Leveraging Large Language Models for Aspect-Based Sentiment Classification Using GPT-4

Sami Alanazi[1] and Xiuwen Liu[2]

[1,2] Department of Computer Science, Florida State University, Tallahassee

ABSTRACT

Aspect-based sentiment analysis (ABSA) is an essential task in natural language processing (NLP) that aims to extract both aspects and their corresponding sentiment polarities from textual data. Despite its importance, ABSA faces challenges in accurately capturing nuanced sentiments, particularly in complex and context-dependent scenarios. This paper introduces a novel technique that employs few-shot learning with the GPT-4 model to determine the sentiment of specific aspects within the text. By using a template prompt along with few-shot examples, we enhance ChatGPT's in-context learning capability, thereby improving the effectiveness of aspect-based sentiment classification. Our tests and evaluations demonstrate that this method achieves state-of-the-art results in ABSA. This research represents a significant advancement in sentiment analysis methodologies, especially in capturing nuanced sentiments toward specific aspects.

KEYWORDS

Sentiment Analysis, Aspect Based Sentiment Analysis, Natural Language Processing, ChatGPT.

1 INTRODUCTION

Aspect-based sentiment analysis (ABSA) is an essential task in natural language processing that seeks to extract nuanced sentiments related to various aspects within the same content. As part of fine-grained opinion mining, ABSA focuses on analyzing opinions at a more granular level in order to provide a more accurate representation of the underlying sentiment [14].One of the main sub-tasks of ABSA is aspect-based sentiment classification (SC), which seeks to determine the sentiment polarity of specific aspects referenced within a sentence. With the increasing demand for more in-depth analysis of opinions expressed in various domains, SC has recently gained significant attention, and multiple research studies have been conducted to develop aspect-specific feature induction methods that improve its accuracy, such as encoder and decoder models [40],[8]. To illustrate the practical application of SC, consider the sentence "The price is reasonable although the service is poor", as shown in Figure 1. In this instance, SC requires a model to assign a positive sentiment to the aspect word "price ". The underlined word in Figure 1 represents the explicit expression of the sentiment relevant to the specified aspect. This example demonstrates SC's fundamental role within ABSA, emphasizing its significance in predicting sentiment polarities associated with designated aspects in diverse content.

DOI : 10.5121/ijait.2024.14402

ChatGPT[1] has recently become a popular choice among researchers and businesses since its launch in November 2022. As a large language model, it has gained considerable attention for its capacity to produce coherent and human-like text across diverse topics. A key factor in its success is the implementation of reinforcement learning from human feedback (RLHF) methodologies[20], which helps align its responses with human preferences. This AI-powered language model excels in engaging in smooth conversations, generating code and poetry, solving complex mathematical problems [35], and demonstrating a variety of emerging capabilities [6]. These features have attracted significant public interest.

Although ChatGPT has achieved significant success, it is essential to explore its capability boundaries to understand where it excels and falls short. This exploration is crucial for further development. The question arises as to whether ChatGPT can enhance ABSA by providing accurate insight into sentiment towards specific aspects and improving the performance of ABSC models.

To address this question, this paper proposes a novel technique to enhance ChatGPT's efficacy in aspect-based sentiment classification. Our approach aims to improve the model's ability to accurately capture nuanced sentiment variations associated with specific aspects within the text. The objective is to enable ChatGPT to not only understand the overall sentiment of a piece of text but also discern and appropriately respond to sentiments associated with individual aspects. We conducted extensive experiments on the SemEval 2014, 2015, and 2016 datasets by [21, 22, 23] for ABSC task, which includes the laptops and restaurants domains. Our methods outperformed state-of-the-art (SOTA) approaches in both domains. Specifically, our approach achieved an accuracy of 88.66% for Laptop14, 87.19% for Restaurant14, 93.56% for Restaurant15, and 94.51% for Restaurant16.

Our contributions are as follows:

- As part of our research on prompt engineering, we have developed a specialized prompt template specifically designed for aspect-based sentiment classification (ABSC). This template enhances the model's ability to accurately identify relevant aspects within a given context, thereby improving its effectiveness in ABSC.

- Our technique involves the strategic selection of limited training data for in-context learning. By employing this method, we achieve a higher performance compared to existing techniques in aspect-based sentiment classification.

- Our method is adept at identifying both explicit and implicit aspects. For explicit aspects, we utilize precise categorization, while for implicit aspects, we introduce a novel labeling strategy by designating them as "noaspect term". This dual approach ensures comprehensive aspect identification in sentiment analysis.

This paper is organized as follows: Section 2 reviews related work, outlining existing methodologies and their limitations in aspect-based sentiment analysis. Section 3 details our technique's methodology, including prompt learning with GPT-4 and data preprocessing. Section 4 describes the experimental setup and validation datasets. In Section 5, we compare our model's results with state-of-the-art baselines. Section 6 concludes with a summary of our findings and explores potential future research directions.

[1] OpenAI
https://platform.openai.com/docs/models/gpt-4.

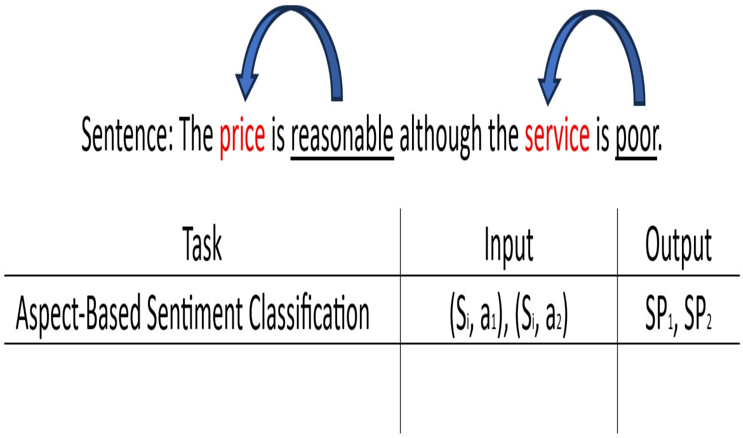

Fig. 1. Illustrates a sample of sentence. The aspects are indicated by red highlights, while the corresponding opinions are represented by underlining words.

2 RELATED WORK

The ABSA field has gained significant attention in fine-grained opinion mining. ABSA involves identifying the sentiment polarity associated with each specific aspect within sentences. One of the biggest challenges in ABSC is capturing sentiment changes based on the surrounding context. Another challenge is recognizing aspects explicitly or implicitly mentioned in the text. Finally, resolving ambiguity when a term refers to multiple aspects is also a major challenge. Various approaches have been proposed to enhance the efficacy of ABSA, such as memory networks [29, 32], convolutional networks [13], attentional networks [33, 16], and graph-based networks [39, 30]. These techniques have been extensively explored in recent years. The latest advancements in ABSA performance include using Pre-trained Language Models (PLMs) such as BERT [3] and RoBERTa [15] that employ context-encoder schemes. These models, as reported in studies by [12, 37, 28],and [38], have achieved state-of-the-art results.

However, context modeling using PLMs may lack aspect-specificity. To address this, various approaches have been suggested. One approach is to use aspect-specific input transformations, such as aspect companion, aspect prompt, and aspect marker, to promote PLMs to pay more attention to an aspect-specific context [17]. Another approach is to use sentiment knowledge-enhanced prompts, like SentiPrompt to tune the language model in a unified framework, explicitly modeling term relations and achieving better performance [10]. These methods have shown effectiveness and robustness, achieving state-of-the-art performance on sentiment polarity prediction and opinion span extraction tasks in ABSA [26].

Prompt engineering, a recent technique in natural language processing, involves carefully designing high-quality task prompts. It serves as an interface for users to interact with language learning models (LLMs) such as decoder-only GPT-3 [1] and encoder-decoder T5 [24]. [25] uses an instruction learning paradigm for the ABSA. ChatGPT, a recently introduced decoder-only LLM, has gained significant attention due to its remarkable ability to generate coherent responses across diverse NLP tasks, including translation [7], information extraction [36], and sentiment analysis [34].

Recent studies have explored the potential of ChatGPT in enhancing sentiment analysis. Fatouros et al. [5] find that ChatGPT outperformed FinBERT in financial sentiment analysis. Similarly, Erfina and Nurul and Korkmaz et al. in [4] and [9] reported positive user sentiments towards ChatGPT, with the majority expressing satisfaction. These findings suggest that ChatGPT can be a valuable tool in sentiment analysis. However, it still has room for further improvements in aspect-based sentiment analysis.

Our approach addresses ABSC by utilizing prompts and ChatGPT to capture aspect-specific context specifically. We initiate the process by distinguishing between explicit and implicit aspects and then apply our methods to analyze the input dataset. By leveraging our approach, we aim to achieve better performance in sentiment polarity prediction tasks in ABSA. Furthermore, our approach can potentially improve aspect-based sentiment classification (ABSC) models, allowing few-shot learning scenarios.

3 METHODOLOGY

This section provides an overview of the Aspect-Based Sentiment Classification in Subsection 3.1. Then it elaborates on a novel method to augment the capabilities of ChatGPT for ABSA. The improvement is rooted in the application of few-shot learning, which is achieved by embedding training data within the template of the prompt, thus promoting in-context learning. This strategy and its implications are further expounded upon in Subsection 3.2

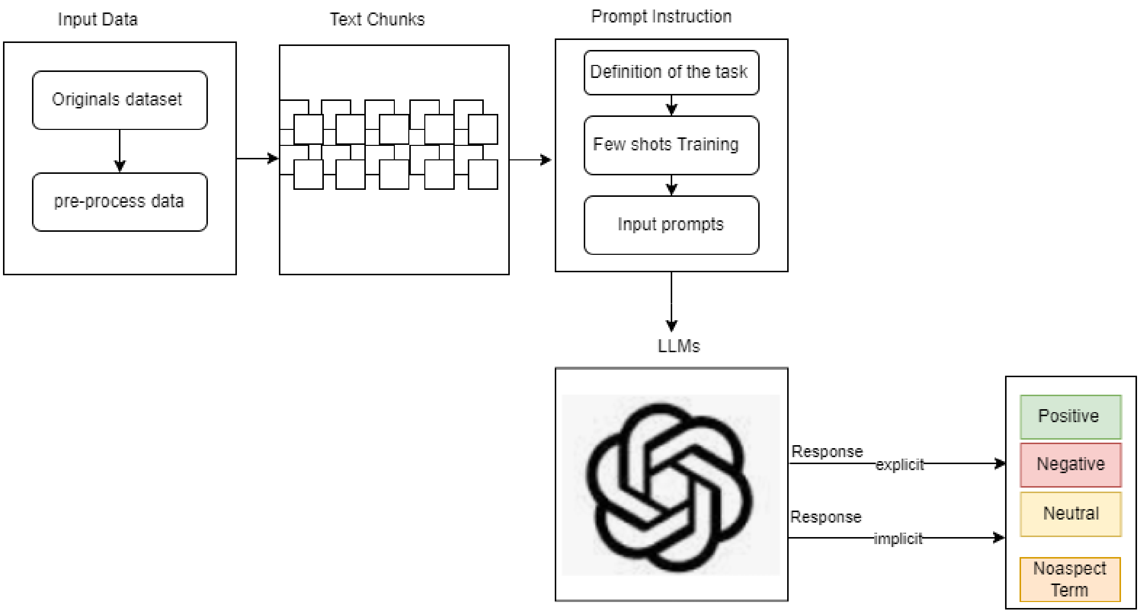

Fig. 2. The overall architecture of the method.

3.1 Aspect-based sentiment Classification

In our study, we explore ABSC, a specific aspect of sentiment analysis that aims to identify the emotional tones or sentiments linked to particular aspect terms in a text, like a review sentence. The primary objective of ABSC is to determine the sentiment polarities connected to each aspect term in a sentence.

To clarify, let's take a review sentence denoted as S_i. Within this sentence, there exist n distinct aspect terms, each possibly conveying a distinct sentiment polarity. These polarities are denoted as $SP_i = \{sp_{1i}, sp_{2i}, \ldots, sp_{ni}\}$, where each sp_{ki} indicates the sentiment polarity associated with the k^{th} aspect term ak_i in the sentence S_i. This procedure converts the initial sentence into several instances, each illustrating a different aspect term along with its respective sentiment polarity.

Our model's approach is designed to manage this complexity. For every sentence S_i and its related aspect term ak_i, the model utilizes a function, represented as $ABSC(S_i, ak_i)$, to process and ascertain the sentiment polarity sp_{ki} linked to that particular aspect term. This procedure is iterated for each aspect term in the sentence, culminating in a thorough analysis where each aspect term is paired with its corresponding sentiment polarity.

Through this systematic method, the model can thoroughly examine each sentence, evaluating and categorizing the sentiments associated with specific aspect terms, thus offering a nuanced comprehension of the sentiments conveyed in the text. This strategy is crucial for precisely assessing sentiment in situations where multiple aspects are deliberated, each potentially evoking distinct emotional reactions.

3.2 Proposed Methods

In this section, we explain in detail how our aspect-based sentiment classification model works. Firstly, we implement prompt learning by partnering with GPT-4. Subsequently, we generate prompts from the data. Following this, we utilize the model via an API key to categorize the specific aspect within the provided context in chunks. In addition, we present the overall architecture of this model in Figure 2, with a detailed description to follow in the subsequent sections.

Preprocessing Our initial step involves the utilization of the original dataset, as outlined in Table 1. Following this, our attention shifted towards the meticulous cleaning and preparation of the data, aiming to create structured inputs suitable for training and testing datasets.

In the process, we restructured the original dataset, organizing the input format based on individual sentences and their corresponding aspect terms requiring polarity classification. This restructuring was implemented to enhance the clarity and effectiveness of the input structure for our analytical purposes. As depicted in Figure 3, our approach involves utilizing sentence S_i and the associated aspect to predict the sentiment polarity for each aspect term.

Consequently, we incorporate sentence S_i and integrate the prompt, wherein the aspect term is denoted as *aspect*. This ensures that our analysis is conducted with a well-prepared and systematically organized dataset, facilitating accurate sentiment polarity predictions for specific aspect terms.

Prompt Template Certainly, ChatGPT demonstrates remarkable proficiency across various zero-shot and few-shot tasks. However, its performance in task-specific contexts heavily depends on the quality of prompts used for guidance, as highlighted by [42]. Consequently, a specialized prompt template is crafted specifically for the purpose of conducting sentiment classification for a particular aspect. This task is one of the aspect-based sentiment analysis.

Initially, we established the task and provided essential definitions and instructions about ABSC. Due to the need to classify specific aspects from the context and output format, we utilized a few examples from the training data. As a result, the input data will

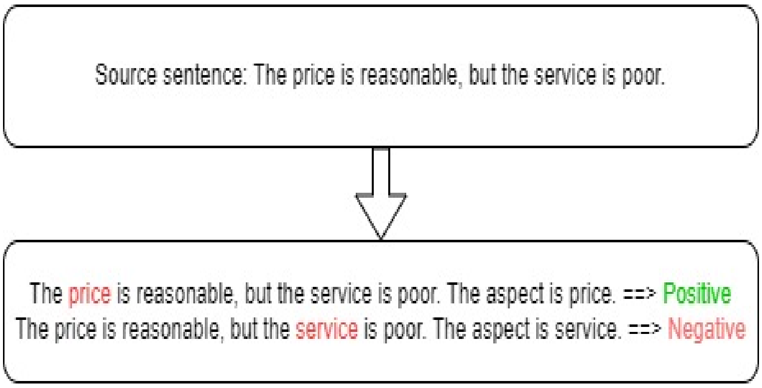

Fig. 3. Example of data preprocessing for aspect-based sentiment analysis.

be structured in a way that follows the format of the training samples, as shown in Figure 4.

Few-shot Learning In order for ChatGPT to learn within a few-shot scenarios, it needs examples to learn from. However, due to the limited space for prompts in ChatGPT, it is not possible to provide all examples from the training data at once. A significant question is whether we must select a few examples from the training data for each test example. Recent research [19] suggests that accurate demonstrations are not always necessary, and even randomly replacing labels in the demonstrations has a minimal impact on performance across tasks. Our experiments support this observation as shown in Figure 5. Therefore, this study does not consider label balance, such as sentiment polarity. We only use the first 10 reviews from the training dataset for each domain.

4 EXPERIMENTS

In this section, we validate the effectiveness of our approach through experiments conducted on four widely used public datasets. We also provide a detailed description of the experimental setup for our model, offering valuable insights for future research. Additionally, we compare our results with state-of-the-art baseline models.

4.1 Datasets

We gather four publicly available datasets to evaluate the effectiveness of ChatGPT compared to specialized methods for aspect-based sentiment classification. These datasets serve as a benchmark for ABSA tasks and comprise customer reviews from two domains, namely laptops (Lapt14) and restaurants (Rest14, Rest15, and Rest16) [21, 22, 23]. Table 1 displays the statistics of these four datasets. Each sample in these datasets consists of three parts: a single sentence that describes the product review, a particular aspect extracted from the review, and its corresponding sentiment polarity. It is worth noting that each sample in the datasets has only one sentiment polarity, which can be either positive, negative, or neutral.

4.2 Experimental Setup

To evaluate the ABSC task performance of the model, we use the API key to send a prompt message to the gpt-4-0613, a highly efficient and cost-effective GPT-4 model. The API's

According to the following sentiment elements definition:

- The 'sentiment polarity' refers to the degree of positivity, negativity, or neutrality expressed in the opinion towards a particular aspect or feature of a product or service, and the available polarities include: 'positive', 'negative', and 'neutral'.

- If the aspect is noaspectTerm, the sentiment polarity refers to none.

I would like to give you examples as training for you with the following format:
raw_text;aspectTerms where aspectTerms has polarity

Please note that one raw_text can have many aspect terms.

The following are the training data:

{shots_output}

Now, based on what you learned, please define the polarity of the aspect with the same format as the training data.

{raw_texts}

Fig. 4. The prompt template.

temperature is set at 0, yielding mostly deterministic outputs, with the other parameters kept at their default values.

To ensure a thorough assessment of Model's ABSC task performance, we start with instructions following the provided guidelines. However, due to token limit constraints in OpenAI's processing, we divide the procedure into multiple parts. Each part includes 100 test data instances, ten from the training set, and the prompt's instruction. We repeat this process for all input datasets. Subsequently, we consolidate all parts into one file and evaluate them against the ground truth label. This method enables us to effectively assess ChatGPT's ABSC task performance.

Baseline Methods To thoroughly evaluate the effectiveness of the method proposed in this paper, we have compared our model with the state-of-the-art baseline models. Here are the selected models with a brief introduction to each.

Table 1. Statistics of Datasets

Dataset	**Positive**		**Negative**		**Neutral**	
	Train	**Test**	**Train**	**Test**	**Train**	**Test**
Laptop14	994	341	870	128	464	169
Restaurant14	2164	728	807	196	637	196
Restaurant15	912	326	256	182	36	34
Restaurant16	1240	469	439	117	69	30

- In 2020, Wang et al. [31] introduced the Relational Graph Attention Network (R-GAT), which is designed to encode a novel tree structure extracted from a conventional dependency parse tree. This network enhances the representation of the new tree for improved information processing in context.
- In 2021, Li et al. [11] In their paper, they present DualGCN, a dual graph convolutional network that combines the strengths of both syntax structures and semantic correlations for effective integration.
- In 2022, Zhang et al. [41] have proposed a neural network model called dotGCN, which is syntax-based and aims to identify the relevant contextual information in relation to the aspect term.
- In 2023, Cao et al. [2] introduced HRLN-RoBERTa, a model that aims to address the challenge of confirming the mapping between the aspect and the core context. This model comprises a heterogeneous network module and a reinforcement learning module based on a knowledge graph, which work together to improve the accuracy of aspect-based sentiment analysis.
- In 2021, Mao et al. [18] define two machine reading comprehension (MRC) problems and tackle all related subtasks by simultaneously training two BERT-MRC models with shared parameters.
- In 2023, Shi et al. [27] proposed PRoGCN that combines RoBERTa with a Graph Convolution Network by inserting task-specific prompting words into raw text to guide the model. This prompted feature representation aids in generating a textual knowledge graph, enhancing both syntactic and semantic features.
- in 2023, Scaria et al. [25] added positive, negative, and neutral examples to each training sample, and utilized a Tk-Instruct model to tune the model for ABSA subtasks.

Evaluation Metrics We employed accuracy metrics to assess Aspect-Based Sentiment Classification (ABSC). In presenting our findings, we report the results of various compared methods as documented in the original studies [25, 27].

5 RESULTS

This section presents the results of the baseline methods applied to the aspect-based sentiment classification datasets (Laptop14, Restaurant14, Restaurant15, and Restaurant16), as shown in Table 2. Additionally, we compare these results with the performance of our proposed model.

In evaluating aspect-based sentiment classification, our proposed model consistently outperforms baseline methods across all datasets. The main results highlight the effectiveness of our model, with remarkable accuracy percentages: 88.66% for Laptop14, 87.19% for Restaurant14, 93.56% for Restaurant15, and an outstanding 94.51% for Restaurant16.

The improvements observed in our model's performance for aspect-based sentiment classification can be attributed to several advancements. First, our preprocessing of the datasets ensures that the input data is of high quality, noise-free, and appropriately formatted. This step enhances the model's ability to accurately detect and categorize sentiments associated with specific aspects.

Second, by leveraging few-shot learning techniques, our model benefits from improved in-context learning capabilities. Embedding relevant training examples within the prompt allows GPT-4 to better understand the task requirements and generalize from limited data, leading to better performance.

Additionally, the use of carefully crafted input prompts tailored to the aspect-based sentiment classification task guides the model to focus on the relevant aspects and sentiments. This approach increases the accuracy and reliability of the model's responses. Clearly defining the task and conducting targeted few-shot training sessions enable the model to capture nuanced sentiment variations associated with different aspects. This training improves the model's ability to discern and respond to specific sentiments effectively.

Finally, extensive experiments conducted on the SemEval 2014, 2015, and 2016 datasets provide a robust evaluation framework. The consistent performance improvements across multiple datasets, including Laptop14, Restaurant14, Restaurant15, and Restaurant16, demonstrate the model's generalizability and robustness. These advancements collectively contribute to the performance gains of our proposed model, as evidenced by its higher accuracy rates compared to state-of-the-art baseline methods.

5.1 Few-shot Results

In this section, we analyze a few-shot learning results in the ABSC context using the "Restaurants 14" dataset and discuss GPT-3.5[2] and GPT-4's[3] observed performance based on the number of examples available.

Our few-shot investigation of the GPT-3.5 and GPT-4 models give interesting results, as shown in Figure 5. Analysis was done using the first 100 examples "Restaurants 14" dataset to look at how well the model can do ABSC when trained with a different number of training examples.

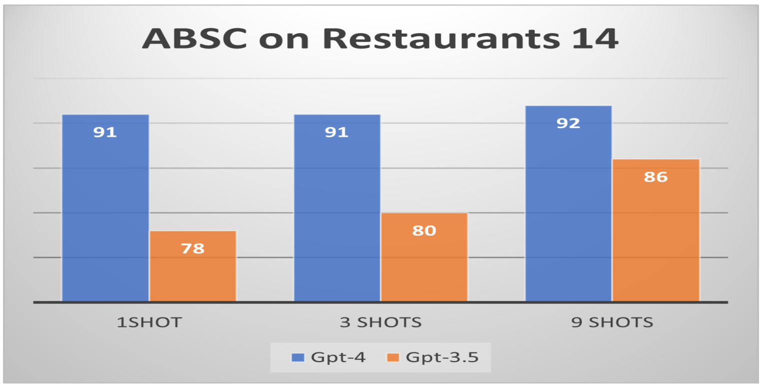

Fig. 5. Few-shot example results on ABSC on 100 input data.

One-Shot Learning Performance of GPT-4 in one-shot learning is exceptionally strong, indicating an advanced initial grasp of tasks, potentially due to more sophisticated

[2] gpt-3.5-turbo-1106
[3] gpt-4-0613

pre-training or improved generalization capabilities. This is in contrast to previous models like GPT-3.5, which do not exhibit similar out-of-the-box efficiency.

Three-Shots Learning In a three-shot learning scenario, GPT-3.5 shows significant improvement over its one-shot performance, but still lags behind GPT-4. This suggests that GPT-3.5 may require more examples to effectively adapt to new tasks, reaching near the competency of GPT-4 with additional context.

Nine-Shots Learning With nine examples, both models demonstrate improved performance compared to one-shot learning. However, GPT-4 maintains a consistent level of performance from three to nine shots, whereas GPT-3.5 shows only a minor gain. This indicates that while GPT-3.5 benefits from more examples, its learning curve is less steep compared to GPT-4.

Consistency Across Few-Shot Scenarios: Across all few-shot scenarios (from one to nine shots), our methodology exhibited consistent performance for each model. This consistency is a testament to the robustness of the approach and suggests that the models' performance is stable across a range of few-shot learning contexts.

5.2 Comparative Results

In this section, we delve into the comparative analysis of various state-of-the-art models, benchmarking their performances across four public datasets. Each model under consideration has been rigorously evaluated for its proficiency in discerning sentiment or extracting pertinent information within specific domains. Notably, our proposed model distinguishes itself by demonstrating exceptional effectiveness in all datasets examined.

Table 2. Comparing the results of experiments across four benchmark datasets on ABSC.

Model	Laptop14	Restaurant14	Restaurant15	Restaurant16	Average
R-GAT[a]	78.21	86.60	-	-	82.41
DaulGCN[a]	78.48	84.27	-	-	81.38
dotGCN[a]	81.03	86.16	85.24	93.18	86.65
HRLN-RoBERTa[a]	66.2	66.9	60.5	76.5	67.8
BERT-MRC[a]	75.97	82.04	73.59	-	77.2
InstructABSA[b]	81.56	85.17	84.50	89.43	85.41
PRoGCN[b]	81.82	**87.32**	86.53	93.02	87.42
Our model[b]	**88.66**	87.19	**93.56**	**94.51**	**91.23**

[a]Supervised Learning.
[b]Few-shot Learning.

Overall Performance Improvement:
Across all datasets, our proposed model demonstrates a consistent improvement in aspect-based sentiment classification. The average performance across the four benchmark datasets is 91.23%, indicating the robustness and efficacy of our approach compared to the baseline methods.

Domain-specific Achievements:
In the aspect-based sentiment classification task on the Laptop14 dataset, our model achieves an accuracy of 88.66%, outperforming dotGCN and PRoGCN by 7.63% and 6.84%, respectively.

For the Restaurant14 dataset, our model achieves an accuracy of 87.19%, outperforming dotGCN by 1.03% and closely matching the PRoGCN performance. On the Restaurant15 dataset, our model achieves an accuracy of 93.56%, surpassing PRoGCN by 7.03%.

In the evaluation of the Restaurant16 dataset, our model demonstrates the performance with an accuracy of 94.51%, outperforming dotGCN by 1.33% and PRoGCN by 1.49%. The consistent performance of our model across these distinct domains highlights its robustness and effectiveness in aspect-based sentiment classification.

In this study, we observed differences between few-shot learning and supervised learning methods in aspect-based sentiment analysis (ABSA). Few-shot learning with ChatGPT, particularly using a template prompt integrated with examples, showed better performance compared to traditional supervised learning. This improvement is attributed to ChatGPT's enhanced in-context learning ability, allowing it to more accurately classify sentiments related to specific aspects in texts. The method developed in this study leverages the adaptive learning capabilities of ChatGPT, making it more efficient in capturing nuanced sentiments without the extensive labeled data typically required in supervised learning.

6 CONCLUSION

This paper introduces a method to improve the performance of aspect-based sentiment classification by leveraging prompt engineering and advanced linguistic feature analysis.Combining these techniques enables the accurate categorization and interpretation of complex sentiments associated with various features within a text, resulting in improved classification accuracy.

The developed methodology achieves state-of-the-art results, demonstrating the feasibility and effectiveness of combining prompt engineering with sophisticated natural language processing models. The findings highlight potential improvements in the accuracy of sentiment analysis and underscore the importance of continuous improvements and adaptations in analytical methodologies to address evolving linguistic nuances and complexities.

Our study significantly contributes to the field of sentiment analysis by highlighting the importance of ongoing refinement in sentiment analysis techniques. The progress outlined in this study sets the stage for future research and development, fostering the continual advancement of sentiment analysis methodologies to better understand and interpret the complexities of human language.

Future research in aspect-based sentiment classification aims to enhance the field by focusing on several key areas. First, integrating context-aware models will improve the detection of nuanced sentiments across various domains by considering the broader context of the text. Second, developing sophisticated techniques to identify implicit aspects and their associated sentiments will tackle one of the more challenging tasks in sentiment analysis. Finally, expanding datasets to encompass more diverse and multilingual data will enable thorough evaluation and enhance the applicability of our methods across different languages and cultures.

References

[1] Tom Brown et al. "Language models are few-shot learners". In: *Advances in neural information processing systems* 33 (2020), pp. 1877–1901.

[2] Yukun Cao et al. "Heterogeneous Reinforcement Learning Network for Aspect-based Sentiment Classification with External Knowledge". In: *IEEE Transactions on Affective Computing* (2023).

[3] Jacob Devlin et al. "Bert: Pre-training of deep bidirectional transformers for language understanding". In: *arXiv preprint arXiv:1810.04805* (2018).

[4] Adhitia Erfina and Muhamad Rifki Nurul. "Implementation of Naive Bayes classification algorithm for Twitter user sentiment analysis on ChatGPT using Python programming language". In: *Data & Metadata* 2 (2023), pp. 45–45.

[5] Georgios Fatouros et al. "Transforming sentiment analysis in the financial domain with ChatGPT". In: *Machine Learning with Applications* 14 (2023), p. 100508.

[6] Simon Frieder et al. "Mathematical capabilities of chatgpt". In: *arXiv preprint arXiv:2301.13867* (2023).

[7] Yuan Gao, Ruili Wang, and Feng Hou. "Unleashing the Power of ChatGPT for Translation: An Empirical Study". In: *arXiv preprint arXiv:2304.02182* (2023).

[8] Qingnan Jiang et al. "A challenge dataset and effective models for aspect-based sentiment analysis". In: *Proceedings of the 2019 conference on empirical methods in natural language processing and the 9th international joint conference on natural language processing (EMNLP-IJCNLP)*. 2019, pp. 6280–6285.

[9] Adem Korkmaz, Cemal Aktürk, and TARIK TALAN. "Analyzing the User's Sentiments of ChatGPT Using Twitter Data". In: *Iraqi Journal For Computer Science and Mathematics* 4.2 (2023), pp. 202–214.

[10] Chengxi Li et al. "Sentiprompt: Sentiment knowledge enhanced prompt-tuning for aspect-based sentiment analysis". In: *arXiv preprint arXiv:2109.08306* (2021).

[11] Ruifan Li et al. "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)*. 2021, pp. 6319–6329.

[12] Xin Li et al. "Exploiting BERT for end-to-end aspect-based sentiment analysis". In: *arXiv preprint arXiv:1910.00883* (2019).

[13] Xin Li et al. "Transformation networks for target-oriented sentiment classification". In: *arXiv preprint arXiv:1805.01086* (2018).

[14] Bing Liu. *Sentiment analysis and opinion mining*. Springer Nature, 2022.

[15] Yinhan Liu et al. "Roberta: A robustly optimized bert pretraining approach". In: *arXiv preprint arXiv:1907.11692* (2019).

[16] Dehong Ma et al. "Interactive attention networks for aspect-level sentiment classification". In: *arXiv preprint arXiv:1709.00893* (2017).

[17] Fang Ma et al. "Aspect-specific context modeling for aspect-based sentiment analysis". In: *CCF International Conference on Natural Language Processing and Chinese Computing*. Springer. 2022, pp. 513–526.

[18] Yue Mao et al. "A joint training dual-mrc framework for aspect based sentiment analysis". In: *Proceedings of the AAAI conference on artificial intelligence*. Vol. 35. 15. 2021, pp. 13543–13551.

[19] Sewon Min et al. "Rethinking the role of demonstrations: What makes in-context learning work?" In: *arXiv preprint arXiv:2202.12837* (2022).

[20] Long Ouyang et al. "Training language models to follow instructions with human feedback". In: *Advances in Neural Information Processing Systems* 35 (2022), pp. 27730–27744.

[21] Maria Pontiki et al. "SemEval-2014 Task 4: Aspect Based Sentiment Analysis". In: *Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014)*. Ed. by Preslav Nakov and Torsten Zesch. Dublin, Ireland: Association for Computational Linguistics, Aug. 2014, pp. 27–35. DOI: `10.3115/v1/S14-2004`. URL: `https://aclanthology.org/S14-2004`.

[22] Maria Pontiki et al. "Semeval-2015 task 12: Aspect based sentiment analysis". In: *Proceedings of the 9th international workshop on semantic evaluation (SemEval 2015)*. 2015, pp. 486–495.

[23] Maria Pontiki et al. "Semeval-2016 task 5: Aspect based sentiment analysis". In: *ProWorkshop on Semantic Evaluation (SemEval-2016)*. Association for Computational Linguistics. 2016, pp. 19–30.

[24] Colin Raffel et al. "Exploring the limits of transfer learning with a unified text-to-text transformer". In: *The Journal of Machine Learning Research* 21.1 (2020), pp. 5485–5551.

[25] Kevin Scaria et al. "Instructabsa: Instruction learning for aspect based sentiment analysis". In: *arXiv preprint arXiv:2302.08624* (2023).

[26] Ronald Seoh et al. "Open aspect target sentiment classification with natural language prompts". In: *arXiv preprint arXiv:2109.03685* (2021).

[27] Xuefeng Shi et al. "Prompted and integrated textual information enhancing aspect-based sentiment analysis". In: *Journal of Intelligent Information Systems* (2023), pp. 1–25.

[28] Youwei Song et al. "Utilizing BERT intermediate layers for aspect based sentiment analysis and natural language inference". In: *arXiv preprint arXiv:2002.04815* (2020).

[29] Duyu Tang, Bing Qin, and Ting Liu. "Aspect level sentiment classification with deep memory network". In: *arXiv preprint arXiv:1605.08900* (2016).

[30] Hao Tang et al. "Dependency graph enhanced dual-transformer structure for aspect-based sentiment classification". In: *Proceedings of the 58th annual meeting of the association for computational linguistics*. 2020, pp. 6578–6588.

[31] Kai Wang et al. "Relational graph attention network for aspect-based sentiment analysis". In: *arXiv preprint arXiv:2004.12362* (2020).

[32] Shuai Wang et al. "Target-sensitive memory networks for aspect sentiment classification". In: *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. 2018.

[33] Yequan Wang et al. "Attention-based LSTM for aspect-level sentiment classification". In: *Proceedings of the 2016 conference on empirical methods in natural language processing*. 2016, pp. 606–615.

[34] Zengzhi Wang et al. "Is ChatGPT a good sentiment analyzer? A preliminary study". In: *arXiv preprint arXiv:2304.04339* (2023).

[35] Jason Wei et al. "Emergent abilities of large language models". In: *arXiv preprint arXiv:2206.07682* (2022).

[36] Xiang Wei et al. "Zero-shot information extraction via chatting with chatgpt". In: *arXiv preprint arXiv:2302.10205* (2023).

[37] Hu Xu et al. "BERT post-training for review reading comprehension and aspect-based sentiment analysis". In: *arXiv preprint arXiv:1904.02232* (2019).

[38] Rohan K Yadav et al. "Human-level interpretable learning for aspect-based sentiment analysis". In: *Proceedings of the AAAI Conference on Artificial Intelligence*. Vol. 35. 16. 2021, pp. 14203–14212.

[39] Chen Zhang, Qiuchi Li, and Dawei Song. "Aspect-based sentiment classification with aspect-specific graph convolutional networks". In: *arXiv preprint arXiv:1909.03477* (2019).

[40] Mi Zhang and Tieyun Qian. "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)*. 2020, pp. 3540–3549.

[41] Richong Zhang et al. "Aspect-Level Sentiment Analysis via a Syntax-Based Neural Network". In: *IEEE/ACM Transactions on Audio, Speech, and Language Processing* 30 (2022), pp. 2568–2583.

[42] Yongchao Zhou et al. "Large language models are human-level prompt engineers". In: *arXiv preprint arXiv:2211.01910* (2022).

Authors

Sami Alanazi received a Master of Science in Computer Science from the University of Colorado Denver. Completed a Bachelor of Science in Computer Science from Northern Border University. Currently pursuing a PhD in Computer Science from Florida State University. Research interests encompass Natural Language Processing, Sentiment Analysis, and Aspect-Based Sentiment Analysis.

Xiuwen Liu received PhD from the Ohio State University in Computer and Information Science. Currently, he is a full professor in Computer Science at Florida State University. His research areas include deep learning, cyber security, and their applications.