sets. The results reported are the averaged scores of 10 runs. As there is no standard train-test split available for $\mathbb{D}_{\text{T}}$, we follow (Chen et al. 2022b) and report the ten-fold cross-validation results. The statistics of these three datasets are presented in Table 1.

Methods Evaluated

We compare our model with the following baselines:

SPAN-BERT (Hu et al. 2019) propose a pipeline model for ALSA. It employs BERT as its backbone network, and a multi-target extractor is utilized to detect the aspect boundaries. Subsequently, a polarity classifier is applied to determine the sentiment polarity of each aspect. **DOER** (Luo et al. 2019) present a dual cross-shared recurrent neural network for ALSA. which incorporates a cross-shared unit to effectively analyze the interplay between AP and SP, thus achieving promising results. **SPRM** (Lin and Yang 2020) is a span-based model to address AP and SP simultaneously. It utilizes a combination of private and shared representations (i.e., expert network) to address the ALSA task, with the objective of capturing the interconnections between the tasks. **S-AESC** (Lv et al. 2021) is a shared span-based learning framework that utilize both dual gated recurrent units and an interaction layer to model the inter-task correlations. **HI-ASA** (Chen et al. 2022b) develop a hierarchical multi-task framework that integrates expert interactions and task-level regularization interactions to enhance the associations

Dataset	Method	Precision	Recall	F_1
$\mathbb{D}_{\mathrm{R}}$	DOER$^\diamond$	0.8032	0.6654	0.7278
	S-AEAC$^\diamond$	0.7826	0.7050	0.7418
	SPAN-BERT$^\diamond$	0.7614	0.7334	0.7492
	SPRM†	0.7754	0.7945	0.7846
	HI-ASA†	0.7859	0.7942	0.7879
	ChatGPT†	0.7023	0.6531	0.6762
	AIFI (Ours)	**0.7919***	**0.8065***	**0.7991***
$\mathbb{D}_{\mathrm{L}}$	DOER$^\diamond$	0.6143	0.5931	0.6035
	S-AEAC$^\diamond$	0.6687	0.6492	0.6588
	SPAN-BERT$^\diamond$	0.6946	0.6672	0.6806
	SPRM†	0.6854	0.6872	0.6851
	HI-ASA†	0.6877	0.6872	0.6871
	ChatGPT†	0.4921	0.4832	0.4872
	AIFI (Ours)	**0.7019***	**0.6909***	**0.6963***
$\mathbb{D}_{\mathrm{T}}$	DOER$^\diamond$	0.5554	0.5479	0.5137
	S-AEAC$^\diamond$	0.5586	0.5374	0.5473
	SPAN-BERT$^\diamond$	0.6072	0.5502	0.5769
	SPRM†	0.6002	0.5825	0.5912
	HI-ASA†	0.6140	0.5879	0.6004
	ChatGPT†	0.4325	0.4114	0.4216
	AIFI (Ours)	**0.6322***	**0.5953***	**0.6130***

Table 2: Overall performance comparison on three datasets. The results with $\diamond$ are taken from their original papers. The † means that we reproduce the models using released code with original parameters on the dataset. And the bold indicates the best results, $*$ denotes the statistical significance for p-value $\leq$ 0.05 compared with HI-ASA.

between AP and SP. It is the current state-of-the-art model for joint aspect-sentiment prediction. **ChatGPT** (Wang et al. 2023) is a recently launched large language model (LLM), which has shown superior performance in various natural language processing (NLP) tasks. In this study, we employ ChatGPT as aspect extractor and sentiment classifier to retrieve the semantic information from the raw text corpus. We carefully develop task-specific prompts (i.e., few-shot learning) to activate its comprehension capabilities.

Evaluation Measures and Configurations

Following previous studies, we employ precision, recall, and $F1$ to assess the ALSA performance of our model. For AP task, $F1$ is also utilized as the evaluation metric. Additionally, we adopt accuracy as the metric for SP.

In the experiment setting, we employ the BERT-Large model as the basic network architecture, comprising 24 transformers with a hidden size of 784. We adopt the Adam optimizer (Kingma and Ba 2014) with a learning rate of 3e-5, a batch size of 64, and a dropout probability of 0.1. Besides, we set τ^{max} to be 1 and τ^{min} to be 0.05. We tune the λ_1, λ_2 and λ_3 in the range [0,1].

Experiment Results

Overall Performance (RQ1). Table 2 presents the comparison results of ALSA between AIFI and the baseline models. Evidently, the proposed AIFI model consistently achieves state-of-the-art results on three real-world datasets

Sub-Task	Method	$\mathbb{D}_{\mathrm{R}}$	$\mathbb{D}_{\mathrm{L}}$	$\mathbb{D}_{\mathrm{T}}$
AP	SPAN-BERT	0.8238	0.8335	0.7528
	S-AEAC	0.8420	0.8519	0.7604
	SPRM	0.8665	0.8425	0.6960
	HI-ASA	0.8592	0.8573	**0.7621**
	AIFI (ours)	**0.8701**	**0.8597**	0.7523
SP	SPAN-BERT	0.8995	0.8139	0.7516
	T-GCN$^\sharp$	-	0.8179	0.7803
	SPRM	0.9035	0.8150	0.7834
	HI-ASA	0.9246	0.8375	0.8382
	AIFI (ours)	**0.9284**	**0.8486**	**0.8549**

Table 3: Sub-task results on three real-world datasets.The $\sharp$ means the results are from (Tian, Chen, and Song 2021).

when compared to existing methods. According to the significance test, the model demonstrates excellent performance with p-value $\leq$ 0.05. More precisely, AIFI outperforms the HI-ASA model (the current best method) by $+1.22\%$, $+0.92\%$ and $+1.26\%$ in $F1$ score on $\mathbb{D}_{\mathrm{R}}$, $\mathbb{D}_{\mathrm{L}}$, and $\mathbb{D}_{\mathrm{T}}$, respectively. This advantage primarily stems from the inter-task contrastive learning paradigm, which explicitly captures proper knowledge across tasks, thereby enhancing representation learning for AP-SP. Additionally, the adaptive temperature parameter aids in the contrastive representation integration, enabling better capturing of task-relatedness when compared to other interaction models.

Furthermore, to validate ChatGPT's performance on ALSA tasks, we offer demonstration examples in the input to fully exploit the capabilities of ChatGPT through in-context learning. We perform few-shot prompting experiments on ALSA using various examples. However, the results indicate that ChatGPT still lags significantly behind state-of-the-art models on three datasets. This suggests that enhancing performance in these low-resource domains continues to be a challenging task.

Results for Individual Sub-Task (RQ2). To further investigate the effectiveness of AIFI in sub-tasks, we conducted a comparative analysis between AIFI and previous methodologies developed for both sub-tasks. The results are presented in Table 3. Basically, our model has achieved good results, especially in the SP task, with performance improvements of $+0.38\%$, $+1.11\%$, and $+1.67\%$ on three datasets, respectively, demonstrating the effectiveness of AIFI. However, the performance in AP task is lower than HI-ASA model on $\mathbb{D}_{\mathrm{T}}$. This happened because the sentence length in Twitter reviews is relatively short, which may limit the contextual information contained in the text. It is challenging to estimate the correlation strength between task-specific features due to this limited context.

Ablation Study (RQ3). For detailed model analysis, an ablation study is performed, to study and observe the contribution and feasibility of different parts of the AIFI model. From Table 4, the observation can be made that all the parts of the AIFI model positively contributed to enhancing the ALSA performance. But the greatest contribution comes

Task	Method	$\mathbb{D}_R$	$\mathbb{D}_L$	$\mathbb{D}_T$
AP	-w/o CL	0.8644	0.8068	0.7373
	-w/o Adaptive CL	0.8647	0.8248	0.7436
	AIFI (ours)	**0.8701**	**0.8378**	**0.7523**
SP	-w/o CL	0.9214	0.8202	0.8344
	-w/o Adaptive CL	0.9270	0.8407	0.8288
	AIFI (ours)	**0.9284**	**0.8486**	**0.8549**
ALSC	-w/o CL	0.7931	0.6602	0.5887
	-w/o Adaptive CL	0.7955	0.6847	0.6086
	AIFI (ours)	**0.7991**	**0.6963**	**0.6130**

Table 4: Ablation study results on three datasets. "- w/o CL" means without contrastive learning, and "- w/o adaptive CL" means without adaptive contrastive learning.

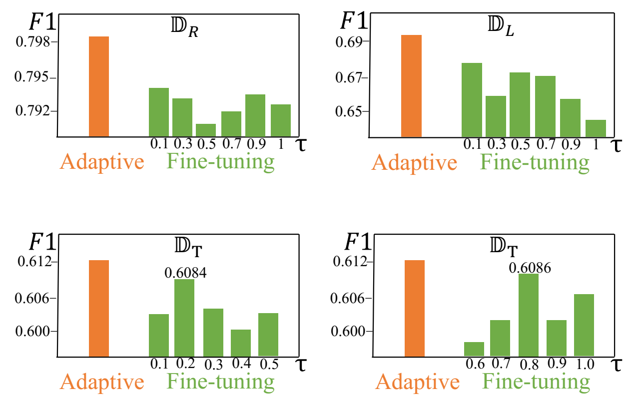

Figure 3: $F1$ results for adaptive τ and fine-tuning τ method.

	Reviews	HI-ASA	AIFI (Ours)
(1)	Not only was the food outstanding, but the little perks were great.	(food, positive) ✓ (perks, positive) ✓	(food, positive) ✓ (perks, positive) ✓
(2)	I asked for seltzer with lime, no ice.	(perseltzer with lime, positive) ×	(perseltzer with lime, neutral) ✓
(3)	I bought a protector for my key pad and it works great.	(foprotector, positive) × (works, positive) ✓	(foprotector, neutral) ✓ (works, positive) ✓
	Error Analysis		
(4)	I The first time the sushi was outstanding, the second time it was a little bland.	(sushi, Positive) ✓ (none) ×	(sushi, Positive) ✓ (none) ×
(5)	I am satisfied with the com-puter, except for the processor.	(computer, positive) ✓ (processor, negative) ×	(computer, positive) ✓ (processor, negative) ×

Figure 4: Some case studies and error analysis. ✓ and × denote correct and incorrect predictions. The aspects are outlined with rectangular boxes.

from the contrast-based alignment strategy: when the contrastive learning module is removed, there is a significant drop in the model's performance. This demonstrates that contrast-based alignment strategy highly influence the performance and encourage learning proper interaction features between two sub-task through contrastive learning. The previous alignment strategies fail to distinguish task-specific features, resulting in suboptimal performance.

Effect of Parameter τ (RQ4). In AIFI, the parameter τ is crucial. We derive it through a correlation prediction network but wondered if manual tuning across datasets could enhance performance. To validate this hypothesis, we experiment with three datasets (as illustrated in Figure 3). Specifically, we adjust τ in the range [0,1] for three datasets.

In the first two datasets, we observe that the average performance of the latter half (i.e., $\tau \geq 0.5$) in the fine-tuning method is generally lower than that of the first half. This pattern arises because as the parameter τ increases, the correlation strength in contrastive learning between aspect and sentiment features decreases, leading to a decrease in performance. For the $\mathbb{D}_T$ dataset, the fine-tuning approach achieves its highest performance with an optimal τ value of 0.8. Importantly, this value exceeds the optimal values of 0.1 observed in the other two datasets. This difference can be attributed to the relatively high proportion of neutral reviews in $\mathbb{D}_T$. Although manual adjustment of τ can yield satisfactory outcomes, a notable disparity persists when compared to our adaptive method. Our belief is that not all sentences within a dataset inherently exhibit consistent task-related features. So an effective approach for joint aspect and sentiment prediction might involve varying contrastive strengths.

Case Study and Error Analysis (RQ5). Figure 4 displays prediction results of HI-ASA and AIFI models on some example reviews that are selected from datasets. In the 1st review, both HI-ASA and AIFI models successfully extract aspects and corresponding sentiments. But further, in the 2nd review, HI-ASA model does not handle the aspect-sentiment correlations very well in neutral review and fails to predict the sentiment. However, our model can accurately extract this pair "(seltzer with lime, neutral)", which indicates that proper interactions between AP and SP truly benefit the task. Similarly, in the 3rd review, HI-ASA model becomes unfit for detecting multiple aspects with different sentiments.

Despite the fact that the AIFI model is capable of settling the negative transfer issues to some extent, it is necessary to conduct error analysis to assess the constraints of the model. In the 4th review, our model proves inadequate in identifying multiple opinion terms linked to a single aspect, thus struggling to differentiate sentiments. In the 5th review, the sentiment expressions of "processor" are overly implicit, leading to the model's inability to make accurate predictions.

Conclusion

In this paper, we explore the effective prediction of aspects and sentiments in reviews through the utilization of inter-task proper interactions. We present a contrast-based alignment framework, which is able to learn sufficient inter-task interactive features while avoiding negative transfer issues. To dynamically adjust the alignment process, we devise a predictive network dedicated to estimating correlation strength between aspect and sentiment features. Through experiments conducted on three datasets, we demonstrate that AIFI surpasses the current state-of-the-art methods.

Acknowledgments

We sincerely thank all the anonymous reviewers for their valuable comments to improve this paper. The research work is supported by the National Key Research and Development Program of China under Grant No. 2021ZD0113602, the National Natural Science Foundation of China under Grant Nos. 62176014, the Fundamental Research Funds for the Central Universities, National Natural Science Foundation of China under Grant No. 62206266, the National Natural Science Foundation of China (No. 62176029).

References

Bai, T.; Xiao, Y.; Wu, B.; Yang, G.; Yu, H.; and Nie, J.-Y. 2022. A Contrastive Sharing Model for Multi-Task Recommendation. In *Proceedings of the ACM Web Conference 2022*, 3239–3247.

Blu, T.; Thévenaz, P.; and Unser, M. 2004. Linear interpolation revitalized. *IEEE Transactions on Image Processing*, 13(5): 710–719.

Chen, S.; Shi, X.; Li, J.; Wu, S.; Fei, H.; Li, F.; and Ji, D. 2022a. Joint Alignment of Multi-Task Feature and Label Spaces for Emotion Cause Pair Extraction. In *Proceedings of the 29th International Conference on Computational Linguistics*, 6955–6965.

Chen, W.; Du, J.; Zhang, Z.; Zhuang, F.; and He, Z. 2022b. A Hierarchical Interactive Network for Joint Span-based Aspect-Sentiment Analysis. In *Proceedings of the 29th International Conference on Computational Linguistics*, 7013–7019.

Chen, W.; Yu, W.; He, G.; Jiang, N.; and He, G. 2020. Coarse-to-Fine Attention Network via Opinion Approximate Representation for Aspect-Level Sentiment Classification. In *Neural Information Processing: 27th International Conference, ICONIP 2020, Bangkok, Thailand, November 23–27, 2020, Proceedings, Part I 27*, 704–715. Springer.

Cui, J.; Wang, Z.; Ho, S.-B.; and Cambria, E. 2023. Survey on sentiment analysis: evolution of research methods and topics. *Artificial Intelligence Review*, 1–42.

Devlin, J.; Chang, M.-W.; Lee, K.; and Toutanova, K. 2019. 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)*, 4171–4186.

Fan, Z.; Wu, Z.; Dai, X.; Huang, S.; and Chen, J. 2019. 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)*, 2509–2518.

Giorgi, J.; Nitski, O.; Wang, B.; and Bader, G. 2021. DeCLUTR: Deep Contrastive Learning for Unsupervised Textual Representations. 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)*, 879–895.

Hu, M.; Peng, Y.; Huang, Z.; Li, D.; and Lv, Y. 2019. Open-Domain Targeted Sentiment Analysis via Span-Based Extraction and Classification. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, 537–546.

Huang, H.; Ye, D.; Shen, L.; and Liu, W. 2023. Curriculum-Based Asymmetric Multi-Task Reinforcement Learning. *IEEE transactions on pattern analysis and machine intelligence*, 45(6): 7258–7269.

Kingma, D. P.; and Ba, J. 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

Lee, G.; Yang, E.; and Hwang, S. 2016. Asymmetric multi-task learning based on task relatedness and loss. In *International conference on machine learning*, 230–238. PMLR.

Li, H.; and Lu, W. 2017. Learning Latent Sentiment Scopes for Entity-Level Sentiment Analysis. In *AAAI*, 3482–3489.

Liang, B.; Luo, W.; Li, X.; Gui, L.; Yang, M.; Yu, X.; and Xu, R. 2021. Enhancing aspect-based sentiment analysis with supervised contrastive learning. In *Proceedings of the 30th ACM international conference on information & knowledge management*, 3242–3247.

Liang, S.; Wei, W.; Mao, X.-L.; Fu, Y.; Fang, R.; and Chen, D. 2023. STAGE: Span Tagging and Greedy Inference Scheme for Aspect Sentiment Triplet Extraction. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, 13174–13182.

Lin, P.; and Yang, M. 2020. A Shared-Private Representation Model with Coarse-to-Fine Extraction for Target Sentiment Analysis. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: Findings*, 4280–4289.

Lin, Z.; Yang, X.; Peng, X.; Zhao, W.; Liu, S.; Wang, L.; and Zheng, B. 2022. Modeling Adaptive Fine-grained Task Relatedness for Joint CTR-CVR Estimation.

Lingling, X.; Haoran, X.; Zongxi, L.; WANG, F. L.; Weiming, W.; and Qing, L. 2023. Contrastive Learning Models for Sentence Representations. *ACM Transactions on Intelligent Systems and Technology*.

Liu, B. 2012. Sentiment Analysis and Opinion Mining. In *Synthesis Lectures on Human Language Technologies 5.1 (2012): 1-167.*

Liu, Y.; Zhou, Y.; Li, Z.; Wang, J.; Zhou, W.; and Hu, S. 2023. HIM: An End-to-end Hierarchical Interaction Model for Aspect Sentiment Triplet Extraction. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*.

Luo, H.; Li, T.; Liu, B.; and Zhang, J. 2019. DOER: Dual Cross-Shared RNN for Aspect Term-Polarity Co-Extraction. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, 591–601.

Lv, H.; Liu, J.; Wang, H.; Wang, Y.; Luo, J.; and Liu, Y. 2023. Efficient hybrid generation framework for aspect-based sentiment analysis. In *Proceedings of the 17th Conference of the European Chapter of the Association for Computational Linguistics*, 1007–1018.

Lv, Y.; Wei, F.; Zheng, Y.; Wang, C.; Wan, C.; and Wang, C. 2021. A span-based model for aspect terms extraction and aspect sentiment classification. *Neural Computing and Applications*, 33(8): 3769–3779.

Ma, J.; Zhao, Z.; Yi, X.; Chen, J.; Hong, L.; and Chi, E. H. 2018. Modeling task relationships in multi-task learning with multi-gate mixture-of-experts. In *Proceedings of the 24th ACM SIGKDD international conference on knowledge discovery & data mining*, 1930–1939.

Mitchell, M.; Aguilar, J.; Wilson, T.; and Van Durme, B. 2013. Open domain targeted sentiment. In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*, 1643–1654.

Oord, A. v. d.; Li, Y.; and Vinyals, O. 2018. Representation learning with contrastive predictive coding. *arXiv preprint arXiv:1807.03748*.

Pontiki, M.; Galanis, D.; Papageorgiou, H.; Androutsopoulos, I.; Manandhar, S.; Mohammad, A.-S.; Al-Ayyoub, M.; Zhao, Y.; Qin, B.; De Clercq, O.; et al. 2016. Semeval-2016 task 5: Aspect based sentiment analysis. In *Proceedings of the 10th international workshop on semantic evaluation (SemEval-2016)*, 19–30.

Pontiki, M.; Galanis, D.; Papageorgiou, H.; Manandhar, S.; and Androutsopoulos, I. 2015. Semeval-2015 task 12: Aspect based sentiment analysis. In *Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015)*, 486–495.

Pontiki, M.; Galanis, D.; Pavlopoulos, J.; Papageorgiou, H.; Androutsopoulos, I.; and Manandhar, S. 2014. SemEval-2014 Task 4: Aspect Based Sentiment Analysis. In *Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014)*, 27–35.

Rethmeier, N.; and Augenstein, I. 2023. A Primer on Contrastive Pretraining in Language Processing: Methods, Lessons Learned, and Perspectives. *ACM Computing Surveys*, 55(10): 1–17.

Tian, Y.; Chen, G.; and Song, Y. 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*, 2910–2922.

Wang, F.; and Liu, H. 2021. Understanding the behaviour of contrastive loss. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2495–2504.

Wang, T.; and Isola, P. 2020. Understanding contrastive representation learning through alignment and uniformity on the hypersphere. In *International Conference on Machine Learning*, 9929–9939. PMLR.

Wang, Z.; Xie, Q.; Ding, Z.; Feng, Y.; and Xia, R. 2023. Is ChatGPT a good sentiment analyzer? A preliminary study. *arXiv preprint arXiv:2304.04339*.

Xiong, H.; Yan, Z.; Zhao, H.; Huang, Z.; and Xue, Y. 2022. Triplet Contrastive Learning for Aspect Level Sentiment Classification. *Mathematics*, 10(21): 4099.

Yang, Y.; Li, K.; Quan, X.; Shen, W.; and Su, Q. 2020. Constituency lattice encoding for aspect term extraction. In *Proceedings of the 28th international conference on computational linguistics*, 844–855.

Zhang, O.; Wu, M.; Bayrooti, J.; and Goodman, N. 2021. Temperature as uncertainty in contrastive learning. *arXiv preprint arXiv:2110.04403*.

Zhang, R.; Ji, Y.; Zhang, Y.; and Passonneau, R. J. 2022a. Contrastive data and learning for natural language processing. In *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies: Tutorial Abstracts*, 39–47.

Zhang, W.; Li, X.; Deng, Y.; Bing, L.; and Lam, W. 2022b. A Survey on Aspect-Based Sentiment Analysis: Tasks, Methods, and Challenges. *arXiv preprint arXiv:2203.01054*.

Zhang, Y.; Yang, Y.; Li, Y.; Liang, B.; Chen, S.; Dang, Y.; Yang, M.; and Xu, R. 2022c. Boundary-driven table-filling for aspect sentiment triplet extraction. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, 6485–6498.

Zhang, Y.; Zhang, R.; Mensah, S.; Liu, X.; and Mao, Y. 2022d. Unsupervised sentence representation via contrastive learning with mixing negatives. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, 11730–11738.

Zhou, C.; Li, Q.; Li, C.; Yu, J.; Liu, Y.; Wang, G.; Zhang, K.; Ji, C.; Yan, Q.; He, L.; et al. 2023. A comprehensive survey on pretrained foundation models: A history from bert to chatgpt. *arXiv preprint arXiv:2302.09419*.