Large Language Model is a Better Context Extractor for Aspect-Based Sentiment Analysis

Anonymous ACL submission

Abstract

Previous Aspect-Based Sentiment Analysis (ABSA) studies have often incorporated syntactic information to connect contextual details with the designated aspect. These methods rely on complex model design to obtain syntactic structure information, further acquiring crucial semantic insights. Considering the potent contextualization abilities of the Large Language Model (LLM), we present the Low-Rank Adaptation plus In-domain Dynamic Examplar (LoRA-IDE) framework. This framework effectively aligns the task and sentence context information with the target aspect, leveraging the power of LLM. Specifically, we employ the LoRA training strategy to enable LLM to learn the context information of ABSA and promote the model's understanding of the connection between sentence context and aspects through the use of curated, designed instructions with IDE. Experimental results demonstrate that our proposed approach not only improves the performance of LLM on ABSA but also outperforms the current state-of-the-art model on two benchmarks at a large scale. The codes will be released upon the acceptance of this paper.

1 Introduction

Aspect-based sentiment analysis (ABSA) is a fine-grained sentiment analysis that aims to extract detailed sentiment information regarding specific aspects (Pontiki et al., 2014). For example, for the sentence "The food is fresh and piping hot." and the aspect of interest, "food", the task is to detect the positive sentiment expressed towards "food."

In this task, the most challenging part is accurately recognizing the contextual information to the relevant aspects (Ma et al., 2023). Several studies investigated refining the dependency trees of the context (Chen et al., 2020; Zhou et al., 2021; Chen et al., 2022). Others tried to utilize Graph Neural Networks (GNNs) in conjunction with dependency trees to better exploit syntax information (Zhang and Qian, 2020; Wang et al., 2020; Tang et al., 2020; Xiao et al., 2021; Zhang et al., 2022). These models convert the syntactic dependency relation with the context into a graph representation, then encoded using a combination of attention or convolution mechanisms. This approach enables the models to effectively leverage syntactic information to obtain the relevant contextual details to specific aspect terms, leading to better performance.

Although the model can benefit from the syntactic information, it often requires complicated design (Ma et al., 2023). Additionally, incorporating syntactic information that relies on dependency parsers introduces inherent inaccuracies (Wang et al., 2020) and further leads to errors occurring in ABSA. This paper proposes employing the LLM in the ABSA to align aspect terms and contextual information directly. Meanwhile, LLM naturally excels at understanding context thanks to the massive amount of parameters and text training data (Touvron et al., 2023; Brown et al., 2020). However, using LLM directly in ABSA does not yield the optimal results (Liu et al., 2022). To overcome this limitation, we introduce a novel Low-Rank Adaptation with an In-domain Dynamic Examplar (LoRA-IDE) framework tailored to the LLM to extract context information for the ABSA task. More specifically, we adopt the LoRA method (Hu et al., 2021) to implement a parameter-efficient fine-tuning strategy on LLM, facilitating the model to learn the context information. During the tuning phase, we introduce the dynamic inclusion of in-domain examples.

This strategy optimizes the model's ability to align specific aspects with corresponding sentiments. The experimental results on commonly used ABSA datasets indicate that our method significantly surpassed previous methods and considerably boosted the performance of LLM in ABSA. We have made the following contributions:

- We propose a novel method that leverages the power of LLMs to align contextual information with targeted aspects without relying on syntactic information. This method avoids the possible errors caused by improper parsing or incorrect use of syntactic information.

- We introduce a model-agnostic approach that can be easily applied to any open-sourced LLM. This approach offers flexibility and can be seamlessly integrated into various LLM architectures.

- The extensive experiments demonstrate superior results compared to previous state-of-the-art methods. Specifically, we achieved an impressive improvement of 5.3% and 5.9% in the F1 score for the laptop and restaurant datasets (Pontiki et al., 2014), respectively.

2 Related Works

ABSA is the commonly used term in literature to describe sentiment analysis at the aspect level. The term “aspect” refers to the entities, persons, events, features, objects, or targets mentioned in a sentence that is relevant to the sentiment being expressed (Pang and Lee, 2008). To explore the sentiment information expressed in the context, earlier studies utilized features such as bag-of-words, part of speech, and word position (Saias, 2015; Wang et al., 2013), which is ineffective in capturing contextual information associated with specific aspects. Therefore, some studies have combined attention and memory networks into deep neural network (DNN) models, enabling the model to comprehend the interdependencies among words throughout a given sentence (Wang et al., 2016, 2017; Ma et al., 2017).

Simultaneously, other researchers in this field have predominantly focused on combining syntactic information to extract contextual cues. For instance, Zhou et al. (2021) proposes enhancing dependency trees through aspect-centric tree structure learning, while Chen et al. (2022) modifies syntactic distances based on aspect-to-context attention scores. Furthermore, several studies have chosen to incorporate dependency graphs into neural networks. These methods, such as Graph Attention Networks (GAT) (Wang et al., 2020) and Graph Convolutional Networks (GCN) (Zhang and Qian, 2020; Xiao et al., 2021; Zhang et al., 2022), effectively reduce the distance between aspects and their associated context, thereby alleviating the long-term dependency problem.

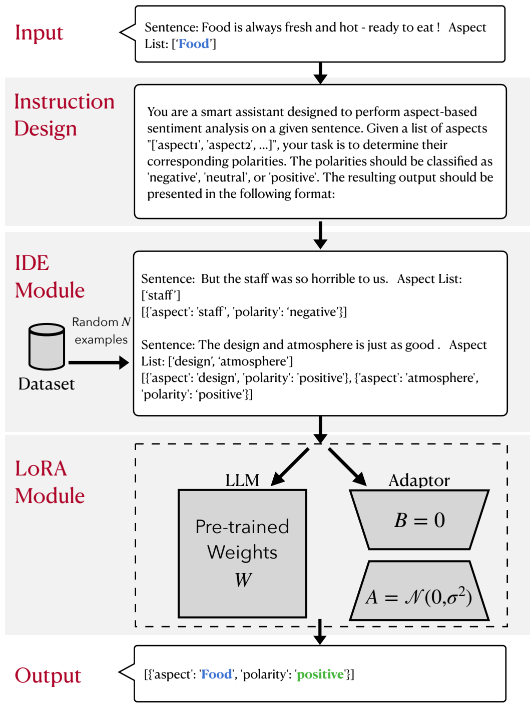


Figure 1: The construction workflow of LoRA-IDE framework.

3 Proposed Methodology

To align contextual information with targeted aspects using LLM, we propose an efficient framework for context extraction. This framework comprises two key components: the LoRA and IDE modules. The LoRA module facilitates the LLM learning context by leveraging adaptors, while the IDE module aids the LLM learning context from its in-context learning ability.

3.1 Problem Definition

The ABSA task seeks to identify the sentiment polarity $SP_i = sp_i^1, sp_i^2 ..., sp_i^m$, with $sp_i^m \in$ [positive, neutral, negative] towards sentence S_i and the given aspects $A_i = a_i^1, a_i^2 ..., a_i^m$, where m represents the number of aspect terms present in the sentence. In the context of LLM, the information from S_i and A_i is incorporated into the prompt P_i along with the task instruction I. Thus, the comprehensive formulation of the ABSA task using LLM can be represented as $[SP_i]$ = LLM($P_i(I, S_i, A_i)$). The primary objectives for LLM are twofold: Firstly, to establish a connection between SP_i and the contextual information regarding the task's purpose P_i. Secondly, to establish a link between LLM and the contextual information of S_i towards A_i.

3.2 LoRA Module

To bolster LLM's grasp of contextual information within the prompt, this study employs a parameter-efficient fine-tuning strategy called Low-Rank Adaptation of Large Language Models (LoRA) (Hu et al., 2021). Instead of retraining the entire model for ABSA, the LoRA methodology involves freezing the weights of LLM and the introduction of smaller trainable matrices into each layer of the Transformer architecture. Figure 1 illustrates the structure of the trainable rank decomposition matrices employed in LoRA. During the training process, the pre-trained weights denoted as W are held constant and do not undergo gradient updates. On the other hand, matrices A and B, characterized by trainable parameters, are subject to updates. As a result, the context information of the task could be learned and stored in these adaptors. Specifically, matrix $B \in \mathbb{R}^{d*r}$, and matrix $A \in \mathbb{R}^{r*k}$, where the dimensions of input and output are maintained. This process can be mathematically represented as the following equation:

$$h = W_0 x + \triangle W x = W_0 x + BAx \quad (1)$$

3.3 IDE Module

To tackle the challenge of acquiring context information about a specific aspect of a sentence, we propose a strategy incorporating an in-domain dynamic examplar technique, capitalizing on LLM's ability to learn from examples (Dong et al., 2023). Our strategy is motivated by two factors. Firstly, in-domain sentences tend to exhibit shared characteristics. Secondly, we aim to prevent the deterioration of LLM's understanding capability and avoid overfitting the data format. To achieve this, we introduce ABSA task instructions before each input. Additionally, we randomly select a dynamic number N of examples from both the in-domain training and development datasets for each input. These examples consist of pairs of input sentences and their corresponding targeted aspects, accompanied by the true polarity labels, as depicted in Figure 1.

4 Experiments

4.1 Datasets

We evaluate our work on three public standard ABSA datasets: Laptop and Restaurant datasets from Pontiki et al. (2014), and Twitter (Dong et al., 2014) dataset. To address the absence of official validation datasets, we randomly allocated 10% of the training set as the validation dataset. Please refer to Appendix A.1 for a detailed statistical breakdown of these datasets.

4.2 Setup

We selected Alpaca-7b (Taori et al., 2023) as the backbone LLM for our framework due to its open-source nature and its moderate performance in LLM. While Alpaca-7b can understand instructions, it does not possess the same level of advanced capabilities as ChatGPT (Ouyang et al., 2022). The Alpaca-7b employed in this study was sourced from the work of Yahma (2023). [1] We use the consistent instruction applied during training to prompt the Language Model (LLM) throughout the testing phase. Additionally, we evaluate the performance of the LLM in two distinct scenarios: zero-shot and few-shots. We conduct tests with shot values of 3, 5, and 8 for the few-shots evaluation, calculating their average performance (details are provided in Appendix A.3). We also test the performance of ChatGPT (GPT-3.5-turbo) by employing the API, [2] with the identical prompt. We adopt F1 score as our evaluation metric. Our experiments are conducted through one NVIDIA A-100 GPU. Additional information regarding hyperparameters can be found in Appendix A.2.

4.3 Results

We thoroughly compare our model with the state-of-the-art models and evaluate our model against a range of GNN-based models: (1) T-GCN (Tian et al., 2021), (2) DualGCN (Li et al., 2021), (3) dot-GCN (Chen et al., 2022), and (4) SSEGCN (Zhang et al., 2022). Additionally, we compare our model with dependency tree-based models, including (5) DGEDT (Tang et al., 2020) and (6)R-GAT (Wang et al., 2020). We also include two recently developed models: (7) TF-BERT (Zhang et al., 2023), which represents context information using sentiment intensities, and (8) APARN (Ma et al., 2023), which focuses on learning the semantic dependencies of the context. Finally, we conduct the same test with an advanced LLM (9) GPT-3.5-turbo for border comparison (Ouyang et al., 2022).

Table 1 showcases the experimental results of our model and the baseline models on the same benchmark. The results clearly demonstrate the

[1] https://huggingface.co/yahma/alpaca-7b-lora
[2] https://api.openai.com/v1/models

Model	Laptop	Restaurant	Twitter
T-GCN (Tian et al., 2021)	77.03	79.95	75.25
DualGCN (Li et al., 2021)	78.10	81.16	76.02
dotGCN (Chen et al., 2022)	78.10	80.49	77.00
SSEGCN (Zhang et al., 2022)	77.96	81.09	76.02
DGEDT (Tang et al., 2020)	75.60	80.00	75.40
R-GAT (Wang et al., 2020)	74.07	81.35	74.88
TF-BERT (Zhang et al., 2023)	78.46	81.15	<u>77.25</u>
APARN (Ma et al., 2023)	79.10	82.44	**78.79**
GPT-3.5-turbo$_{zero-shot}$ (Ouyang et al., 2022)	74.70	83.13	51.47
GPT-3.5-turbo$_{few-shots}$	77.68	84.66	60.17
LoRA-IDE$_{zero-shot}$	**83.27**	<u>87.34</u>	74.45
LoRA-IDE$_{few-shots}$	<u>82.94</u>	**87.88**	74.54

Table 1: The F1 score of the proposed model and previous baselines. The highest score is highlighted in bold font and the second highest score is underlined for clarity.

Model	Laptop	Restaurant	Twitter
• ***zero-shot***			
Base LLM	57.66	61.20	32.49
+LoRA	79.81	85.83	70.31
+LoRA+3ISE	81.97	86.84	70.90
+LoRA+5ISE	**83.56**	<u>87.00</u>	<u>70.12</u>
+LoRA+8ISE	82.59	86.33	73.41
+LoRA+IDE	<u>83.27</u>	**87.34**	**74.45**
• ***few-shots***			
Base LLM	67.37	74.03	51.03
+LoRA	76.44	83.02	71.00
+LoRA+3ISE	82.34	<u>87.58</u>	73.61
+LoRA+5ISE	<u>82.58</u>	87.33	73.41
+LoRA+8ISE	80.28	86.18	**75.43**
+LoRA+IDE	**82.94**	**87.88**	<u>74.54</u>

Table 2: F1 score of ablation results on our framework.

superiority of our LoRA-IDE framework over previous models that heavily rely on syntactic information. Remarkably, our framework even outperforms the previous state-of-the-art model (APARN) on two out of three datasets, exhibiting an increase in the F1 score by 5.3% and 5.9% under the zero-shot and 4.9% and 6.6% under the few-shots. Moreover, our model surpasses one of the most powerful existing LLMs, GPT-3.5-turbo, in all three domains, whether in zero-shot or few-shots circumstances. Notably, it achieves an impressive 44.65% increase in F1 score under the zero-shot and a 23.88% increase in F1 score under the few-shots, specifically in the Twitter domain. These findings provide strong evidence for the effectiveness of our framework in leveraging LLM to extract contextual information for the ABSA task.

4.4 Ablation Study

We performed an ablation study on all three datasets to assess the effectiveness of our framework on LLM. The results are presented in Table 2. The numbers accompanied by **ISE** indicate the usage of N in-domain static examples during training. As anticipated, the LLM incorporating both the LoRA and IDE modules exhibited superior performance across three domains and two testing environments (zero-shot and few-shot).

Based on the findings presented in Table 2, it becomes evident that each module plays a crucial role in enabling the LLM to extract contextual information for ABSA. The inclusion of the LoRA module results in a significant improvement in the model performance, demonstrating the validity of the adaptation process of LLM on ABSA to align it with the context. In particular, the F1 score improvement in the Twitter domain stands out prominently, with an impressive increase of 116.4% (zero-shot). This highlights the inherent limitations of context information stored in LLM and effectively showcases the remarkable capabilities of the LoRA module. Furthermore, the addition of the IDE module further enhances the performance of LLM. The F1 score increase ranges from 1.8% to 5.9% (zero-shot) and 3.5% to 6.5% (few-shots) across the three datasets, demonstrating the effectiveness of the IDE module in aligning LLM with the context of in-domain examples. Notably, the improvement is more pronounced under the few-shots setting, highlighting the IDE module's ability to strengthen the in-context learning capability of LLM. Finally, the gap in F1 scores between the static and dynamic examplar serves as compelling evidence for the effectiveness of our dynamic strategy in mitigating the degradation of the LLM's understanding capability caused by overfitting.

5 Conclusion

In this research, we introduce the LoRA-IDE framework on LLM as a means to leverage its power in extracting contextual information for the ABSA. Unlike previous studies that rely on syntactic information to connect the context with the target aspect, our approach utilizes the LoRA module to enable LLM to learn the ABSA task's context through adaptors and the IDE module to facilitate learning from the context of in-domain examples. The experimental results demonstrate significant improvements, surpassing the previous state-of-the-art baseline by 5.3% and 5.9% on two of three benchmark datasets. Each module of our proposed framework has been shown to be effective.

Limitations

The model's performance in the Twitter domain does not surpass previous state-of-the-art baselines. This can be attributed to two primary factors. First, a significant number of sentences on Twitter are incomplete and grammatically incorrect compared to the sentences in the other two datasets. Second, Twitter's context frequently includes buzzwords and the latest popular abbreviations. These factors hinder the LLM from effectively leveraging the semantic information stored within the text.

References

Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel Ziegler, Jeffrey Wu, Clemens Winter, Chris Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam McCandlish, Alec Radford, Ilya Sutskever, and Dario Amodei. 2020. Language models are few-shot learners. In *Advances in Neural Information Processing Systems*, volume 33, pages 1877–1901. Curran Associates, Inc.

Chenhua Chen, Zhiyang Teng, Zhongqing Wang, and Yue Zhang. 2022. Discrete opinion tree induction for aspect-based sentiment analysis. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 2051–2064, Dublin, Ireland. Association for Computational Linguistics.

Chenhua Chen, Zhiyang Teng, and Yue Zhang. 2020. Inducing target-specific latent structures for aspect sentiment classification. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 5596–5607, Online. Association for Computational Linguistics.

Li Dong, Furu Wei, Chuanqi Tan, Duyu Tang, Ming Zhou, and Ke Xu. 2014. Adaptive recursive neural network for target-dependent Twitter sentiment classification. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 49–54, Baltimore, Maryland. Association for Computational Linguistics.

Qingxiu Dong, Lei Li, Damai Dai, Ce Zheng, Zhiyong Wu, Baobao Chang, Xu Sun, Jingjing Xu, Lei Li, and Zhifang Sui. 2023. A survey on in-context learning.

Edward J. Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, and Weizhu Chen. 2021. Lora: Low-rank adaptation of large language models.

Ruifan Li, Hao Chen, Fangxiang Feng, Zhanyu Ma, Xiaojie Wang, and Eduard Hovy. 2021. Dual graph convolutional networks for aspect-based sentiment analysis. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 6319–6329, Online. Association for Computational Linguistics.

Haokun Liu, Derek Tam, Mohammed Muqeeth, Jay Mohta, Tenghao Huang, Mohit Bansal, and Colin Raffel. 2022. Few-shot parameter-efficient fine-tuning is better and cheaper than in-context learning.

Dehong Ma, Sujian Li, Xiaodong Zhang, and Houfeng Wang. 2017. Interactive attention networks for aspect-level sentiment classification. In *Proceedings of the 26th International Joint Conference on Artificial Intelligence*, IJCAI'17, page 4068–4074. AAAI Press.

Fukun Ma, Xuming Hu, Aiwei Liu, Yawen Yang, Shuang Li, Philip S. Yu, and Lijie Wen. 2023. AMR-based network for aspect-based sentiment analysis. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 322–337, Toronto, Canada. Association for Computational Linguistics.

Long Ouyang, Jeff Wu, Xu Jiang, Diogo Almeida, Carroll L. Wainwright, Pamela Mishkin, Chong Zhang, Sandhini Agarwal, Katarina Slama, Alex Ray, John Schulman, Jacob Hilton, Fraser Kelton, Luke Miller, Maddie Simens, Amanda Askell, Peter Welinder, Paul Christiano, Jan Leike, and Ryan Lowe. 2022. Training language models to follow instructions with human feedback.

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

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.

José Saias. 2015. Sentiue: Target and aspect based sentiment analysis in SemEval-2015 task 12. In *Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015)*, pages 767–771, Denver, Colorado. Association for Computational Linguistics.

Hao Tang, Donghong Ji, Chenliang Li, and Qiji Zhou. 2020. Dependency graph enhanced dual-transformer structure for aspect-based sentiment classification. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 6578–6588, Online. Association for Computational Linguistics.

Rohan Taori, Ishaan Gulrajani, Tianyi Zhang, Yann Dubois, Xuechen Li, Carlos Guestrin, Percy Liang, and Tatsunori B. Hashimoto. 2023. Alpaca: A strong, replicable instruction-following model. https://crfm.stanford.edu/2023/03/13/alpaca.html. Accessed: 2023-03-13.

Yuanhe Tian, Guimin Chen, and Yan Song. 2021. Aspect-based sentiment analysis with type-aware graph convolutional networks and layer ensemble. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 2910–2922, Online. Association for Computational Linguistics.

Hugo Touvron, Thibaut Lavril, Gautier Izacard, Xavier Martinet, Marie-Anne Lachaux, Timothée Lacroix, Baptiste Rozière, Naman Goyal, Eric Hambro, Faisal Azhar, Aurelien Rodriguez, Armand Joulin, Edouard Grave, and Guillaume Lample. 2023. Llama: Open and efficient foundation language models.

Kai Wang, Weizhou Shen, Yunyi Yang, Xiaojun Quan, and Rui Wang. 2020. Relational graph attention network for aspect-based sentiment analysis. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 3229–3238, Online. Association for Computational Linguistics.

Wei Wang, Hua Xu, and Xiaoqiu Huang. 2013. Implicit feature detection via a constrained topic model and SVM. In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*, pages 903–907, Seattle, Washington, USA. Association for Computational Linguistics.

Wenya Wang, Sinno Jialin Pan, Daniel Dahlmeier, and Xiaokui Xiao. 2017. Coupled multi-layer attentions for co-extraction of aspect and opinion terms. *Proceedings of the AAAI Conference on Artificial Intelligence*, 31(1).

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.

Zeguan Xiao, Jiarun Wu, Qingliang Chen, and Congjian Deng. 2021. BERT4GCN: Using BERT intermediate layers to augment GCN for aspect-based sentiment classification. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 9193–9200, Online and Punta Cana, Dominican Republic. Association for Computational Linguistics.

Yahma. 2023. Alpaca-7b-lora.

Mao Zhang, Yongxin Zhu, Zhen Liu, Zhimin Bao, Yunfei Wu, Xing Sun, and Linli Xu. 2023. Span-level aspect-based sentiment analysis via table filling. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 9273–9284, Toronto, Canada. Association for Computational Linguistics.

Mi Zhang and Tieyun Qian. 2020. Convolution over hierarchical syntactic and lexical graphs for aspect level sentiment analysis. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 3540–3549, Online. Association for Computational Linguistics.

Zheng Zhang, Zili Zhou, and Yanna Wang. 2022. SSEGCN: Syntactic and semantic enhanced graph convolutional network for aspect-based sentiment analysis. In *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 4916–4925, Seattle, United States. Association for Computational Linguistics.

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

A Appendix

A.1 Datasets

Dataset	Laptop			Restaurant			Twitter		
	#+	#0	#-	#+	#0	#-	#+	#0	#-
Train	976	455	851	2164	637	807	1507	3016	1528
Test	337	167	128	728	196	196	172	336	169

Table 3: Statistics of three ABSA datasets. The symbols #+, #0, and #- represent the quantities of positive, neutral, and negative sentiments, respectively.

A.2 Hyperparameters

In this study, a standardized set of hyperparameters was utilized across all experiments. To ensure consistency, the dynamic number N was randomly chosen from a range of [0, 8]. Additionally, the learning rate was set to a fixed value of 3e-4, while a warm-up period of 50 steps was incorporated. The training epoch is set at 20. The experimental setup utilizes a batch size of 16 and implements 4 gradient accumulation steps. Evaluation and save steps are uniformly set at 100. The rank r and α value of LoRA are both set as 16. Notably, the optimization strategy involves the application of the Adam optimizer coupled with weight decay.

A.3 Details of Few-Shots Cases

To evaluate the performance of our model under few-shots prompting scenarios, we conducted experiments with shot values of 3, 5, and 8. These shot values represent the number of examples we included in the prompt during the inference stage. The series of figures below illustrate the F1 scores achieved by different models under these few-shots conditions.

The results indicate that our proposed framework is able to boost the LLM's performance on the ABSA task across all three domains. Furthermore, increasing the number of illustrated examples included in the prompt led to a further improvement in the model's performance.

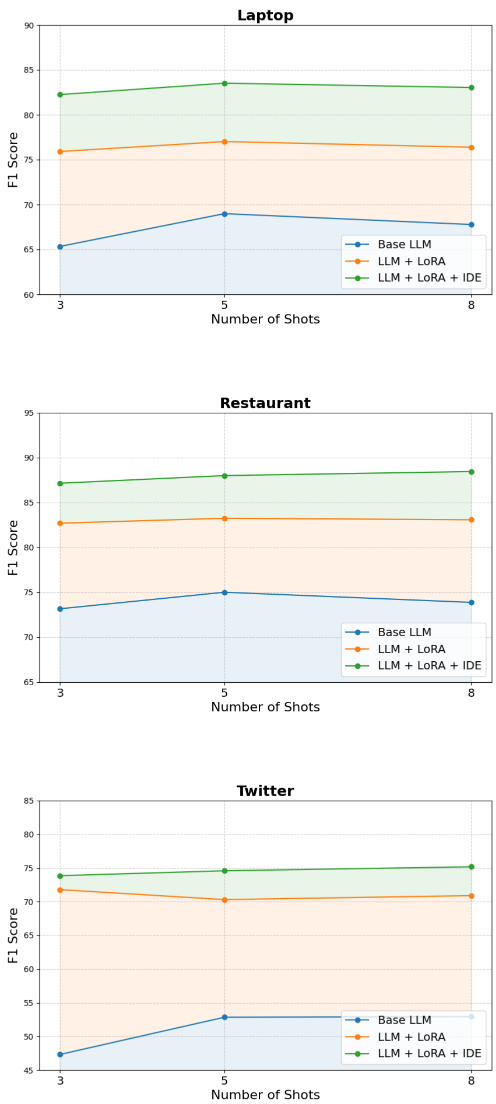
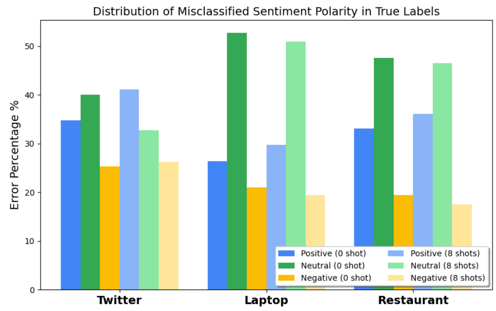


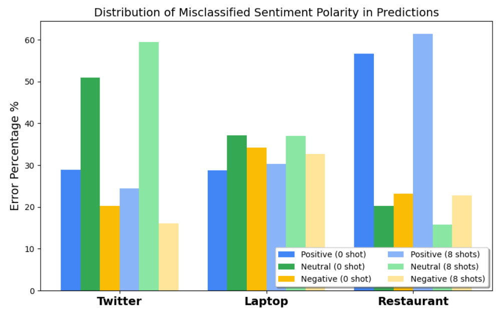
Figure 2: F1 scores across different models under few-shots inference configuration

A.4 Error Analysis

To inform the future research on the utilization of LLMs in ABSA, we conduct an error analysis of our suggested framework. We select the test outcomes from the 0-shot and 8-shots scenarios within the few-shot cases, followed by an analysis of the sentiment polarity distribution within these identified error cases.



(a) Misclassified Sentiment Polarity in True Labels



(b) Misclassified Sentiment Polarity in Predictions

Figure 3: Distribution of Sentiment Polarity in Error Cases

The misclassification of sentiment polarity in the true labels of the test datasets, as illustrated in Figure 3a, indicates that the "neutral" polarity is the most commonly misclassified across all cases. This implies that the “neutral” sentiment poses the most significant challenge, thus enhancing its likelihood of being misclassified as other sentiments.

In Figure 3b, we observe a distinct pattern in the distribution of misclassified sentiment polarity in predictions. Within the Twitter domain, the majority of sentiments are erroneously classified as “neutral”. This stands in stark contrast to the Laptop domain, where the misclassified sentiment polarity is uniformly distributed. In the Restaurant domain, the sentiment polarity of the corresponding aspects is predominantly misclassified as “positive”. This distribution is strikingly similar to the sentiment polarity distribution of the training datasets. The Twitter training dataset contains the highest number of “neutral” sentiment polarities, which is double the amount of the other two sentiments. The Laptop dataset displays a balanced distribution of sentiment polarity, while the Restaurant dataset comprises the highest number of “positive” sentiments, approximately triple the amount of the other two sentiments. This implies that the performance of our framework could be affected during the tuning phase by the skewed distribution of the training dataset. This further highlights the importance of high-quality training datasets for the optimal performance of the framework based on LLMs.