An Instruction Tuning-Based Contrastive Learning Framework for Aspect Sentiment Quad Prediction with Implicit Aspects and Opinions

Hao Zhang[1,2], Yu-N Cheah[1*], Congqing He[1], Feifan Yi[1],

[1]School of Computer Sciences, Universiti Sains Malaysia, Penang, Malaysia,
[2]Cangzhou Normal University, Cangzhou, China
zhanghaousm@gmail.com, yncheah@usm.my
{hecongqing,yifeifan}@student.usm.my

Abstract

Aspect sentiment quad prediction (ASQP) is crucial in aspect-based sentiment analysis (ABSA). It involves identifying a text's aspect, sentiment, opinion, and category. Existing methods have insufficiently explored how to effectively leverage the knowledge of pre-trained language models (PLMs) to handle implicit aspects and opinions, particularly in combinations such as implicit aspect & explicit opinion, explicit aspect & implicit opinion, and implicit aspect & implicit opinion. We introduce ITSCL, a framework leveraging Instruction Tuning and Supervised Contrastive Learning to improve aspect sentiment quad predictions, especially for implicit aspects and opinions. Implementing this approach presents several challenges. First, designing effective instructions and prompts to optimize the model's training is difficult. Second, creating sentiment combination vectors with contrastive learning to enhance the model's discrimination requires further investigation. To address these challenges, ITSCL combines instruction tuning with aligned PLM templates, enabling better knowledge acquisition and identification of implicit sentiments. Additionally, the contrastive learning framework enhances performance by using four fully connected layers to combine sentiments, aspects, opinions, and combinations, maximizing similarity for same-label representations and minimizing it for different labels. Experimental results show our method significantly outperforms previous methods on benchmark datasets.[1]

1 Introduction

ASQP aims to extract four key components from a given text: aspect, sentiment, opinion, and category (as shown in Figure 1). This task is crucial in understanding user opinions and sentiments in various applications such as customer feedback analysis, social media monitoring, and product reviews.

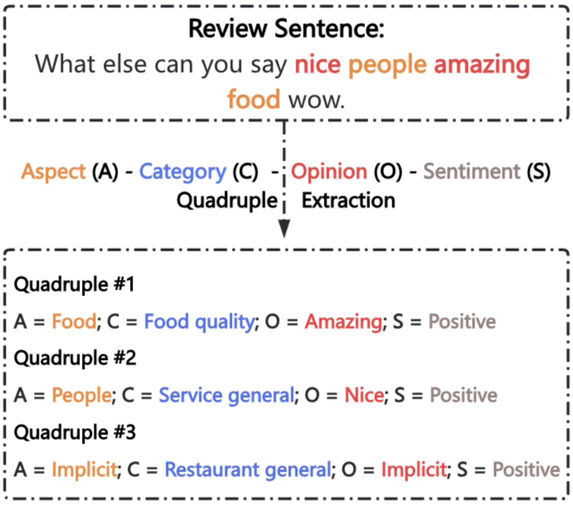

Figure 1: An illustration of the Aspect-Category-Opinion-Sentiment quadruple extraction (ACOSQE), also denoted as the Aspect Sentiment Quad Prediction (ASQP) task.

Surveys (Zhou et al., 2019; Zhang et al., 2022; Zhu et al., 2022) have extensively studied the evolution and trends in ABSA methodologies, particularly those employing deep learning techniques, attention mechanisms, and pre-trained language models (PLMs). Their work highlights that adopting deep learning methods has significantly advanced research in ASQP. Nonetheless, a comprehensive overview of ASQP and its solutions is lacking. Zhang et al. (2024a) provides the first comprehensive review of ASQP, addressing gaps by reclassifying ABSA subtasks, summarizing various PLM methods applied to ASQP, and exploring ChatGPT in sentiment analysis. The field has advanced significantly and is primarily driven by PLM approaches. Transformer-based

[*]Corresponding author: Yu-N Cheah
[1]Our experimental codes and data are available at: https://github.com/sydmou/ASQP-ITSCL

Findings of the Association for Computational Linguistics: EMNLP 2024, pages 7698–7714

models like BERT (Kenton and Toutanova, 2019), BART (Lewis et al., 2020), and T5 (Raffel et al., 2020) have set new benchmarks by leveraging pre-trained language models (PLMs) and fine-tuning techniques to achieve state-of-the-art results.

However, several challenges remain. BERT-based methods (Cai et al., 2021; Zhang et al., 2021a), while pioneering the use of transformer models, face limitations when combined with traditional techniques like CRF, leading to issues such as gradient vanishing and error propagation. BART-based approaches (Xiong et al., 2023; Hoang et al., 2022) introduced encoder-decoder models but lack extensive fine-tuning research and often borrow methods from T5. Although T5-based approaches have recently achieved promising results in the ASQP field, research on implicit sentiment remains limited. Moreover, most methods only research simple input and output template construction (Zhang et al., 2021a; Gao et al., 2022; Bao et al., 2022; Mao et al., 2022; Hu et al., 2022; Gou et al., 2023; Zhang et al., 2024b; Wang et al., 2024) or basic contrastive learning (Xiong et al., 2023; Peper and Wang, 2022).

In this work, we enhance quad extraction accuracy for implicit aspects and opinions using the T5 model, improving ASQP fine-tuning through deeper sentiment instruction and contrastive learning. Implementing this approach faces several challenges, including designing instruction learning templates, constructing multi-dimensional contrastive learning to infer implicit sentiments from single-sentence datasets without additional context, and fine-tuning the T5 model to avoid overfitting or underfitting, thus enhancing accuracy and robustness.

We propose a novel framework, ITSCL (Instruction Tuning and Supervised Contrastive Learning), designed to improve accuracy and robustness, particularly for implicit aspects and opinions. The Instruction Tuning (IT) component uses explicit instruction learning to provide detailed input and output constructions, guiding PLMs and indicating that aspects and opinions can be implicit. This custom template is designed explicitly for ASQP task. The multi-dimensional Supervised Contrastive Learning (SCL) component employs four fully connected layers to optimize similarity for same-label representations and minimize it for different labels (sentiment, aspect, opinion, aspect&opinion). This approach aligns representations and integrates similarities and differences across dimensions, effectively capturing explicit and implicit sentiment. Experiments show our approach significantly outperforms state-of-the-art methods on two benchmark datasets, excelling in explicit sentiment identification and implicit aspect and opinion detection. The key contributions of this paper are:

- We introduce ITSCL, a unified ASQP framework that employs instruction tuning and contrastive learning to improve aspect sentiment quad predictions, especially for implicit aspects and opinions, addressing gaps in current methods.
- We extensively explored the T5-large model in the context of ASQP, marking the first comprehensive application of this model in the ABSA field.
- The experimental results demonstrate that our proposed framework based on T5-large substantially surpasses recent state-of-the-art (SOTA) methods in implicit and explicit sentiment analysis.

2 Methodology

2.1 ASQP Problem Statement

We formulate ASQP as a joint quadruple extraction task following Zhang et al. (2021a) and Cai et al. (2021). The aim is to extract an unordered set of ACOS quadruples $Q_1, Q_2, \ldots, Q_n$ from text T. Each quadruple $Q_i = (a_i, c_i, o_i, s_i)$ consists of an aspect term $\boldsymbol{a} \in \boldsymbol{V}_{\mathrm{x}} \cup$ *{Implicit}*, aspect category $\boldsymbol{c} \in \boldsymbol{V}_{\mathrm{x}}$ (Pre-Defined) of the aspect term, opinion term $\boldsymbol{o} \in \boldsymbol{V}_{\mathrm{x}} \cup$ *{Implicit}*, and sentiment polarity $\boldsymbol{p}$ $\in$ {POS, NEU, NEG}. In some cases, quadruples may lack clear supporting aspect and/or opinion spans, and these instances are marked as *implicit*.

2.2 ASQP as ITSCL Framwork

Figure 2 illustrates the implementation process of the ITSCL method in the ASQP task. The left part includes the two key steps: input prompting and output prompting. Input prompting consists of manual prompting and instruction tuning to fine-tune the T5 model. The input examples are fed into the encoder stack, generating the output from the decoder stack. The right part highlights the implementation process of supervised contrastive learning. Each layer maximizes the similarity between representations with the same label and minimizes the similarity with different labels. This

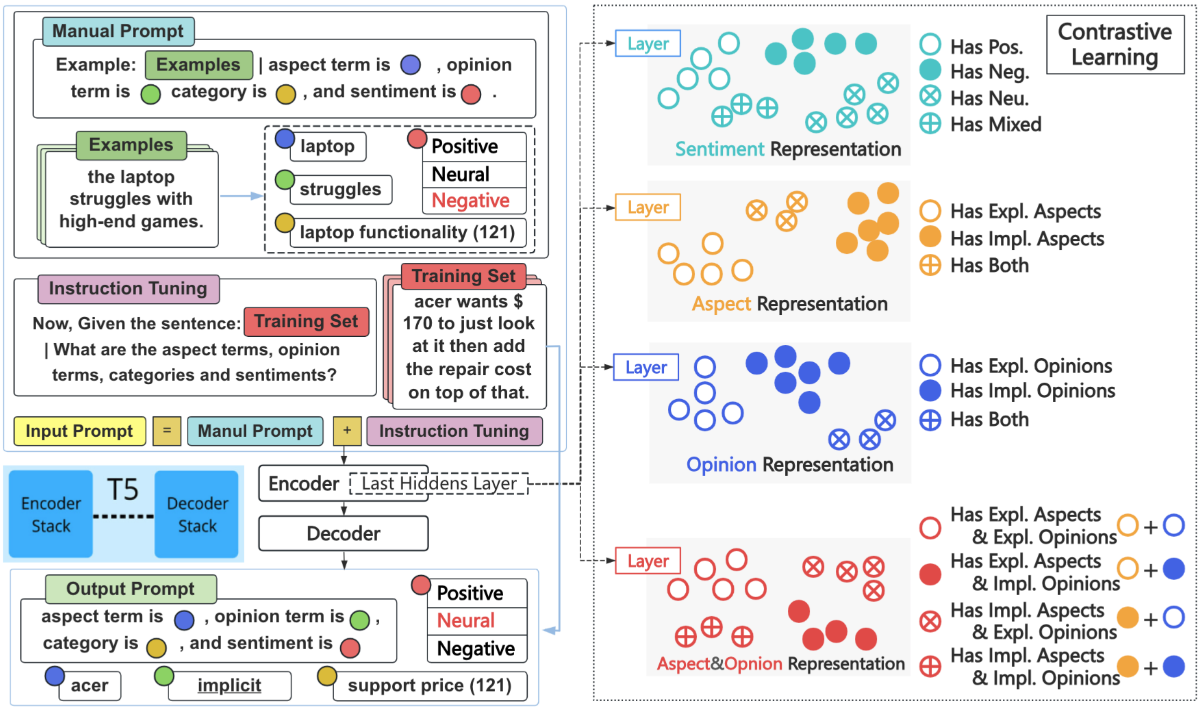

Figure 2: Framework of the ASQP-ITSCL Approach. The left part illustrates instruction tuning and the flow through the T5 encoder-decoder model. The restaurant category has 13 classifications, and the laptop category has 121 types. The right part shows the contrastive learning process, with layers representing sentiments, aspects, opinions, and combinations. These combinations include four types: explicit aspects with explicit opinions, explicit aspects with implicit opinions, implicit aspects with explicit opinions, and implicit aspects with implicit opinions.

Instance

Input: Example: the laptop struggles with high-end games. | aspect term is laptop, opinion term is struggles, category is laptop functionality, and sentiment is negative. Now, Given the sentence: acer wants $ 170 to just look at it then add the repair cost on top of that. | What are the aspect terms, opinion terms, categories and sentiments?
Output: aspect term is acer, opinion term is implicit, category is support price, and sentiment is neutral

Figure 3: Example of the ASQP-IT Method Input and Output Construction.

approach effectively differentiates between positive, negative, and neutral sentiments and explicit and implicit aspects and opinions.

2.3 Unified Instruction Prompt Tuning

During the manual prompting stage, example sentences are provided, annotating aspect terms, opinion terms, categories, and sentiment polarities to help the model understand the task. In the instruction tuning stage, the model is required to analyze the text and answer relevant questions. Combining the content of both stages generates formatted input text, which the T5 model then processes. The model creates complete sentences containing aspect terms, opinion terms, categories, and sentiment polarities in the output prompting stage. If aspects and opinions are implicit, they are marked as implicit. Figure 3 shows an example of the IT method in the ASQP task, including the input and output parts, demonstrating how input examples and instruction tuning are used to generate complete output sentences, showcasing the practical application and effectiveness of the ITSCL method.

2.3.1 Manual Prompt Design

The input template is meticulously structured to provide the model with a clear example of the required analysis. Examples can refer to one or more prefix samples and their corresponding labels for aspect, opinion, sentiment, and category. In this study, we designed only one sample. A manual input prompt example is provided, accompanied by specific instructions to guide the model in comprehending the analysis needed. [Example Sentence] is a placeholder for example sentence T_e with labeled $Q_e = (a_e, c_e, o_e, s_e)$ provided for the model to analyze. `[A]` is the placeholder for the aspect term $a_e \in O_e$, `[O]` is the placeholder for the opinion term $o_e \in O_e$, and `[C]` is the placeholder for the category $c_e \in O_e$. [S] is the placeholder for the sentiment polarity $s_e \in O_e$.

Example: [Example Sentence] | aspect is [A], opinion is [O], the category is [C], and sentiment is [S].

For the experiment, the T_e: "this place has got to be the best Japanese restaurant in the New York area." is used as the input for the Restaurant dataset, as it relates to the restaurant category. Similarly, T_e: "the laptop struggles with high-end games." is used for the laptop dataset regarding laptops. The example sentences for each domain are randomly generated by ChatGPT and manually annotated.

2.3.2 Instruction Tuning Template

The Instruction Tuning Template is a meticulously crafted guide that facilitates the model's understanding and analysis of sentiment in text. It is structured as follows:

Now, analyze the following sentence: [Target Sentence], What are the aspect terms, opinion terms, categories, and sentiments?

This template prompts the model to identify and extract key components of sentiment expression within a given text T_t from the training data. [Target Text] is the placeholder for T_t. By framing the task as a series of questions, this template encourages the model to consider aspect terms, opinion terms, categories, and sentiments, which correspond to the [Example Sentence], enhancing the accuracy and depth of sentiment analysis.

2.3.3 Output Prompt Design

Given T_t with labelled $Q_t = (a_t, c_t, o_t, s_t)$, the output template is designed to map [A], [C], [O], and [S]. This ensures each Q_t is clearly identified and corresponds to the T_t structure from the input Instruction Tuning Template. During training, if a_t or o_t in T_t are implicit, the placeholders [A] and [O] are assigned "*implicit*." This approach enables the model to handle subtle or hidden sentiments effectively, enhancing the accuracy of the analysis. The template is formulated as follows:

Aspect is [A], Opinion is [O], Category is [C] and Sentiment is [S]

This design streamlines the extraction process and ensures that the output is directly aligned with the input template, clearly and concisely representing the model's sentiment analysis results.

2.3.4 Training Loss

This experiment adopted a new instruction-tuning-based prompt engineering approach leveraging the pre-trained T5 model. The model was fine-tuned by designing specific input prompts and optimizing it using cross-entropy loss, aiming to improve the accuracy and efficiency of implicit sentiment analysis. The loss function is defined as follows:

$$L(\theta) = -\sum_{(x,y)\in D} \log p_\theta(y|x) \tag{1}$$

Here: $L(\theta)$ represents the loss function for model parameters θ. D is the training dataset containing pairs of text and labels. x is the input text. y is the corresponding label, including the aspect, opinion, sentiment, and category. $p_\theta(y|x)$ is the probability of predicting label y given input x when the model is parameterized by θ.

2.4 Contrastive Learning Representation

This research uses the general SCL formulation based on the approaches by Sedghamiz et al. (2021) and Peper and Wang (2022). The model generates representations for each example x_i by feeding the sum-pooled encoder representation through four fully connected layers, each corresponding to a characteristic (Sentiment, Aspect, Opinion, Aspect & Opinion), producing a representation h_i^c. The model architecture varies based on the size of the T5 model used. For the T5-large model, fully connected layers with 1024-dimensional input and output are utilized, while for the T5-base model, the layers are configured with 768-dimensional input and 28-dimensional output.

Formulation and Implementation In SCL, the objective is to maximize the representation similarity between samples with the same label and minimize the representation similarity between samples with different labels. Specifically, the loss function calculates the similarity between each sample i and each of its positive pairs p in the set $P(i)$, and takes the negative log of these similarities; in the denominator, it calculates the similarity between sample i and all negative pairs b in the set $B(i)$, where $b \neq i$ indicates that the negative pairs do not include sample i itself. The formula is as follows:

$$\mathcal{L} = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{|P(i)|}\sum_{p\in P(i)} -\log\frac{\exp(\mathrm{sim}(h_i, h_p)/\tau)}{\sum_{b\neq i}\exp(\mathrm{sim}(h_i, h_b)/\tau)} \tag{2}$$

Where: N is the total number of samples in the batch. In this experiment, the training batch size N

	Restaurant	Laptop
#Categories	13	121
#Sentences (S)	2284	4076
#Quads (Q)	3661	5773
#Q/S	1.60	1.42
#EA & EO	2431 (66.40%)	3278 (56.78%)
#IA & EO	530 (14.48%)	912 (15.80%)
#EA & IO	350 (9.56%)	1241 (21.50%)
#EA & IO	350 (9.56%)	342 (5.92%)
#POS	2503	3578
#NEU	151	316
#NEG	1007	1879
#Train	1530	2934
#Dev	171	326
#Test	583	816
#Train (Quads)	2484	4172
#Dev (Quads)	261	440
#Test (Quads)	916	1161

Table 1: Data statistics for the ACOS-Datset. Both datasets feature explicit and implicit aspects and opinions, offering diverse quadruple types (EAEO, EAIO, IAEO, IAIO) and balanced sentiment distributions (#NEG, #NEU, #POS).

is set to 16. $P(i)$ is the set of positive pairs for sample i (samples with the same label). $\text{sim}(h_i, h_p)$ denotes the similarity between sample i and its positive pair p. τ is a temperature scaling parameter that controls the smoothness of the similarity scores. During training, a dropout probability of 0.1 is used to prevent overfitting. Excluding sample i itself from negative pairs is crucial, as self-similarity is always highest and uninformative, which would hinder the learning process. This approach ensures the model focuses on distinguishing features between different samples rather than self-similarity.

Final Training Loss The final loss is defined as follows:

$$\mathcal{L} = \mathcal{L}_{\text{IT}} + \alpha_1 \mathcal{L}_{\text{sent}} + \alpha_2 \mathcal{L}_{\text{aspect}} + \alpha_3 \mathcal{L}_{\text{opinion}} + \alpha_4 \mathcal{L}_{\text{joint}} \tag{3}$$

where α_1, α_2, α_3, and α_4 are hyperparameters.

3 Experimental Setup

3.1 Datasets

ACOS Dataset We incorporate the Restaurant and Laptop datasets from ACOS (Cai et al., 2021), detailed in Table 1. They are divided into training, validation, and testing sets for systematic model training and evaluation.

3.2 Compared Models

We compare our methods with the following two types of previous state-of-the-art methods:

Pipeline model The Double Propagation (DP) method (Qiu et al., 2011), utilized by Cai et al. (2021) for the ASQP task, enhances textual coverage by leveraging relationships between extracted aspects and opinions. JET (Xu et al., 2020) is an end-to-end method identifying aspects, opinions, and sentiment polarities using position-aware tagging. Cai et al. (2021) adapted JET for ASQP by extracting aspect-opinion-sentiment triples and then using a BERT-based model for aspect categories. TAS-BERT (Wan et al., 2020) jointly detects sentiment tuples, while Extract-Classify (Cai et al., 2021) decomposes the ACOS task into two steps. TAS-BERT-ACOS (Pipeline) (Cai et al., 2021) co-extracts category-sentiment conditional aspect-opinion pairs and filters out invalid pairs to form quadruples.

Unified model PARAPHRASE-BART (Xiong et al., 2023) uses BART for ABSA, handling aspect term extraction and sentiment polarity classification. GEN-NAT-SCL-BART (Xiong et al., 2023) enhances BART with natural adversarial training and SCL. BART-CRN (Xiong et al., 2023) combines BART with a convolutional recurrent network for improved aspect and sentiment extraction. BARTABSA (Hoang et al., 2022) handles ABSA sub-tasks separately using BART. GAS (Zhang et al., 2021b) frames ABSA tasks as a generative process. Paraphrase (Zhang et al., 2021a) designs semantic templates with fixed-order tuple elements. Seq2Path (Mao et al., 2022) and Opinion Tree (Bao et al., 2022) generate tuples as tree paths, comprehensively detecting and visualizing sentiment elements. GEN-SCL-NAT (Peper and Wang, 2022) integrates SCL and natural adversarial training with T5 for improved robustness. UnifiedABSA (Wang et al., 2024) uses multi-task instruction learning for a unified framework. Special_Symbol (Hu et al., 2022) enhances representation with special symbols. DLO (Hu et al., 2022) optimizes structure selection based on training set scores. ILO (Hu et al., 2022) selects template orders based on instance context and semantics. Special_Symbols+UAUL, DLO+UAUL, and ILO+UAUL (Hu et al., 2023) combine unsupervised adversarial uncertainty learning (UAUL) to improve robustness. MvP (Gou et al., 2023) improves sentiment tuple prediction by aggregating multi-view results.

3.3 Experiment Details

We employ the T5-base and T5-large models, both from Raffel et al. (2020), available in the Huggingface Transformer library, as our pre-trained generative encoder-decoder models. During training, we set the learning rate to 3e-4 and 9e-5 for T5 and the dropout rate to 0.1 for all contrastive learning (CL) layers. The T5-base model is trained on Nvidia 3090 GPUs, while the T5-large model is trained on Nvidia A40 GPUs. The hyperparameters in Equation 3 are set as follows: $\alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = \alpha$, with $\alpha = 0.05$, and the SCL temperature τ is set to 0.25. These parameters are tuned on the training set. All reported results are the average of five runs with different random seeds. The related values are reported in Table 9.

3.4 Evaluation Metrics

The experiment uses F1 scores (F_1) as the main evaluation metric. A sentiment quad prediction is correct if all predicted elements match the gold labels. Precision (P) and recall (R) scores for the ASQP task are also reported.

4 Results and Discussions

4.1 Main Performance Results

Table 2 reports the overall performance on the ASQP task. Comparing model performance on the Restaurant and Laptop datasets shows that the ASQP-ITSCL framework, particularly with T5-large, performs the best. However, the T5-base variant does not achieve the top results, highlighting the importance of model size in performance. The ASQP-ITSCL (T5-large) model achieved the highest F1 scores on these two datasets, 64.86 and 46.11, respectively. This represents an improvement of 2.03 and 0.67 points over Opinion Tree and 2.24 and 0.95 points over GEN-SCL-NAT. Other models showed that rule-based methods performed the worst; BERT-based methods showed improvements but were still outperformed by BART-based methods. Overall, T5-based models performed the best.

4.2 Explicit and Implicit Sentiment Analysis

As shown in Table 3, ASQP-ITSCL(T5-large) achieved the highest F1 scores on Restaurant (71.8) and Laptop (47.2), significantly outperforming other models in EAEO. For implicit aspects and opinions (IAEO, EAIO, IAIO), on Restaurant, IAEO and EAIO scores were 53.2 and 44.4, slightly lower than GEN-SCL-NAT's 56.5 and 46.2, but IAIO scored 52.2, surpassing GEN-SCL-NAT's 50.7. On Laptop, IAEO scored 61.3, higher than GEN-SCL-NAT's 54.0 and PARAPHRASE's 51.0. While EAIO was close to GEN-SCL-NAT (34.4 vs 34.3), IAIO (39.7) exceeded GEN-SCL-NAT and PARAPHRASE's 39.6. The relatively lower performance of ITSCL on certain metrics, such as IAEO and EAIO in the Restaurant dataset, may be attributed to the uneven distribution and sample sizes between the Restaurant and Laptop datasets. As shown in Table 1, the Laptop dataset contains approximately 382 more IAEO samples and over 891 additional EAIO samples compared to the Restaurant dataset. This discrepancy could be inferred as a reason why our method achieves better results on the Laptop dataset. These results show that ASQP-ITSCL T5-large excels in both explicit and implicit aspects and opinions.

5 Additional Analyses

5.1 Ablation Experiments

As shown in Table 4, ablation studies indicate that removing the sentiment representation significantly reduces performance on both datasets, highlighting its critical role. Removing aspect representation has a smaller impact on the laptop dataset but a larger impact on the restaurant dataset. Removing opinion representation has a relatively smaller yet significant impact on both datasets. When both aspect and opinion representations are removed, F1 scores drop on both datasets, especially the restaurant dataset, to 62.98, showing their combined importance. Removing all representations, leaving only the instruction tuning information (IT), causes a significant performance drop on both datasets, particularly the restaurant dataset. This underscores the importance of all representation components and the IT+SCL method in enhancing performance on sentiment analysis tasks.

5.2 SNE Representations

To understand the impact of the SCL objective on the model's hidden representations, t-SNE visualizations of the mean-pooled final encoder layer were generated (Van der Maaten and Hinton, 2008). These t-SNE plots (Figure 4) show the model's ability to distinguish between different kinds of representations around aspects, opinions, sentiment polarities and aspect&opinion at different training epochs for the restaurant dataset. As training progresses, the model gradually learns to differentiate these features. Data points are chaotic and indistinguishable in the early stages (E=1). By the later stages (E=50), data points are more separated,

Method	Model	Restaurant			Laptop		
		P.	R.	F1.	P.	R.	F1.
Double-Propagation (Cai et al., 2021)	RULE	34.67	15.08	21.04	13.0	5.70	8.0
JET-ACOS (Cai et al., 2021)	BERT	59.81	28.94	39.01	44.52	16.25	23.81
TAS-BERT-ACOS (Cai et al., 2021)	BERT	26.29	46.29	33.53	47.15	19.22	27.31
Extract-Classify (Cai et al., 2021)	BERT	38.54	52.96	44.61	45.56	29.48	35.80
PARAPHRASE-BART (Xiong et al., 2023)	BART	43.62	36.19	39.56	36.36	29.63	32.65
GEN-NAT-SCL-BART (Xiong et al., 2023)	BART	48.93	40.51	44.32	37.13	32.44	34.63
BART-CRN (Xiong et al., 2023)	BART	50.84	47.10	48.90	48.16	31.83	38.32
BARTABSA (Hoang et al., 2022)	BART	56.80	51.09	53.45	41.06	37.89	39.41
GAS (Zhang et al., 2021b)	T5-base	57.09	57.51	57.30	43.45	43.29	43.37
Seq2Path(Mao et al., 2022)	T5-base	-	-	58.41	-	-	42.97
ILO + UAUL (Hu et al., 2023)	T5-base	59.46	59.12	59.29	43.92	43.46	43.69
Special_Symbols+UAUL (Hu et al., 2023)	T5-base	61.22	59.87	60.53	44.38	43.65	44.01
Muti-Task-IT(Wang et al., 2024)	T5-base	-	-	60.60	-	-	42.58
DLO + UAUL (Hu et al., 2023)	T5-base	61.03	60.55	60.78	43.78	43.53	43.65
PARAPHRASE (Zhang et al., 2021a)	T5-base	-	-	60.97	-	-	44.08
MvP (Gou et al., 2023)	T5-base	-	-	61.54	-	-	43.92
GEN-SCL-NAT (Peper and Wang, 2022)	T5-large	-	-	62.62	-	-	45.16
Opinion Tree (Bao et al., 2022)	T5-base	63.96	61.74	62.83	46.11	44.79	45.44
ASQP-ITSCL	T5-base	61.45	60.92	61.18	44.69	44.19	44.43
ASQP-ITSCL	T5-large	**65.56**	**64.19**	**64.86**	**46.31**	**45.91**	**46.11**

Table 2: Comparison of methods on Restaurant and Laptop datasets.

Method	Restaurant (F1.)				Laptop (F1.)			
	EAEO	IAEO	EAIO	IAIO	EAEO	IAEO	EAIO	IAIO
Double-Propagation (Cai et al., 2021)	26.0	N/A	N/A	N/A	9.8	N/A	N/A	N/A
JET-ACOS (Cai et al., 2021)	52.3	N/A	N/A	N/A	35.7	N/A	N/A	N/A
TAS-BERT-ACOS (Cai et al., 2021)	33.6	31.8	14.0	39.8	26.1	41.5	10.9	21.2
Extract-Classify (Cai et al., 2021)	45.0	34.7	23.9	33.7	35.4	39.0	16.8	18.6
PARAPHRASE-BART (Xiong et al., 2023)	38.6	37.8	16.7	38.5	31.3	38.9	21.1	35.6
GEN-NAT-SCL-BART (Xiong et al., 2023)	46.9	30.5	20.5	37.6	35.9	40.7	20.9	30.2
BART-CRN (Xiong et al., 2023)	54.1	50.6	18.9	42.9	38.9	54.3	24.5	**40.7**
BARTABSA(split) (Hoang et al., 2022)	58.5	43.9	20.0	42.9	39.9	52.8	23.4	29.8
PARAPHRASE (Zhang et al., 2021a)	65.4	53.3	45.6	45.6	45.7	51.0	33.0	39.6
GEN-SCL-NAT (Peper and Wang, 2022)	66.5	**56.5**	**46.2**	50.7	45.8	54.0	34.3	39.6
ASQP-ITSCL (T5-base)	69.8	51.2	31.9	45.6	46.4	59.1	30.3	40.0
ASQP-ITSCL (T5-large)	**71.8**	53.2	44.4	**52.2**	**47.2**	**61.3**	**34.4**	39.7

Table 3: Comparison of explicit and implicit analysis methods on Restaurant and Laptop datasets. EA, EO, IA, and IO denote explicit aspects, explicit opinions, implicit aspects, and implicit opinions, respectively. N/A indicates the model cannot handle the corresponding type.

Method	Restaurant	Laptop
	F1.	F1.
BARTABSA	53.45	39.41
PARAPHRASE	60.97	44.08
GEN-SCL-NAT	62.62	45.16
ASQP-ITSCL (T5-large)	**64.86**	**46.11**
-w/o Sentiment Rep.	63.15	44.90
-w/o Aspect Rep.	62.47	45.90
-w/o Opinion Rep.	64.09	44.87
-w/o Aspect&Opinion Rep.	62.98	45.49
-w/o All Rep. (IT)	63.09	44.84

Table 4: Ablation analysis of ITSCL (T5-large) model.

demonstrating significant improvement in identifying and distinguishing implicit and explicit sentiment features, with the best results at E=50. These results and visualizations validate the model's effectiveness in implicit and explicit sentiment analysis tasks. Increased training epochs lead to clearer and more separated cluster structures, highlighting the importance of contrastive learning and fine-tuning.

6 Related Works

Aspect-Base Sentiment Analysis (ABSA) Eearly ABSA research mainly focused on single tasks such as Aspect-Based Sentiment Classification (ABSC) (Wang et al., 2016; Liu and Zhang, 2017; Ma et al., 2019; Tay et al., 2018). As the research focus shifted from Aspect Opinion Co-Extraction (AOCE) (Yin et al., 2016), Aspect-Oriented Opinion Extraction (AOOE) (Fan et al., 2019) to Aspect-Opinion Pair Extraction (AOPE) (Zhao et al., 2020; Wu et al., 2021b), researchers began to explore joint extraction and sentiment prediction (Aspect Sentiment Pair Extraction (ASPE) (Cai et al., 2020; Liu et al., 2021) and Category

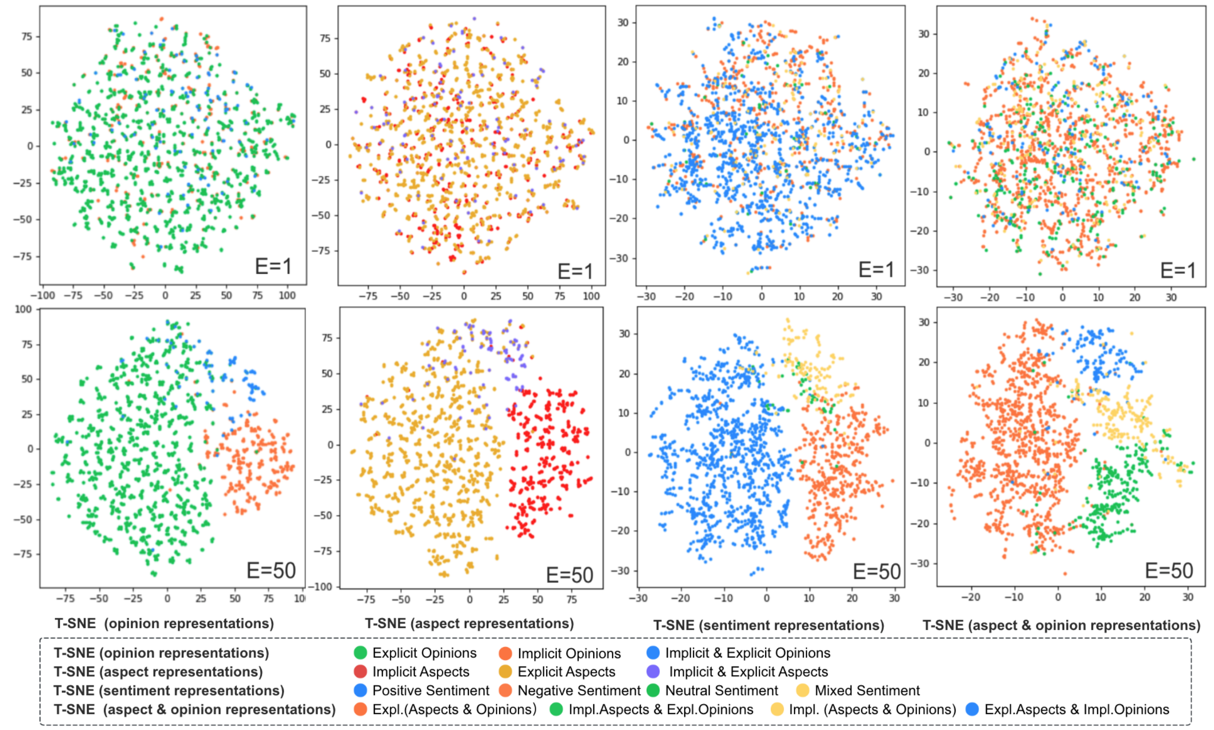

Figure 4: T-SNE visualization of the mean-pooled final encoder layer on the Restaurant dataset. Our ITSCL objective encourages the encoder to produce distinguishable representations of four key input combinations: opinions, aspects, sentiment, and aspects & opinions.

Sentiment Pair Extraction (CSPE) (Wan et al., 2020; Bu et al., 2021; Cai et al., 2020), collectively known as pair ABSA.Recently, the focus has shifted to compound ABSA tasks, including Aspect-Category-Sentiment Triplet Extraction (ACSTE) (Wan et al., 2020; Wu et al., 2021a; Zhang et al., 2021b), Aspect-Opinion-Sentiment Triplet Extraction (AOSTE) (Peng et al., 2020; Xu et al., 2020; Mao et al., 2021; Chen et al., 2021), and Aspect-Category-Opinion-Sentiment Quadruple Extraction (ACOSQE) (Cai et al., 2021).

Instruction Tuning Instruction-based Prompt learning teaches models to follow language instructions and uses text prompts to align pre-training objectives with downstream tasks, improving zero/one/few-shot performance. By designing natural language templates to wrap original inputs and prompt PLMs, researchers have achieved great success in various NLP tasks (Schick and Schütze, 2021; Seoh et al., 2021; Mi et al., 2022). Refer to (Liu et al., 2023; Zhang et al., 2023) for a comprehensive overview of prompt learning. In the ABSA field, Varia et al. (2022) proposed an instruction-tuning framework that converts ABSA subtasks into a question-and-answer format, significantly improving the T5 model's few-shot learning and fine-tuning performance. Wang et al. (2024) referred to their method as multi-task instruction tuning or multi-task prompt training.

Multi-dimensional Contrastive Learning Contrastive Learning is to learn representations by contrasting positive and negative examples, maximizing the similarity of positive pairs and minimizing the similarity of negative pairs. Inspired by its success, many contrastive learning-based models (Li et al., 2021; Liang et al., 2021; Peper and Wang, 2022; Wang et al., 2022; Lin et al., 2023) have been proposed to enhance ABSA performance. Most works (Li et al., 2021; Liang et al., 2021; Peper and Wang, 2022) employ supervised contrastive learning to learn fine-grained sentiment knowledge by aligning sentiment representations with the same sentiment label. Wang et al. (2022) use cross-channel data augmentation strategies and in-domain generators to construct multi-aspect samples for contrastive learning. Lin et al. (2023) use token-level and sentence-level data augmentation strategies and sentiment labels for cross-lingual contrastive learning to enhance ABSA performance. Xu and Wang (2023) presents and compares two commonly used contrastive learning methods to improve ABSA performance.

7 Conclusions

In this paper, we introduced ITSCL, a unified framework designed to enhance the prediction accuracy of aspect sentiment quads, particularly focusing on implicit aspects and opinions. By combining instruction tuning through prompt engineering with highly aligned PLM templates, ITSCL enables models to acquire knowledge more effectively and identify implicit sentiments. The contrastive learning framework also improves model performance by optimizing the similarity between representations with the same label and differentiating those with different labels. Our extensive experiments on benchmark datasets demonstrate that ITSCL significantly outperforms existing methods.

Limitations

The study has the following limitations:

- **Data Diversity:** Most datasets for ASQP are small and limited, failing to capture its complexity and diversity. More comprehensive datasets with detailed annotations of explicit and implicit aspects and opinions are needed.

- **Model Scale:** The effectiveness of the ITSCL framework has only been validated on T5-base and T5-large, both generative models. In future work, we plan to apply it to larger models with updated paradigms and further incorporate additional machine-learning approaches and networks to verify their effectiveness.

- **Explicit and Implicit Sentiment Analysis:** Many sentences contain multiple combinations of explicit and implicit aspects and opinions, which require further study. In particular, the ambiguity, implicitness, complexity of aspects and opinions and the relationships and consistency among quadruples also need further exploration.

- **Parameter Tuning:** The performance heavily relies on fine-tuning hyperparameters, which can be time-consuming and require substantial experimentation.

- **Sample Size Limitation:** Due to the input length requirements and computational limitations of the T5 model, we designed only one standard sample in the manual prompt design. In the future, we can consider adding different types of samples, including explicit and implicit examples, for different types of training data to improve the model's adaptability and effectiveness.

Ethics Statement

We used datasets extensively employed in prior scientific studies in all our experiments. When analyzing the experimental results, we made every effort to uphold fairness and honesty, ensuring that our work does not cause harm to anyone.

Our research adheres to the highest ethical standards in artificial intelligence and natural language processing. All datasets are publicly available, ethically sourced, and anonymized to protect personal information. Our methods and models are designed to be transparent, reproducible, and beneficial to the community, avoiding misuse or harm. We are committed to advancing ASQP by continuously optimizing algorithms and models to improve their accuracy and reliability in practical applications. We pledge to share our research findings and methods openly, promoting collective progress and advocating for ethical and responsible AI research practices.

Acknowledgments

We want to thank the anonymous reviewers for their valuable comments. This work is supported by the Fundamental Research Grant Scheme (FRGS), Ministry of Higher Education Malaysia, under Grant FRGS/1/2023/ICT02/USM/02/4.

References

Xiaoyi Bao, Z Wang, Xiaotong Jiang, Rong Xiao, and Shoushan Li. 2022. Aspect-based sentiment analysis with opinion tree generation. *IJCAI 2022*, pages 4044–4050.

Jiahao Bu, Lei Ren, Shuang Zheng, Yang Yang, Jingang Wang, Fuzheng Zhang, and Wei Wu. 2021. Asap: A chinese review dataset towards aspect category sentiment analysis and rating prediction. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 2069–2079.

Hongjie Cai, Yaofeng Tu, Xiangsheng Zhou, Jianfei Yu, and Rui Xia. 2020. Aspect-category based sentiment analysis with hierarchical graph convolutional network. In *Proceedings of the 28th international conference on computational linguistics*, pages 833–843.

Hongjie Cai, Rui Xia, and Jianfei Yu. 2021. 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)*, pages 340–350.

Shaowei Chen, Yu Wang, Jie Liu, and Yuelin Wang. 2021. Bidirectional machine reading comprehension for aspect sentiment triplet extraction. In *Proceedings of the AAAI conference on artificial intelligence*, volume 35, pages 12666–12674.

Zhifang Fan, Zhen Wu, Xinyu Dai, Shujian Huang, and Jiajun Chen. 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)*, pages 2509–2518.

Tianhao Gao, Jun Fang, Hanyu Liu, Zhiyuan Liu, Chao Liu, Pengzhang Liu, Yongjun Bao, and Weipeng Yan. 2022. LEGO-ABSA: A prompt-based task assemblable unified generative framework for multi-task aspect-based sentiment analysis. In *Proceedings of the 29th International Conference on Computational Linguistics*, pages 7002–7012, Gyeongju, Republic of Korea. International Committee on Computational Linguistics.

Zhibin Gou, Yujiu Yang, et al. 2023. Mvp: Multi-view prompting improves aspect sentiment tuple prediction. In *The 61st Annual Meeting Of The Association For Computational Linguistics*.

Cao Duy Hoang, Quang Vinh Dinh, and Ngoc Hong Tran. 2022. Aspect-category-opinion-sentiment extraction using generative transformer model. In *2022 RIVF International Conference on Computing and Communication Technologies (RIVF)*, pages 1–6. IEEE.

Mengting Hu, Yinhao Bai, Yike Wu, Zhen Zhang, Liqi Zhang, Hang Gao, Shiwan Zhao, and Minlie Huang. 2023. Uncertainty-aware unlikelihood learning improves generative aspect sentiment quad prediction. In *Findings of the Association for Computational Linguistics: ACL 2023*, pages 13481–13494.

Mengting Hu, Yike Wu, Hang Gao, Yinhao Bai, and Shiwan Zhao. 2022. Improving aspect sentiment quad prediction via template-order data augmentation. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, pages 7889–7900.

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

Mike Lewis, Yinhan Liu, Naman Goyal, Marjan Ghazvininejad, Abdelrahman Mohamed, Omer Levy, Veselin Stoyanov, and Luke Zettlemoyer. 2020. Bart: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 7871–7880.

Zhengyan Li, Yicheng Zou, Chong Zhang, Qi Zhang, and Zhongyu Wei. 2021. 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*, pages 246–256.

Bin Liang, Wangda Luo, Xiang Li, Lin Gui, Min Yang, Xiaoqi Yu, and Ruifeng Xu. 2021. Enhancing aspect-based sentiment analysis with supervised contrastive learning. In *Proceedings of the 30th ACM international conference on information & knowledge management*, pages 3242–3247.

Nankai Lin, Yingwen Fu, Xiaotian Lin, Dong Zhou, Aimin Yang, and Shengyi Jiang. 2023. Cl-xabsa: Contrastive learning for cross-lingual aspect-based sentiment analysis. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 31:2935–2946.

Jian Liu, Zhiyang Teng, Leyang Cui, Hanmeng Liu, and Yue Zhang. 2021. Solving aspect category sentiment analysis as a text generation task. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 4406–4416, Online and Punta Cana, Dominican Republic. Association for Computational Linguistics.

Jiangming Liu and Yue Zhang. 2017. 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*, pages 572–577, Valencia, Spain. Association for Computational Linguistics.

Pengfei Liu, Weizhe Yuan, Jinlan Fu, Zhengbao Jiang, Hiroaki Hayashi, and Graham Neubig. 2023. Pre-train, prompt, and predict: A systematic survey of prompting methods in natural language processing. *ACM Computing Surveys*, 55(9):1–35.

Dehong Ma, Sujian Li, Fangzhao Wu, Xing Xie, and Houfeng Wang. 2019. Exploring sequence-to-sequence learning in aspect term extraction. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 3538–3547, Florence, Italy. Association for Computational Linguistics.

Yue Mao, Yi Shen, Jingchao Yang, Xiaoying Zhu, and Longjun Cai. 2022. Seq2path: Generating sentiment tuples as paths of a tree. In *Findings of the Association for Computational Linguistics: ACL 2022*, pages 2215–2225.

Yue Mao, Yi Shen, Chao Yu, and Longjun Cai. 2021. A joint training dual-mrc framework for aspect based sentiment analysis. In *Proceedings of the AAAI conference on artificial intelligence*, volume 35, pages 13543–13551.

Fei Mi, Yasheng Wang, and Yitong Li. 2022. Cins: Comprehensive instruction for few-shot learning in task-oriented dialog systems. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, pages 11076–11084.

Haiyun Peng, Lu Xu, Lidong Bing, Fei Huang, Wei Lu, and Luo Si. 2020. Knowing what, how and why: A near complete solution for aspect-based sentiment analysis. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 8600–8607.

Joseph Peper and Lu Wang. 2022. Generative aspect-based sentiment analysis with contrastive learning and expressive structure. In *Findings of the Conference on Empirical Methods in Natural Language Processing*.

Guang Qiu, Bing Liu, Jiajun Bu, and Chun Chen. 2011. Opinion word expansion and target extraction through double propagation. *Computational linguistics*, 37(1):9–27.

Colin Raffel, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li, and Peter J Liu. 2020. Exploring the limits of transfer learning with a unified text-to-text transformer. *The Journal of Machine Learning Research*, 21(1):5485–5551.

Timo Schick and Hinrich Schütze. 2021. Few-shot text generation with natural language instructions. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 390–402.

Hooman Sedghamiz, Shivam Raval, Enrico Santus, Tuka Alhanai, and Mohammad Ghassemi. 2021. Supcl-seq: Supervised contrastive learning for downstream optimized sequence representations. In *2021 Findings of the Association for Computational Linguistics, Findings of ACL: EMNLP 2021*, pages 3398–3403. Association for Computational Linguistics (ACL).

Ronald Seoh, Ian Birle, Mrinal Tak, Haw-Shiuan Chang, Brian Pinette, and Alfred Hough. 2021. Open aspect target sentiment classification with natural language prompts. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 6311–6322.

Yi Tay, Luu Anh Tuan, and Siu Cheung Hui. 2018. Learning to attend via word-aspect associative fusion for aspect-based sentiment analysis. In *Proceedings of the AAAI conference on artificial intelligence*, volume 32.

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

Siddharth Varia, Shuai Wang, Kishaloy Halder, Robert Vacareanu, Miguel Ballesteros, Yassine Benajiba, Neha Anna John, Rishita Anubhai, Smaranda Muresan, and Dan Roth. 2022. Instruction tuning for few-shot aspect-based sentiment analysis. *arXiv preprint arXiv:2210.06629*.

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

Bing Wang, Liang Ding, Qihuang Zhong, Ximing Li, and Dacheng Tao. 2022. A contrastive cross-channel data augmentation framework for aspect-based sentiment analysis. In *Proceedings of the 29th International Conference on Computational Linguistics*, pages 6691–6704.

Yequan Wang, Minlie Huang, Xiaoyan Zhu, and Li Zhao. 2016. Attention-based LSTM for aspect-level sentiment classification. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pages 606–615, Austin, Texas. Association for Computational Linguistics.

Zengzhi Wang, Rui Xia, and Jianfei Yu. 2024. Unified absa via annotation-decoupled multi-task instruction tuning. *IEEE Transactions on Knowledge and Data Engineering*.

Chao Wu, Qingyu Xiong, Hualing Yi, Yang Yu, Qiwu Zhu, Min Gao, and Jie Chen. 2021a. Multiple-element joint detection for aspect-based sentiment analysis. *Knowledge-Based Systems*, 223:107073.

Shengqiong Wu, Hao Fei, Yafeng Ren, Donghong Ji, and Jingye Li. 2021b. Learn from Syntax: Improving Pair-wise Aspect and Opinion Terms Extraction with Rich Syntactic Knowledge. *IJCAI International Joint Conference on Artificial Intelligence*, pages 3957–3963.

Haoliang Xiong, Zehao Yan, Chuhan Wu, Guojun Lu, Shiguan Pang, Yun Xue, and Qianhua Cai. 2023. Bart-based contrastive and retrospective network for aspect-category-opinion-sentiment quadruple extraction. *International Journal of Machine Learning and Cybernetics*, pages 1–13.

Lingling Xu and Weiming Wang. 2023. Improving aspect-based sentiment analysis with contrastive learning. *Natural Language Processing Journal*, 3:100009.

Lu Xu, Hao Li, Wei Lu, and Lidong Bing. 2020. Position-aware tagging for aspect sentiment triplet extraction. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 2339–2349.

Yichun Yin, Furu Wei, Li Dong, Kaimeng Xu, Ming Zhang, and Ming Zhou. 2016. Unsupervised word and dependency path embeddings for aspect term extraction. In *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence*, pages 2979–2985.

Hao Zhang, Yu-N Cheah, Osamah Mohammed Alyasiri, and Jieyu An. 2024a. Exploring aspect-based sentiment quadruple extraction with implicit aspects, opinions, and chatgpt: a comprehensive survey. *Artificial Intelligence Review*, 57(2):17.

Shengyu Zhang, Linfeng Dong, Xiaoya Li, Sen Zhang, Xiaofei Sun, Shuhe Wang, Jiwei Li, Runyi Hu, Tianwei Zhang, Fei Wu, et al. 2023. Instruction tuning for large language models: A survey. *arXiv preprint arXiv:2308.10792*.

Wenxuan Zhang, Yang Deng, Xin Li, Yifei Yuan, Lidong Bing, and Wai Lam. 2021a. Aspect sentiment quad prediction as paraphrase generation. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 9209–9219.

Wenxuan Zhang, Xin Li, Yang Deng, Lidong Bing, and Wai Lam. 2021b. 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)*, pages 504–510.

Wenxuan Zhang, Xin Li, Yang Deng, Lidong Bing, and Wai Lam. 2022. A survey on aspect-based sentiment analysis: tasks, methods, and challenges. *IEEE Transactions on Knowledge and Data Engineering*.

Wenyuan Zhang, Xinghua Zhang, Shiyao Cui, Kun Huang, Xuebin Wang, and Tingwen Liu. 2024b. Adaptive data augmentation for aspect sentiment quad prediction. *arXiv preprint arXiv:2401.06394*.

He Zhao, Longtao Huang, Rong Zhang, Quan Lu, and Hui Xue. 2020. Spanmlt: A span-based multi-task learning framework for pair-wise aspect and opinion terms extraction. In *Proceedings of the 58th annual meeting of the association for computational linguistics*, pages 3239–3248.

Jie Zhou, Jimmy Xiangji Huang, Qin Chen, Qinmin Vivian Hu, Tingting Wang, and Liang He. 2019. Deep learning for aspect-level sentiment classification: survey, vision, and challenges. *IEEE access*, 7:78454–78483.

Linan Zhu, Minhao Xu, Yinwei Bao, Yifei Xu, and Xiangjie Kong. 2022. Deep learning for aspect-based sentiment analysis: a review. *PeerJ Computer Science*, 8:e1044.

A Software and Hardware

Efficient execution of ASQP requires the right software and hardware setup. This section details the computational platforms and tools used in our research.

- **Software Configuration:** The research operates on Ubuntu 9.4.0 for stability and compatibility. Python 3.8 is used for its powerful libraries and community support. The framework for neural network modelling is PyTorch 1.7.0, with Transformers 4.14.1, sentencepiece 0.1.97, and PyTorch Lightning 0.8.1, simplifying the training processes.
- **Hardware Configuration:**
  - **Instance A:** Equipped with 12 vCPUs (Intel 8255C), 43GB RAM, and an NVIDIA RTX 3090 GPU (24GB). Used for T5-base (220M) model training.
  - **Instance B:** Equipped with 15 vCPUs (AMD EPYC 7543), 80GB RAM, and an NVIDIA A40 GPU (48GB). Used for T5-large (770M) model training.

B Results of IT and IT+SCL Methods

Table 5 compares the IT and IT+SCL methods across Restaurant and Laptop datasets, highlighting metrics such as loss, precision (P), recall (R), and F1 scores for T5-base and T5-large models. IT+SCL consistently outperforms IT alone, indicating SCL's enhancement of sentiment analysis. For example, in the Restaurant dataset, IT+SCL with T5-base achieves an F1 score of 61.18 at epoch 50, compared to 60.42 for IT. The T5-large model performs better, with IT+SCL achieving an F1 score of 64.86 at epoch 30, significantly higher than the T5-base's 61.18 at epoch 50. Similar improvements are seen across all datasets, with the largest gains in the Restaurant dataset. In the Laptop dataset, IT+SCL with T5-large achieves an F1 score of 46.11 at epoch 35, compared to 44.84 for IT. Fine-tuning the learning rate to 9×10^{-5} also results in notable performance gains. Performance generally improves with more training epochs until gains stabilize or slightly fluctuate. These findings demonstrate the effectiveness of incorporating SCL into the IT framework, the benefits of using larger models, and the importance of fine-tuning.

C Results of ASQP-ITSCL with Implicit and Explicit Sentiment Combinations

The analysis of Table 6 reveals that T5-large generally outperforms T5-base across various configurations on the Restaurant and Laptop datasets. Both models achieve their highest F1 scores with explicit aspect and opinion combinations (EAEO). T5-large performs best at Epoch 30 for the Restaurant dataset and Epoch 35 for the Laptop dataset, while T5-base performs best at Epoch 50 for both datasets. The models show moderate performance in handling implicit aspects with explicit opinions (IAEO) and implicit aspects with implicit opinions (IAIO), but they particularly struggle with explicit aspects and implicit opinions (EAIO). The Restaurant dataset results are consistently better than those for the Laptop dataset, suggesting that the dataset's nature or quality may impact performance.

D Mutil-Epoch SNE Representations

To understand the SCL objective's impact on hidden representations, t-SNE visualizations of the mean-pooled final encoder layer were generated (Van der Maaten and Hinton, 2008). These plots (Figures 5 to 8) show the model's ability to distinguish between implicit and explicit aspects, opinions, and sentiment polarities at different training epochs for the restaurant dataset. As training progresses, the model gradually improves in distinguishing various aspects, opinions, and sentiment polarities. Early epochs (e.g., E=1, E=5) show chaotic data distributions, while later epochs (e.g., E=40, E=55) show clearer separation. At E=50, the model achieves the best results, with the highest F1 score, as shown in Table 6. These results validate the model's effectiveness in implicit and explicit sentiment analysis. With more training epochs, the model forms clearer and more distinct clusters, highlighting the importance of contrastive learning and fine-tuning. This demonstrates the significant advantages of the ITSCL method.

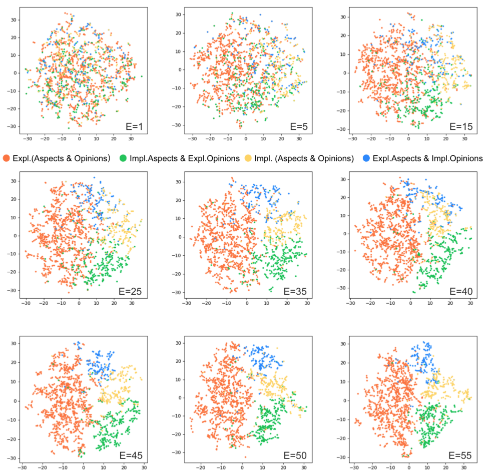

Figure 5: T-SNE visualization of aspect & opinion representations on the Restaurant dataset.

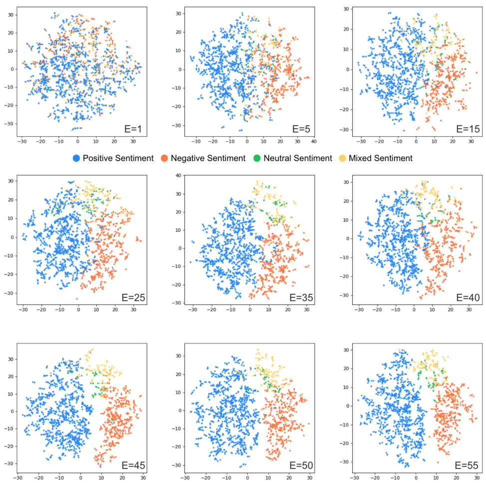

Figure 6: T-SNE visualization of sentiment representations on the Restaurant dataset.

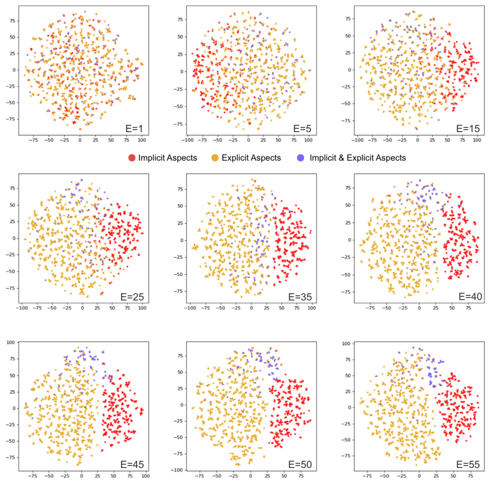

Figure 7: T-SNE visualization of aspect representations on the Restaurant dataset.

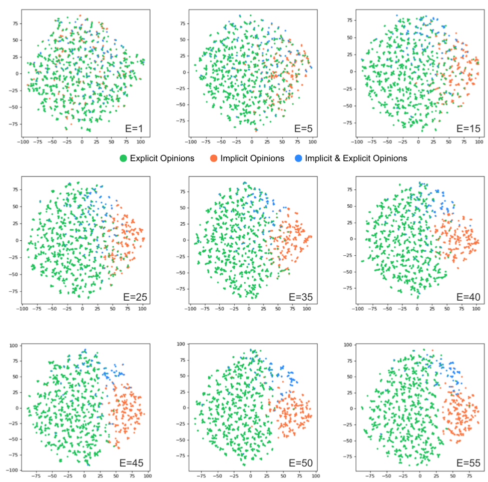

Figure 8: T-SNE visualization of opinion representations on the Restaurant dataset.

E Origin and new Category Labels for Rest and Laptop datasets

Comparative analysis of original and new prompt formats reveals that simplifying category descriptions, from ‘RESTAURANT#GENERAL‘ to ‘restaurant general‘, improves model efficiency without sacrificing comprehension. Therefore, this experiment will use the New Category Label instead of the origin Category Label using human-readable descriptive category labels, as shown in Table 7.

F Manual Prompt Design Detail

The input template provides a clear example, including a sample sentence and labels for aspect, opinion, sentiment, and category. A manual input prompt with specific instructions is shown. For instance, in "This place has got to be the best Japanese restaurant in the New York area," the model identifies the Aspect ("restaurant"), Opinion ("best"), Sentiment ("positive"), and Category ("restaurant general"). Table 8 shows the methodology with prefix examples for context, categorizing items into restaurant and laptop reviews. For the experiment, "This place has got to be the best Japanese restaurant in the New York area" is used for the Restaurant dataset, and "The laptop struggles with high-end games" for the Laptop dataset. These sentences are generated by ChatGPT and manually annotated. A suffix prompt with $TEXT as a placeholder structures the contextual information.

G Hyperparameters in the SCL

We report the parameters used in the supervised contrastive learning (SCL) objective in Table 9. α is the loss weighting factor, and τ is the temperature value determining how severely to punish hard negative examples. A learning rate of 3×10^{-4} enables faster convergence with fewer training cycles but is prone to overfitting. In contrast, a learning rate of 9×10^{-5} results in slower, more stable convergence over longer training cycles.

H Impact of Temperature τ

Figure 9 shows the F1 score variation with different temperature parameters (τ), which determines how severely to punish hard negative examples. The F1 score is highest, and the model performs best when τ is between 0.1 and 0.3. The F1 score is highest at $\tau = 0.25$. An excessively high τ (e.g., 0.3) significantly decreases performance. Properly adjusting τ is crucial for optimizing model performance in capturing implicit sentiments and opinions.

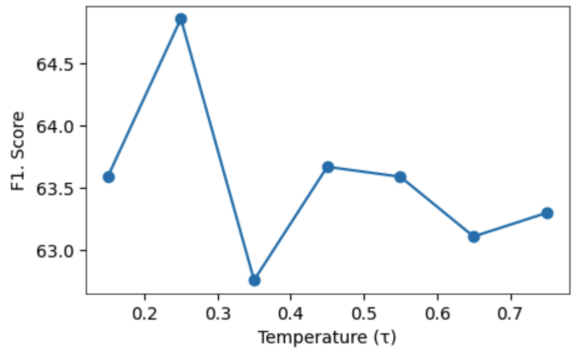

Figure 9: F1 Score Variation with Temperature (τ) in Restaurant Dataset.

I Algorithm

In this appendix, we provide a detailed description of the implementation process of the ITSCL algorithm. The Algorithm 1 are the specific steps of the ITSCL algorithm. This algorithm aims to enhance the accuracy and robustness of sentiment analysis by combining instruction tuning and supervised contrastive learning methods. The core idea of the algorithm is to use manual prompt design and multi-layered feature representations to optimize the model's learning effectiveness, especially in handling implicit sentiments.

Algorithm 1 Inference

Require: Dataset $D = \{D_{\text{train}}, D_{\text{val}}, D_{\text{test}}\}$, Pre-trained Model M, Temperature τ
Ensure: Fine-tuned Model M with Manual Prompts and SCL

Initialize model M with pre-trained weights
Set learning rate, loss function, and number of epochs
for each epoch e from 1 to E **do**
    for each sample (x, y) in D_{train} **do**
        Step 1: Manual Prompt Design and Instruction Tuning
        Define the input prompt template P_{input} for x (e.g., "Analyze the sentiment, aspect, opinion, and category for the following text: x")
        Define the output prompt template P_{output} for y (e.g., "Aspect: a, Opinion: o, Sentiment: s, Category: c")
        Encode input x using the prompt P_{input} to get x_P
        Forward pass x_P through M to get raw prediction $\hat{y}$
        Format $\hat{y}$ to match the output prompt template P_{output} to get formatted prediction $\hat{y}_P$
        Compute cross-entropy loss $\mathcal{L}_{\text{CE}}$ using:

$$\mathcal{L}_{\text{CE}} = -\sum_{i=1}^{N} y_i \log(\hat{y}_{P,i}) \tag{4}$$

    end for
    Step 2: Supervised Contrastive Learning (SCL) with Layer-specific Losses
    for each batch B in D_{train} **do**
        Forward pass B through M to get representations h_i for each layer
        Compute sentiment characteristic-specific SCL loss $\mathcal{L}_{\text{sent}}$
        Compute aspect characteristic-specific SCL loss $\mathcal{L}_{\text{aspect}}$
        Compute opinion characteristic-specific SCL loss $\mathcal{L}_{\text{opinion}}$
        Compute joint representation characteristic-specific SCL loss $\mathcal{L}_{\text{joint}}$
        Compute total SCL loss $\mathcal{L}_{\text{SCL}}$ using:

$$\mathcal{L}_{\text{SCL}} = \mathcal{L}_{\text{sent}} + \mathcal{L}_{\text{aspect}} + \mathcal{L}_{\text{opinion}} + \mathcal{L}_{\text{joint}} \tag{5}$$

        Compute total loss $\mathcal{L}$ using:

$$\mathcal{L} = \mathcal{L}_{\text{CE}} + \mathcal{L}_{\text{SCL}} \tag{6}$$

        Backpropagate total loss $\mathcal{L}$ and update model weights
    end for
    Step 3: Validate on Validation Set
    Validate M on validation set D_{val}
    Adjust manual prompt templates P_{input} and P_{output} and model parameters based on validation performance metrics (e.g., accuracy, F1 score)
end for
Step 4: Final Evaluation
Evaluate the fine-tuned model M on the test dataset D_{test} and record the performance metrics
    return Fine-tuned Model M

Epoch	Restaurant (IT) (T5-base)				(IT+SCL) (T5-base)			
	loss	P.	R.	F1.	loss	P.	R.	F1.
E=5	0.038	53.66	47.27	50.26	14.31	57.84	54.37	56.05
E=10	0.024	58.44	54.80	56.56	8.333	60.25	56.77	58.46
E=15	0.015	58.39	55.46	56.89	5.935	59.71	57.75	58.71
E=20	0.010	60.11	58.41	59.25	4.445	60.18	59.06	59.61
E=25	0.006	60.00	58.62	59.30	3.335	61.14	59.93	60.53
E=30	0.005	59.51	58.08	58.78	2.924	61.65	60.37	61.00
E=35	0.003	59.00	57.97	58.48	2.522	60.85	59.39	60.11
E=40	0.0028	**60.68**	**60.15**	**60.42**	2.421	61.09	60.15	60.62
E=45	0.0023	59.73	58.62	59.17	2.392	60.11	59.06	59.58
E=50	0.0019	58.75	58.30	58.52	2.246	**61.45**	**60.92**	**61.18**
E=55	0.0018	59.76	58.84	59.30	2.056	60.51	60.04	60.27
	Restaurant (IT) (T5-large)				(IT+SCL) (T5-large)			
	loss	P.	R.	F1.	loss	P.	R.	F1.
E=5	0.064	64.73	60.92	62.77	6.750	63.38	60.26	61.78
E=10	0.017	63.81	61.03	62.39	3.528	63.99	61.68	62.81
E=15	0.009	64.39	61.79	63.06	2.595	63.40	61.46	62.42
E=20	0.004	64.05	61.46	62.73	2.444	63.35	61.90	62.62
E=25	0.0017	62.91	61.46	62.18	2.087	62.93	62.45	62.68
E=30	0.0014	**63.97**	**62.23**	**63.09**	2.006	**65.56**	**64.19**	**64.86**
E=35	0.0018	63.11	62.01	62.56	1.966	62.89	61.79	62.33
Epoch	Laptop (IT) (T5-base)				(IT+SCL) (T5-base)			
	loss	P.	R.	F1.	loss	P.	R.	F1.
E=5	0.038	40.45	37.38	38.85	15.56	42.01	41.00	41.50
E=10	0.023	43.72	41.69	42.68	9.984	44.30	44.44	44.37
E=15	0.014	44.24	43.67	43.95	7.404	44.37	44.10	44.23
E=20	0.011	44.20	43.67	43.93	5.066	44.22	43.50	43.86
E=25	0.007	43.95	43.50	43.72	3.851	42.52	42.38	42.45
E=30	0.004	**44.26**	**43.84**	**44.05**	3.128	43.01	42.46	42.73
E=35	0.003	43.13	42.98	43.05	2.859	43.03	42.55	42.79
E=40	0.0027	42.32	41.77	42.05	2.404	43.34	42.89	43.12
E=45	0.0018	43.25	43.07	43.16	2.312	43.87	43.50	43.69
E=50	0.0025	43.45	43.41	43.43	2.102	**44.69**	**44.19**	**44.43**
E=55	0.0015	42.88	42.55	42.72	2.153	44.49	44.19	44.34
	Laptop (IT) (T5-large)				(IT+SCL) (T5-large)			
	loss	P.	R.	F1.	loss	P.	R.	F1.
E=5	0.064	45.14	44,36	44.74	9.351	45.21	44.70	44.95
E=10	0.026	43.40	43.58	43.49	4.631	45.53	44.70	45.11
E=15	0.010	44.89	44.27	44.58	2.966	44.67	44.79	44.73
E=20	0.005	43.88	43.84	43.86	2.310	44.73	44.27	44.50
E=25	0.0029	44.71	44.44	44.58	2.044	44.58	44.62	44.60
E=30	0.0025	43.68	43.76	43.72	2.017	44.70	44.27	44.48
E=35	0.0015	**44.97**	**44.70**	**44.84**	1.935	**46.31**	**45.91**	**46.11**

Table 5: Comparison of IT and IT+SCL methods. (learning rate set to 9e-5 and τ set to 0.25)

Dataset	Model	Type	Gold	Pred.	Hit	P.	R.	F1.
Restaurant	T5-base	EAEO	596	625	426	68.16	71.48	69.78
Restaurant	T5-base	IAEO	122	128	64	50.00	52.46	51.20
Restaurant	T5-base	EAIO	107	75	29	38.67	27.10	31.87
Restaurant	T5-base	IAIO	91	80	39	48.75	42.86	45.61
Total (Epoch=50)			916	908	558	61.45	60.92	61.18
Restaurant	T5-large	EAEO	596	633	441	69.67	73.99	**71.77**
Restaurant	T5-large	IAEO	122	130	67	51.54	54.92	**53.17**
Restaurant	T5-large	EAIO	107	64	38	59.38	35.51	**44.44**
Restaurant	T5-large	IAIO	91	70	42	60.00	46.15	**52.17**
Total (Epoch=30)			916	897	**588**	65.56	64.19	**64.86**
Laptop	T5-base	EAEO	673	714	322	45.10	47.85	46.43
Laptop	T5-base	IAEO	169	146	93	63.70	55.03	59.05
Laptop	T5-base	EAIO	253	229	73	31.88	28.85	30.29
Laptop	T5-base	IAIO	66	59	25	42.37	37.88	**40.00**
Total (Epoch=50)			1161	1148	513	44.69	44.19	44.43
Laptop	T5-large	EAEO	673	712	327	45.93	48.59	**47.22**
Laptop	T5-large	IAEO	169	154	99	64.29	58.58	**61.30**
Laptop	T5-large	EAIO	253	230	83	36.09	32.81	**34.37**
Laptop	T5-large	IAIO	66	55	24	43.64	36.36	39.67
Total (Epoch=35)			1161	1151	**533**	46.31	45.91	**46.11**

Table 6: Results of ASQP-SCL Model with EAEO, IAEO, EAIO and IAIO in Epoch=50.

Dataset	**Origin Category Label**	**New Category Label**
REST	RESTAURANT#GENERAL FOOD#STYLE_OPTIONS FOOD#QUALITY	restaurant general food style_options food quality
LAPTOP	LAPTOP#OPERATION_PERFORMANCE OS#DESIGN_FEATURES SHIPPING#GENERAL	laptop functionality operating system features shipping general

Table 7: Origin and new Category Labels for REST and LAPTOP Datasets.

Ablation	**Input Prompt**
Prefix	Example: this place has got to be the best japanese restaurant in the new york area. |
Restaurant(Origin)	aspect term is place, opinion term is best, category is RESTAURANT#GENERAL, and sentiment is postive.
Restaurant(New)	aspect term is place, opinion term is best, category is restaurant general, and sentiment is postive.
Prefix	Example: the laptop struggles with high-end games. |
Laptop(Origin)	aspect term is laptop, opinion term is struggles, category is LAPTOP#OPERATION_PERFORMANCE, and sentiment is negative.
Laptop(New)	aspect term is laptop, opinion term is struggles, category is laptop functionality, and sentiment is negative.
Suffix	Now, Given the sentence: $TEXT $TEXT is the placeholder for the ASQP processing sentence.

Table 8: Examples of ASQP-IT input template.

	α	τ	learning rate	T5-base Epoch	T5-large Epoch	Batch size
RESTAURANT	0.05	0.25	3e-4&9e-5	(5, 10, …, 55)	(5, 10, …, 35)	16
LAPTOP	0.05	0.25	3e-4&9e-5	(5, 10, …, 55)	(5, 10, …, 35)	16

Table 9: The hyperparameters used in the SCL.