ISAPrompt: Advanced Prompting for Implicit Sentiment Analysis

Anonymous ACL submission

Abstract

With the increasing social demand for fine-grained sentiment analysis (SA), implicit sentiment analysis (ISA) has become crucial, as its opinion cues are often presented in an implicit and obscure manner. Therefore, extracting hidden opinion words is necessary to detect implicit sentiment. Recently, Prompt-oriented fine-tuning (POFT) has achieved promising results in a variety of text classification tasks. However, most of them lack attention to detect implicit sentiment. In this paper, we introduce ISAPrompt which is an improved method based on POFT, designed to extract hidden human perspectives for ISA. Improvements were made to the structure of prompts and the design of verbalizers. We also designed the ISAPrompt ensemble trick, which can significantly improve the overall performance of our method. The effectiveness of the proposed ISAPrompt and ISAPrompt ensemble was validated on two benchmark datasets. The results showcase that the approach achieves state-of-the-art (SOTA) performance in implicit sentiment analysis tasks.

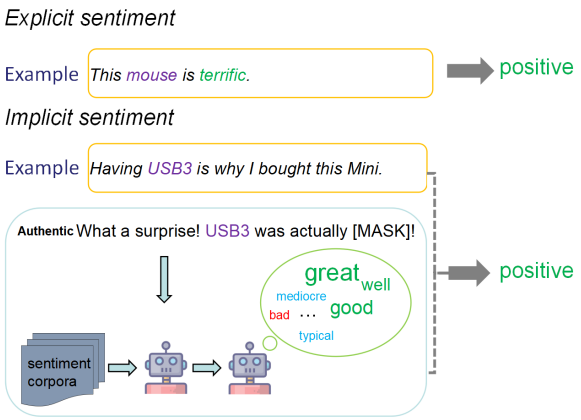

Figure 1: Explicit sentiment expression allows for straightforward inference, whereas detecting implicit sentiment requires deeper analysis to identify hidden opinion words. In this context, ISAPrompt significantly enhances the model's capacity to identify and interpret implicit sentiments, thereby contributing to a more nuanced understanding of affective content.

1 Introduction

Sentiment analysis (SA) focused on identifying and extracting emotional information from text. Traditionally, sentiment analysis (SA) can be categorized into explicit sentiment analysis (ESA) and implicit sentiment analysis (ISA). ESA, which focuses on directly expressed sentiments, is currently the mainstream task in the field (Pontiki et al., 2014). In recent years, Aspect-based Sentiment Analysis (ABSA), a specialized branch of SA (Cui et al., 2023), has garnered significant interest within the academic community. The predominant methodologies for tackling Aspect-Based Sentiment Analysis (ABSA) can be categorized into two types: methods that utilize attention mechanisms (Ma et al., 2017) and approaches that focus on constructing opinion trees (Bao et al., 2023). Both types methods are suitable for handling ESA, aiming to identify opinion words related to the target in the input text, which are clear indicators of sentiment polarity. However, these approaches pay little attention to ISA situations. ISA involves sentiment expressions that consist only of factual descriptions and lack explicit opinions, yet still convey clear sentiment polarity that can be discerned in context (Russo et al., 2015). For instance, given a text from the laptop dataset "*Having USB3 is why I bought this Mini.*" . This sentence primarily outlines the motivations for acquiring the 'Mini' model without explicitly employing language that signifies sentiment towards USB3. Nonetheless, a distinctly positive interpretation can be discerned. Furthermore, implicit sentiment expressions are prevalent in aspect-based sentiment recognition. As demonstrated in Table 2, 27.47% of reviews in the Restaurant dataset and 30.09% in the Laptop dataset exhibit implicit sentiment. Consequently, the analysis of implicit sentiment is both crucial

and challenging.

Indeed, pinpointing hidden opinions is crucial for the precision of Implicit Sentiment Analysis (ISA). As illustrated in Figure 1, in the example of Explicit Sentiment Analysis (ESA), the presence of distinct opinion words associated with the target "mouse" enables us to accurately deduce the positive sentiment directed towards it. However, in the ISA task, the target's opinion words are not provided directly. To precisely ascertain the implicit opinions regarding target words in ISA, a wealth of knowledge and a profound level of comprehension are indispensable.

Fortunately, the successful development of large-scale pre-trained language models (LLMs) can offer substantial support in this regard. LLMs possess extensive world knowledge and demonstrate remarkable capabilities in common-sense understanding (Paranjape et al., 2021), which makes them particularly effective for ISA task. To leverage the full potential of LLMs, many text classification efforts have adopted prompt-oriented fine-tuning (Schick and Schütze, 2020; Liu et al., 2023) to help better use knowledge in LLMs. Most prompt-oriented fine-tuning combines the original input with a specially crafted prompt template to predict [MASK], as illustrated in the implicit sentiment example in Figure 1. Then, LLMs based on the surrounding context return the probability distribution over the given label words for the [MASK]. The mapping from label words to the specific class called verbalizer (Hu et al., 2021). It also fine-tunes LLMs using these templated inputs, optimizing its ability to understand and accurately fill in the [MASK] position.

ISAPrompt is a new prompt-oriented fine-tuning method. Different from conventional POFT, ISAPrompt has two major improvements: Firstly, according to the above definition of ISA, we can analyze the real opinion words associated with the target words by examining the differences in sentiment and factual conditions regarding the literal content in implicit texts. To this end, we designed prompts that focus on guiding LLMs to pay attention to the actual intent in the text, while avoiding situations where the LLMs overly focuses on opinions not explicitly present, which can lead to errors in recognizing clearly sentiment-laden opinion words. Take the implicit sentiment example in Figure 1, a well-crafted prompt like "What a surprise! USB3 was actually [MASK]! " mimics the speaker's impending revelation of the truth to encourage the model to convey the true intent. Secondly, Hu et al. (2021) point out that words mapped by a verbalizer should have two essential qualities: broad coverage and minimal subjective bias. Such a comprehensive projection is vital for effectively mimicking pre-training. Therefore, we designed a diverse set of label words through dataset analysis with assistance from ChatGPT, to make the label words more aligned with actual conditions. Furthermore, We apply ISAPrompt with multiple different prompt compositions to the same model to produce a more powerful and comprehensive model.

To sum up, our contributions can be listed as follows: (1) We propose ISAPrompt, an improved method based on prompt-oriented fine-tuning, with enhancements to prompts construction and the verbalizer, to better enable large language models to recognize implicit sentiment within text. (2) We introduce ISAPrompt ensemble, an enhanced approach derived from ISAPrompt, which significantly boosts the overall efficacy of our methodology. (3) The evaluation outcomes on two benchmark datasets demonstrate that using only ISAPrompt allows one 355M large language model to achieve ISA performance comparable to an 11B large language model at the state-of-the-art level, while employing ISAPrompt-ensemble further achieves state-of-the-art results in ISA performance.

2 Related Work

The task of SA (either ESA or ISA) can be defined as: given a sentence s and a target term $t \subset s$, we need to determine whether the sentiment towards t is positive, negative, or neutral. In this section, we review the literature relevant to our work, focusing on implicit sentiment analysis and the application of prompt-oriented fine-tuning techniques to leverage the capabilities of LLMs.

2.1 Implicit Sentiment Analysis

Sentiment analysis has long been a popular research topic within the NLP (Natural Language Processing) community (Pang et al., 2008). Most of the SA works above do not focus on ISA. However, ISA is often more ubiquitous in realistic scenarios (Fei et al., 2023). ISA need to detect sentiments that are subtly conveyed rather than explicitly stated, discerning emotions through contextual cues, tone, or nuanced expressions, etc. In the beginning, some research try to infer implicit senti-

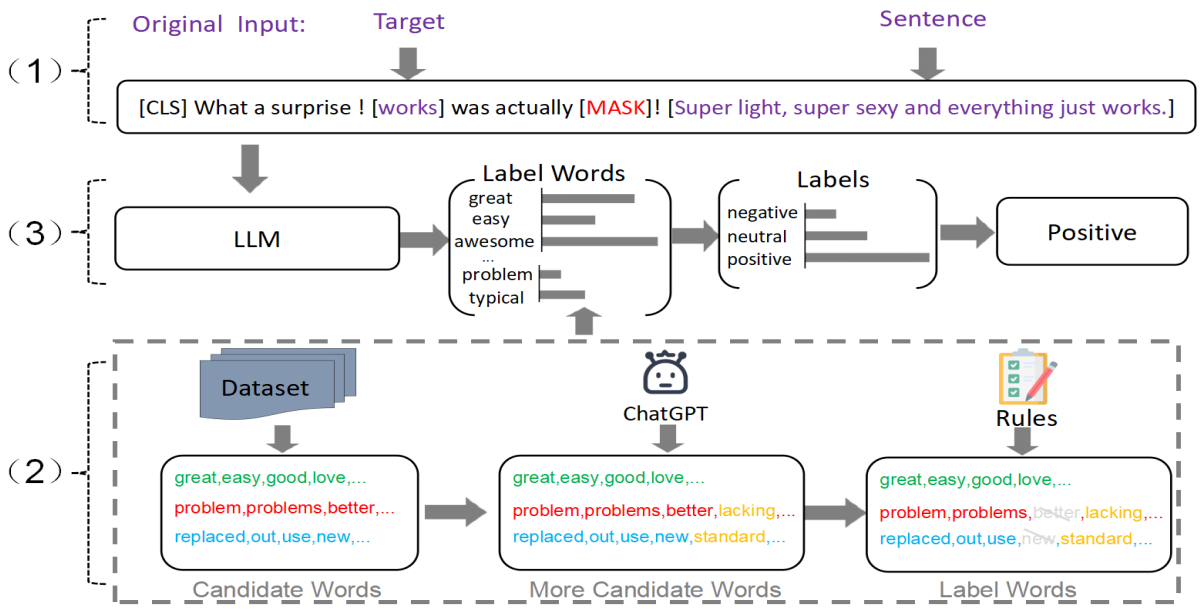

Figure 2: The illustration of ISAprompt. Let's take our designed prompt T_3 as an example. (1) Combine the original input with the constructed prompt. (2) Finding label words, the dashed box illustrates the process where the verbalizer obtains the final label words. (3) Verbalizer Mapping, the verbalizer maps the LLM's predictions over the label words into labels.

ment through these details, Deng and Wiebe (2014) identified implicit sentiment through inference beyond explicit sentiment expressions. Liao et al. (2019) examine the characteristics of sentences that convey fact-based implicit sentiment and propose a hierarchical semantic fusion approach to learn feature representations for detecting implicit sentiment through deep feature learning. Recently, Large-scale pre-trained language models have received increasing research attention for enhancing the performance of downstream applications (Devlin, 2018), particularly in tasks such as implicit sentiment analysis. Xu et al. (2022) focus on implicit sentiment analysis that uses the sentiment of events in a sentence to determine its overall sentiment polarity. Experiments on both BiLSTM and BERT models showed that BERT, as a representative of LLMs, achieved the best performance. Li et al. (2021) adopt Supervised Contrastive Pre-training on large-scale sentiment annotated corpora. Aims to enable deep neural models to capture the contrast between implicit and explicit sentiments. Besides, to capture hidden information and find the true sentiment, pre-trained models trained on large-scale datasets are tried and applied to implicit sentiment analysis. Most recently, Wei et al. (2022) proves that chain of thoughts (CoT) can improve the ability of large language models to perform reasoning. CoT is applied by crafting prompts that encourage the model to break down its reasoning into clear, logical steps. After that, Fei et al. (2023) introduced a CoT framework to simulate the human reasoning process of implicit sentiemnt analysis. Using supervised fine-tuning of large language models and carefully designed prompts that guide thought chains to persuasively mine and predict latent emotions. However, this method rely on larger-scale LLMs, their good performance can only be achieved when the scale of LLMs reaches 11 billion parameters or more. So, this requires a significant amount of time and computational resources. Ouyang et al. (2024) generate explicit sentiment augmentations with a LLM to enhance implicit classification tendencies. Their method surpasses THOR on smaller-scale LLMs, but still lags significantly behind THOR when applied to LLMs with 11 billion parameters. This indicates that there is still significant room for improvement in using smaller-scale LLMs to handle implicit sentiment analysis tasks. Hence, ISAPrompt aims to further leverage the capabilities of LLMs to analyze implicit sentiments.

2.2 Prompt-oriented Fine-tuning

Inspired by the recent works utilizing language prompts to probe the knowledge in LLMs (Brown, 2020), more and more works are trying to adopt POFT. Moreover, many POFT methods have shown

impressive results in SA tasks, such as: Schick and Schütze (2020) introduce manual prompts to guide the model to understand specific tasks sentiment analysis. Chen et al. (2022) try to inject latent knowledge contained in relation labels into prompt for relation extraction to let LLMs better identify emotions. Seo et al. (2024) designed specific prompts to induce one LLM to generate training text, which is then used to fine-tune another LLM, thereby enhancing its ability to better judge emotions.

These works optimize POFT by designing prompts to fully leverage the capabilities of large language models. Besides developing effective prompts, the verbalizer plays a crucial role in POFT and significantly impacts its performance (Gao et al., 2020). Additionally, some approaches use Prompt-tuning (PT), and compared to POFT. PT freezes all the LLM parameters and only tunes the prompts to save computing resources (Gu et al., 2021). As the verbalizer serves an equivalent role in both PT and POFT, enhancements implemented in the verbalizer for PT can be directly transferred to POFT. Consequently, we have also taken into account the advancements made to the verbalizer in PT. Shin et al. (2020) develop automatic methods for searching verbalizers to optimize verbalizer selection. By this, the model can better capture subtle sentiment nuances, leading to improved classification accuracy. Hu et al. (2021) incorporating denoised external knowledge into the verbalizer to enhance and stabilize the performance of prompt tuning. As a result, LLMs have access to a wide range of words with clear emotional meanings to choose from, enabling them to provide more accurate responses. Liu et al. (2023) by analyzing the dataset, multiple label words can be assigned to each class. Additionally, incorporating sarcasm-related knowledge into the verbalizer allows the model to better recognize and distinguish sarcastic expressions.

Therefore, our method ISAPrompt aims to improve the structure of prompt to make LLMs focus on hidden opinions within sentences and developed a verbalizer that enables LLMs to express a broader range of perspectives, thereby enhancing the performance of implicit sentiment analysis tasks more efficiently. ISAPrompt not only reduces the reliance on model size, but also enhances the model's ability to uncover hidden opinion.

3 METHOD

In this section, we will describe our Prompt-oriented fine-tuning model for implicit sentiment recognition (ISAPrompt) in detail. The overview of ISAPrompt is shown in Figure 2. We start by introducing the construction of prompts. Next, we elaborate on verbalizer construction and utilization. And then elucidate the training objective. The above is the content of ISAPrompt. Finally, we present the ISAPrompt ensemble, which can greatly enhance the overall effectiveness of our approach.

$T_1(x, w) = x$ Believe it or not, w was actually $[mask]$!

$T_2(x, w) = x$ Who would have thought that w was actually $[mask]$!

$T_3(x, w) =$ What a surprise! w was actually $[mask]$! x

$T_4(x, w) =$ No one expected that w was actually $[mask]$! x

$T_5(x, w) = x$ It's hard to believe, but w was $[mask]$.

$T_6(x, w) =$ Surprisingly, w was so $[mask]$! x

Figure 3: Clash Prompts.

3.1 Prompt Construction

Recall that the representative characteristic of implicit sentiment analysis is that the text does not express the target sentiment words directly. The actual intention is often hidden behind the literal content. For implicit texts, an effective prompt template should reflect the underlying sentiment embedded within the original text, while for non-implicit texts, it should align with the intended meaning. Consequently, we use sentences related to facts in such a way that they emphasize attention to the opinions hidden within the sentences, without interfering with the extraction of sentiments from non-implicit sentiment texts. In addition, we use "!" in many places to enhance the semantic expression. We imitate speaker and design six clash prompts $T_1(x, w) \sim T_6(x, w)$. Phrases like "Believe it or not" and "What a surprise" are employed to dig out the true feelings about the target term w in the sentence x. All the prompts we designed have been presented in Figure 3.

3.2 Verbalizer Construction

By merging the original input with the prompt template, we can obtain a new input x_p. LLMs will predict the word at the masked position in the new input x_p. Predicting the masked word based on the context is not a single-choice task, meaning there isn't a single correct answer. And a variety

Table 1: Examples of the final label words for different datasets.

Dataset	Label	Label Words
Restaurant	Positive	good, great, delicious, excellent, best, friendly, fresh, nice...
	Negative	small, bad, rude, slow, worst, horrible, terrible, expensive...
	Neutral	average, mediocre, ok, OK, modern, Westernized, typical...
Laptop	Positive	great, easy, good, love, fast, well, nice, like, better...
	Negative	problem, problems, bad, slow, stopped, issues, poor, short, hard...
	Neutral	replaced, out, use, run, up, working, more, most, average...

of words from a rich vocabulary can appropriately fit the context. Therefore, verbalizer construction is very important. The aim of verbalizer construction is to identify the label words that should be placed in the masked position and then establish how these label words should be mapped to their corresponding labels. We followed the principle of 'words mapped by a verbalizer should have two essential qualities: broad coverage and minimal subjective bias' to construct the label words set. In this section, we introduce how we get label words and the detail about verbalizer mapping.

3.2.1 Finding label words

Candidate Words. We calculated the occurrence rates of opinion words that have been labeled to explicitly indicate sentiment target words in the restaurant and laptop datasets. Since we are using a single mask, for opinion words composed of multiple words like "solid sleek beautiful", we will split them into individual words for counting. After that we counted the number of times opinion words appeared and select the top 30 words as candidate words for the label words of each dataset. This not only avoids the interference of low-frequency words that do not fit the label, but also has good coverage. Besides, we choose top 30 high-frequency words is that some opinion words express sarcasm or humor, such as texts labeled with "negative" but containing opinion words like "better". If we choose "better" as a label word for the "negative" label, it would create significant ambiguity. We found that the lower the word frequency, the more likely such situations arise when organizing candidate words. Therefore, we prefer to select high-frequency words as much as possible. In this way, we get a corresponding reasonable set of candidate words for each dataset.

More Candidate Words. Due to the limited number of marked negative and neutral words in the restaurant and laptop datasets, many of the selected candidate words are unsuitable for the intended role and will be discarded. To ensure that the label words have extensive coverage, we design prompts and use ChatGPT to help us expand the list of negative and neutral candidate words. We use label as seeds to seek more words that fit the category. So that, our prompt is: "Now I want to perform aspect-based sentiment analysis on the [restaurant/laptop] dataset. Please provide some common words with similar meanings to '[positive/negatie/neutral]' to be used as candidate words for the predicted target opinion word." . This prompt can let ChatGPT return words that fit the theme of the dataset. We cannot allow words suitable for describing restaurants, such as "delicious", to be used to describe a laptop. And we select words whose meanings have not appeared in candidate words into candidate words.

Final Label Words. The collected words still can be noisy. For this, we processed the candidate words according to the established rules: (1) Each label word needs to fit its own label and can be fully encoded as a known word in LLMs; otherwise, discard them. (2) We only keep words with the parts of speech being adjectives, nouns, verbs, or adverbs to ensure they fit the prompt template. Finally, We get label words for each dataset and defined in Table 1. Then LLMs can output a probability distribution over the finely detailed label words set. In our method, the processing of each dataset differs only in the label words. Therefore, in the following sections, we will use the method on the Restaurant dataset as an example to further explain our approach, and we define V to represent the label words set for the Restaurant.

3.2.2 Verbalizer Mapping

After establishing the label words, we need to map the predicted probability of label words to the fi-

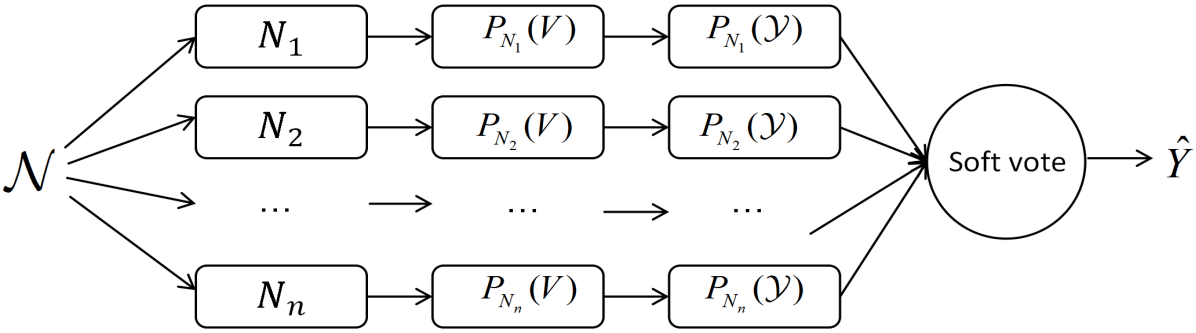

Figure 4: An example of ISAPrompt ensemble.

nal label y. Let M be a LLM, and let V_y denote the subset of V that is mapped to a specific label y, M can give the probability of each word v in the vocabulary being filled in $[MASK]$ token $P_M(\text{[MASK]} = v|\mathbf{x}_p)$. We presume that every label word within a class has an equal influence on the prediction of the label. So that, we use the average of label words predicted scores. And the probability of each label is:

$$P(\hat{y}) = \frac{\sum_{v \in V_y} P_M(\text{[MASK]} = v|\mathbf{x}_p)}{|V_y|} \quad (1)$$

3.3 Training Objective

Since the training set is available and our ISAPrompt was a prompt-oriented fine-tuning approach, we fine-tuned ISAPrompt under supervised learning. In the following text, we refer to these as ISAPrompt-tuned models. We computed the loss based on the supervised labels of the datasets and trained the model accordingly. The training objective considers cross-entropy loss in order to ensure the basic ability of sentiment recognition. And the specific formula is as follows:

$$Loss = \sum \text{cross-entropy}(y, P(\hat{y})) \quad (2)$$

We fine-tuned the pre-trained model for each prompt. Then we got a model set $\mathcal{M}$. Since we have designed six prompts, $\mathcal{M}$ contains six ISAPrompt-tuned models.

3.4 ISAPrompt ensemble

Inspired by Schick and Schütze (2020), we try to combine the effects of the prompts we designed to achieve better performance than any single ISAPrompt-tuned model, and we adopted Soft Voting to determine the final result.

We use all ISAPrompt-tuned models to annotate unlabeled data, and each ISAPrompt-tuned model will return a probability distribution over the three labels: positive, negative, and neutral, as described in Section 3.3. For the final prediction, we apply soft voting, where we average the probability distributions across all models and select the label with the highest average probability as the final prediction. Additionally, we can not only combine all models but also select specific models for combination. Choosing different combinations of ISAPrompt-tuned models also affects the final prediction. Since the number of ISAPrompt-tuned models is not large, we experimented with all possible combinations. We define $\mathcal{N}$ as one of the combinations from the set of ISAPrompt-tuned model combinations, where N represents a single ISAPrompt-tuned model that is part of the combination $\mathcal{N}$. Therefore, the ISAPrompt ensemble can be illustrated as shown in Figure 4, and for an ISAPrompt-tuned model combination, the predicted probability of each class is:

$$P(\hat{y}) = \max \frac{\sum_{N \in \mathcal{N}} P_N(y)}{|\mathcal{N}|} \quad (3)$$

4 EXPERIMENT

4.1 Datasets

In the experiments, we assess the results using the Restaurant and Laptop datasets from SemEval-2014 (Pontiki et al., 2014). All the instances in these two datasets are split into explicit and implicit sentiment by Li et al. (2021), and the statistics of the dataset are presented in Table 2.

4.2 Baselines

In this subsection, we introduce the baselines used for comparison. Baselines primarily include fine-tuning approaches based on pre-trained models and various state-of-the-art (SOTA) techniques.

Table 2: Statistics on the datasets.

Dataset	Positive	Neutral	Negative	Total	IS (%)
Restaurant-train	2164	805	633	3602	28.59
Restaurant-test	728	196	196	1120	23.84
Restaurant	2892	1001	829	4722	27.47
Laptop-train	987	866	460	2313	30.87
Laptop-test	341	128	169	638	27.27
Laptop	1328	994	629	2951	30.09

BERT+SPC (Devlin, 2018) A method that combines BERT with a sentence pair classification strategy, aims to classify the sentiment of a sentence towards a specific target or aspect.

BERT+ADA (Rietzler et al., 2019) The approach combines BERT's ability to capture rich contextual representations with an adaptive attention mechanism that focuses on the most relevant parts of a sentence according to a specific target or aspect.

BERT+RGAT (Wang et al., 2020) A hybrid model that combines the strengths of BERT and Relational Graph Attention Networks (RGAT) to enhance performance.

BERT$_{\text{Asp}}$+CEPT (Li et al., 2021) pre-trained on the large-scale sentiment aspect aware annotation data which retrieved from domain-specific corpora such as Yelp and Amazon reviews.

BERT$_{\text{Asp}}$+SCAPT (Li et al., 2021) uses a supervised contrastive learning loss, while BERTAsp+CEPT uses a cross-entropy loss. This is the only difference between them.

THOR (Fei et al., 2023) This approach stimulates performance based on CoT prompting with three-step generation, achieving better results after supervised fine-tuning of Flan-T5 11B.

ABSA-ESA (Ouyang et al., 2024) This method try to solve the implicit sentiment issue by generating explicit sentiment augmentations.

4.3 Implementation Details

Our experiments are based on OpenPromt (Ding et al., 2021), which is an open-source toolkit designed for conducting prompt learning. We use ChatGPT 4 to get more candidate words and use ROBERTA-Large (Liu et al., 2019) for the remaining experiments. RoBERTa-large is a transformer-based language model built on the BERT architecture and is regarded as suitable for natural language understanding. We consider the supervised fine-tuning, and for test metrics, we use Micro-F1 in all experiments. Our choice of hyperparameters is thus based on choices made in previous work and practical considerations. We use a learning rate of $1 \cdot 10^{-5}$, a batch size of 16 and a maximum sequence length of 128. we sample 10% instances of each class from the original training set to form the validation set. We tune the entire model for 10 epochs and choose the checkpoint with the best validation performance to test. Our experiments are conducted with a NVIDIA RTX 3090 GPU. All our experiments including ablation experiment, are based on the above rules.

4.4 Results on Supervised Fine-tuning

In this section, we present the experimental results to demonstrate the effectiveness of the proposed ISAPrompt method and ISAPrompt ensemble. Table 3 shows the macro F1 scores of our framework in comparison to the benchmark methods in standard supervised settings. We provide both the best result for ISAPrompt and the best result for the ISAPrompt ensemble on the test set. The results show that ISAPrompt lags only slightly behind the SOTA THOR in ISA performance, which uses larger-scale LLMs and multiple prompts for inference, 0.02% (=79.67-79.73) on restaurant and 0.21% (82.43-82.22) on Laptop. However, ISAPrompts ensemble, after combining the capabilities of multiple ISAPrompts, significantly boosts ISA performance. 3.48% (83.15-79.67) on Restaurant and 2.15% (84.37-82.22) on Laptop.

4.5 Results Analysis

The optimal performance of our method, ISAPrompt, was achieved by utilizing the third designed prompt. Through our improvements, ISAPrompt, using only a single prompt and a 355M-sized model, can achieve performance comparable to that of multiple prompts reasoning and models with a scale of 11B. Additionally, by applying the prompts ensemble trick, our method shows a significant lead in implicit sentiment analysis. Although the prompts ensemble utilizes multiple models, it still requires fewer computational resources, time, and other resources compared to the baseline method, THOR. This demonstrates the effectiveness and practicality of our method.

4.6 Ablation Study

In order to verify the effectiveness of ISAPrompt proposed in this paper, we performed an ablation study. As presented in Table 4, we consider the regular POFT method, which uses the class name

Table 3: F1 results on supervised fine-tuning setup. Best results are marked in bold. The results with † and * are obtained from Fei et al. (2023) and Ouyang et al. (2024).

Method	Restaurant		Laptop	
	All F1	ISA F1	All F1	ISA F1
BERT+SPC† (110M)	77.16	65.54	73.45	69.54
BERT+ADA† (110M)	80.05	65.92	74.18	71.04
BERT+RGAT† (110M)	81.35	67.79	74.07	72.99
BERT_{Asp}+CEPT † (110M)	82.07	67.79	78.38	75.86
BERT_{Asp}+SCAPT † (110M)	83.79	72.28	79.15	77.59
Flan-T5 + THOR† (11B)	87.45	79.73	**85.16**	82.43
Flan-T5 + ABSA-ESA* (250M)	83.79	73.76	81.78	77.91
ROBERTA-large+ISAPrompt (355M)	86.54	79.67	82.71	82.22
ISAPrompt ensemble	**88.13**	**83.15**	83.60	**84.37**

Table 4: Ablation Experimental Results.

Method	Restaurant		Laptop	
	All F1	ISA F1	All F1	ISA F1
POFT	83.58	72.03	81.57	80.05
ISAPrompt-V	85.92	79.13	81.52	**83.80**
ISAPrompt	**86.54**	**79.67**	**82.71**	82.22

$T_1(x, w) = w$ was $[mask]$. x

$T_2(x, w) = w$ just $[mask]$! x

$T_3(x, w) = x$ All in all, w was $[mask]$.

$T_4(x, w) = x$ In summary, the w is $[mask]$.

Figure 5: Regular Prompts.

as the only label word for each class, without focusing on prompts designed to extract hidden opinion words and nuances. This was used in (Schick and Schütze, 2020). In the ablation experiment on POFT, the construction of the prompts was modified based on their designed prompts, changing the sentiment analysis from sentence-level to target word-level to better suit the task. We conducted POFT on all of these prompts and provide the best result on the test set to compare our method. The modified regular prompts were shown in Figure 5. The best performance of our method ISAPrompt on both test sets is achieved by using the third prompt we designed. Therefore, we also compared the regular POFT verbalizer construction method on the third prompt, where the method uses the class name as the sole label word for each class. We will refer to it as ISAPrompt-V.

From the ablation experiment, we can see that our ISAPrompt is effective in implicit sentiment analysis tasks. Our designed prompts that effectively guide the model to focus on hidden opinions and compared with the regular POFT, it is a huge improvement in ISA performance, 7.64%= (79.67-72.03) on restaurant and 2.17%= (82.22-80.05). This proves that it is feasible to better obtain hidden views by designing prompts. In addition, our improvements to the verbalizer also helped a little. These findings emphasize the importance of focusing on hidden opinions to enhance the reliability and accuracy of implicit sentiment detection across various contexts.

5 Conclusion

In this paper, we introduce ISAPrompt, a straightforward and efficient method for POFT aimed at analysis implicit sentiment within text. By mimicking the speaker emphasize the hidden opinions within the texts, expands and advance the verbalizer in POFT, it is possible for relatively smaller-scale LLMs to rival larger-scale LLMs that use more advanced methods. In addition, by adopting the ISAPrompt ensemble, an even more effective ensemble model can be achieved. Empirical experiments demonstrate that our method achieves better performance in implicit sentiment analysis compared to the optimal baseline, while utilizing fewer resources.

6 Limitations

While our method has achieved state-of-the-art performance on several benchmark datasets for implicit sentiment analysis, there are still some limitations, particularly in the following areas.

Firstly, the verbalizer design in ISAPrompt is manually constructed, which requires a little human effort and time, and we plan to develop an automatic role-aligned verbalizer in the future. Moreover, although our constructed prompts have decent applicability, being fixed prompts, they will inevitably encounter scenarios where they are not well-suited. Tackling these issues will be crucial for improving ISAPrompt's flexibility, scalability, and applicability across a wide range of tasks and environments.

References

Xiaoyi Bao, Xiaotong Jiang, Zhongqing Wang, Yue Zhang, and Guodong Zhou. 2023. Opinion tree parsing for aspect-based sentiment analysis. *arXiv preprint arXiv:2306.08925*.

Tom B Brown. 2020. Language models are few-shot learners. *arXiv preprint arXiv:2005.14165*.

Xiang Chen, Ningyu Zhang, Xin Xie, Shumin Deng, Yunzhi Yao, Chuanqi Tan, Fei Huang, Luo Si, and Huajun Chen. 2022. Knowprompt: Knowledge-aware prompt-tuning with synergistic optimization for relation extraction. In *Proceedings of the ACM Web Conference 2022*, WWW '22. ACM.

Jingfeng Cui, Zhaoxia Wang, Seng-Beng Ho, and Erik Cambria. 2023. Survey on sentiment analysis: evolution of research methods and topics. *Artificial Intelligence Review*, 56(8):8469–8510.

Lingjia Deng and Janyce Wiebe. 2014. Sentiment propagation via implicature constraints. In *Proceedings of the 14th Conference of the European Chapter of the Association for Computational Linguistics*, pages 377–385, Gothenburg, Sweden. Association for Computational Linguistics.

Jacob Devlin. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.

Ning Ding, Shengding Hu, Weilin Zhao, Yulin Chen, Zhiyuan Liu, Hai-Tao Zheng, and Maosong Sun. 2021. Openprompt: An open-source framework for prompt-learning. *arXiv preprint arXiv:2111.01998*.

Hao Fei, Bobo Li, Qian Liu, Lidong Bing, Fei Li, and Tat-Seng Chua. 2023. Reasoning implicit sentiment with chain-of-thought prompting. *arXiv preprint arXiv:2305.11255*.

Tianyu Gao, Adam Fisch, and Danqi Chen. 2020. Making pre-trained language models better few-shot learners. *arXiv preprint arXiv:2012.15723*.

Yuxian Gu, Xu Han, Zhiyuan Liu, and Minlie Huang. 2021. Ppt: Pre-trained prompt tuning for few-shot learning. *arXiv preprint arXiv:2109.04332*.

Shengding Hu, Ning Ding, Huadong Wang, Zhiyuan Liu, Jingang Wang, Juanzi Li, Wei Wu, and Maosong Sun. 2021. Knowledgeable prompt-tuning: Incorporating knowledge into prompt verbalizer for text classification. *arXiv preprint arXiv:2108.02035*.

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.

Jian Liao, Suge Wang, and Deyu Li. 2019. Identification of fact-implied implicit sentiment based on multi-level semantic fused representation. *Knowledge-Based Systems*, 165:197–207.

Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2019. Roberta: A robustly optimized bert pretraining approach.

Yiyi Liu, Ruqing Zhang, Yixing Fan, Jiafeng Guo, and Xueqi Cheng. 2023. Prompt tuning with contradictory intentions for sarcasm recognition. In *Proceedings of the 17th Conference of the European Chapter of the Association for Computational Linguistics*, pages 328–339.

Dehong Ma, Sujian Li, Xiaodong Zhang, and Houfeng Wang. 2017. Interactive attention networks for aspect-level sentiment classification. *arXiv preprint arXiv:1709.00893*.

Jihong Ouyang, Zhiyao Yang, Silong Liang, Bing Wang, Yimeng Wang, and Ximing Li. 2024. Aspect-based sentiment analysis with explicit sentiment augmentations. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pages 18842–18850.

Bo Pang, Lillian Lee, et al. 2008. Opinion mining and sentiment analysis. *Foundations and Trends® in information retrieval*, 2(1–2):1–135.

Bhargavi Paranjape, Julian Michael, Marjan Ghazvininejad, Luke Zettlemoyer, and Hannaneh Hajishirzi. 2021. Prompting contrastive explanations for commonsense reasoning tasks. *arXiv preprint arXiv:2106.06823*.

Maria Pontiki, Dimitris Galanis, John Pavlopoulos, Harris Papageorgiou, Ion Androutsopoulos, and Suresh Manandhar. 2014. SemEval-2014 task 4: Aspect based sentiment analysis. In *Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014)*, pages 27–35, Dublin, Ireland. Association for Computational Linguistics.

Alexander Rietzler, Sebastian Stabinger, Paul Opitz, and Stefan Engl. 2019. Adapt or get left behind: Domain adaptation through bert language model finetuning for aspect-target sentiment classification. *arXiv preprint arXiv:1908.11860*.

Irene Russo, Tommaso Caselli, and Carlo Strapparava. 2015. SemEval-2015 task 9: CLIPEval implicit polarity of events. In *Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015)*, pages 443–450, Denver, Colorado. Association for Computational Linguistics.

Timo Schick and Hinrich Schütze. 2020. Exploiting cloze questions for few shot text classification and natural language inference. *arXiv preprint arXiv:2001.07676*.

Yongsik Seo, Sungwon Song, Ryang Heo, Jieyong Kim, and Dongha Lee. 2024. Make compound sentences simple to analyze: Learning to split sentences for aspect-based sentiment analysis.

Taylor Shin, Yasaman Razeghi, Robert L Logan IV, Eric Wallace, and Sameer Singh. 2020. Autoprompt: Eliciting knowledge from language models with automatically generated prompts. *arXiv preprint arXiv:2010.15980*.

Kai Wang, Weizhou Shen, Yunyi Yang, Xiaojun Quan, and Rui Wang. 2020. Relational graph attention network for aspect-based sentiment analysis. *arXiv preprint arXiv:2004.12362*.

Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Fei Xia, Ed Chi, Quoc V Le, Denny Zhou, et al. 2022. Chain-of-thought prompting elicits reasoning in large language models. *Advances in neural information processing systems*, 35:24824–24837.

Minghao Xu, Daling Wang, Shi Feng, Zhenfei Yang, and Yifei Zhang. 2022. Kc-isa: An implicit sentiment analysis model combining knowledge enhancement and context features. In *Proceedings of the 29th International Conference on Computational Linguistics*, pages 6906–6915.