Current Science
CS 5 (2), 1836-1846 (2025)
Received 23 **Jan** 2025 | **Revised 22 Feb** 2025 | **Accepted** 16 **March**. 2025 | **Online Available 15 April** 2025
https://doi.org/ 10.52845/CS/2025-5-2-46

Current Science
Current Science

OPEN ACCESS
ISSN(O)2795-8639

ORIGINAL ARTICLE

Textual Aspect Sentiment Analysis Method Based on Llama-3 Model Optimization

Yilei Xia[1], Huiying Ru[1*], Chao Li[1], Jingqi Zhang[1]

[1]Department of Information Engineering, Hebei University of Architecture, Zhangjiakou, Hebei Province 075000, China

[*]Corresponding Author: Huiying Ru

Abstract

This paper proposes a textual aspect sentiment analysis method based on Llama-3 model optimization to address challenges in fine-grained sentiment classification. The approach integrates two key innovations: (1) Prompt engineering to reconstruct datasets into structured "Aspect-Polarity" pairs, enhancing data quality and reducing noise; (2) LoRA-based fine-tuning, which introduces low-rank adapters to efficiently update model weights while minimizing computational costs. Experiments on the Semeval-2016 dataset demonstrate superior performance, with the optimized Llama-3 model achieving 92.29% accuracy and 90.48% F1-score, outperforming both open-source and closed-source models. Ablation studies confirm the synergistic benefits of prompt engineering and LoRA fine-tuning. The method balances efficiency and accuracy, offering a practical solution for aspect sentiment analysis tasks.

Keywords: Aspect Sentiment Analysis; Llama-3 Model; Prompt Engineering; LoRA Fine-Tuning; Textual Sentiment Classification; Pre-trained Language Models

Introduction

Aspect Sentiment Analysis aims to extract the sentiment tendency of a particular aspect from a text, i.e., to identify whether the sentiment expressed in the text is positive, negative, or neutral with respect to a particular aspect of a particular object. For example, in the sentence "The screen display of this cell phone is very good, but the battery life needs to be improved", "screen display" and "battery life" are two different aspects. Aspects of sentiment analysis to determine the emotional attitude towards these two aspects, the former is a positive sentiment, the latter is a negative sentiment. Aspect Sentiment Analysis can help enterprises and organizations understand more precisely how users feel about products, services or topics by mining users' emotional tendencies towards different aspects in a fine-grained way, so as to optimize products, improve services, enhance user experience and support better decision-making, which has a wide range of application prospects.

As ChatGPT[1] (a conversational large language model released by OpenAI, Inc.) catches fire, pre-trained large language models based on Transformer have shown impressive performance in several natural language processing tasks[2]. Aspect sentiment analysis, a fine-grained sentiment classification task, requires fine-tuning of large language models to accurately recognize the sentiment polarity of aspects in the input text content[3]. However, the scarcity of high-quality fine-grained annotated textual data and the exponential growth of training parameters in large language models make stable fine-tuning increasingly challenging in practical applications. In addition, Transformer-based pre-trained models are often considered as uninterpretable black-box models[4], which makes it difficult to directly modify the internal architecture of large language models to introduce additional knowledge. In recent research, prompt learning has gradually gained attention as a method to guide large language models to understand downstream tasks.

liu[5] et al. propose a prompt engineering approach for zero-sample and interpretable log analysis, which attempts to solve the interpretability problem in log analysis through the mechanism of prompt + llm, using simple cot prompts as well as contextual information, as well as the labeling problems, and achieved better results on tasks such as anomaly detection.Battle[6] et al. explored the effects of different prompt engineering strategies on the ability of a large language model to solve elementary school math problems, pointing out that by letting the language model design its own optimal prompt is a feasible alternative, and given some examples and quantitative success metrics, the performance of automatically generated prompts often better and faster than the optimal prompt found through trial-and-error methods.

In order to overcome the above challenges of the big language model in the field of aspectual sentiment analysis, this paper proposes a textual aspectual sentiment analysis method based on the optimization of the Llama-3 model, which is divided into two parts: the first part of the method is designed to guide the model in generating the responses by designing a suitable prompt through prompt engineering and optimizing the original dataset by reconstructing the original dataset through the designed prompts, so as to classify the words of the text in the dataset classify aspectual categories so that the trained model can better understand the aspectual sentiment analysis task; the second part fine-tunes the large language model by adding additional neural network layers and training it using the optimized dataset, which significantly improves the aspectual sentiment analysis capability of the large language model. The model chosen for this paper is the Llama-3-8B model, which is chosen because it is the latest open source model from Meta, and its open source nature makes it easy to obtain and use; secondly, due to its strong model performance, it has high scores on several evaluation datasets, and it is close to the top closed-source model, GPT-3.5. The model parameter of 8B is used, which is smaller than the Llama-3-70B model, and the model parameter of 8B is used. The model parameter is 8B, compared with the Llama-3-70B model, choosing a smaller model for training requires less video memory and is easier for personal experimental research.

2 Method

2.1 Optimization of Aspect Sentiment Data Based on Prompt Engineering

In the field of fine-tuning large language models, the quality of a dataset is far more important than its quantity[7], and high-quality aspectual sentiment analysis datasets are a key foundation for model performance improvement. In fine-grained sentiment analysis tasks, datasets need to be accurately labeled with the sentiment polarity (e.g., positive, negative, or neutral) of multiple aspects in the text, and at the same time, they are required to cover diverse linguistic expressions and domain scenarios. However, existing datasets generally face challenges such as high labeling cost, noise interference, and poor domain generalization, which may lead to model learning bias if models are trained directly without processing these datasets.

To solve the above problems, this study introduces the prompt engineering technique to optimally reconstruct the original dataset. Prompt engineering guides the large language model to understand the task requirements by designing structured templates and generates labeled data that meets the requirements. The core idea is to utilize the model's a priori knowledge to transform the raw text into Prompt-Response Pair (PRP), so as to automate the generation of high-quality annotations. The following section will focus on the optimization of the prompt engineering designed in this chapter for the dataset.

The main studies of aspectual sentiment analysis usually include two key components, which are goals and sentiments. Among them, goals can be described in terms of aspect categories (Aspect Category, C) or aspect terms (Aspect Term, A), whereas affect involves detailed expression of opinions-Opinion Term (O) and general affective orientation-Sentiment Polarity (P)[8]. These sentiment elements form the main line of research in aspectual sentiment analysis, which can avoid research confusion. Specifically, it is shown in Figure 1. In this section, by designing the prompt engineering template, the text form of the dataset in non-standardized format is transformed into the "A-P" form of sentiment analysis text, i.e., an aspect term corresponds to the form of a sentiment polarity, which greatly facilitates the in-depth understanding of the content of the text of the sentiment analysis in the big language model .

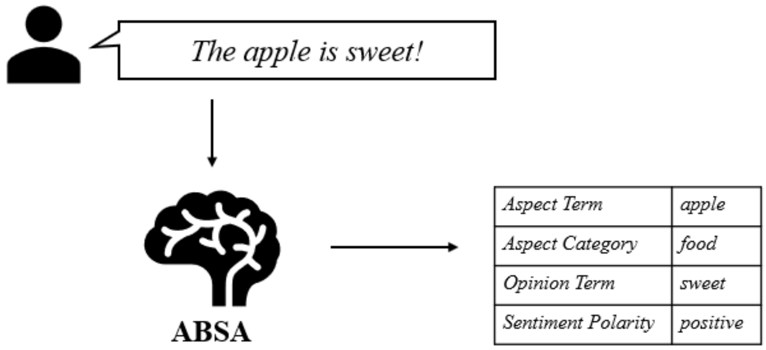

Figure 1. Four key sentiment elements of ABSA

Suppose the original dataset contains text sequences $X = \{x_1, x_2, \dots, x_N\}$, where each sample x_i corresponds to unlabeled aspectual sentiment information. By designing a cue template Γ , x_i can be transformed into a structured cue p_i and the model can be used to generate a response r_i to form the optimized dataset $D' = \{(p_i, r_i)\}_{i=1}^{M}$. The details are as follows:

First, a prompt template Γ was designed to define the "A-P" form of the prompt template Γ , which contains contextual descriptions, question instructions, and answer formats according to the needs of the aspectual sentiment analysis task. The design is as follows:

$$\begin{gathered}\Gamma(x_i) = \text{"You are an aspect} - \text{based sentiment analysis model.} \\ \text{Please analyze the Aspect in the sentence and their Sentiment polarity.} \\ x_i \\ \rightarrow \text{answer: aspect: } \{Aspect\}\text{; sentiment polarity: } \{Polarity\}\text{"}\end{gathered}$$

Input $\Gamma(x_i)$ into the model to generate data to generate responses through conditional probabilities. The formula is shown in (2-1).

$$r_i = argmax_y P\big(y \big| \Gamma(x_i)\big) \#(2-1)$$

Where y represents the output of the model, i.e., the response generated by the model, and $y \in Y$, Y is the legitimate output space containing the set of aspect-sentiment pairs. This formula enables the model to select the response with the highest probability from the legitimate space Y as the final output.

Next, the generated data needs to be filtered and post-processed to select high-quality samples from all the responses generated by the model to ensure the quality of the final dataset. The formula is shown in (2-2).

$$D' = \{(p_i, r_i) | Confidence(r_i) \geq \tau\} \#(2-2)$$

Where D' denotes the filtered high quality dataset where each sample is a cue-response pair (p_i, r_i) . $Confidence(r_i)$ denotes the confidence level of the model-generated response r_i . The confidence level is calculated by the probability distribution of the model output, which is shown in (2-3). τ denotes the confidence threshold, which is used to filter low confidence responses.

$$Confidence(r_i) = P(r_i | \Gamma(x_i)) \#(2-3)$$

where $P(r_i | \Gamma(x_i))$ is the probability that the model generates a response r_i given the cue $\Gamma(x_i)$.

After being processed by the prompt engineering designed above, the accuracy and consistency of the dataset can be effectively improved, and the negative impact of noisy data on model training can be reduced. Table 1 shows some text data comparisons before and after optimization. For example, the original sentence "The plot of this movie is very exciting, but the special effects are a bit fake." can be accurately parsed by the model after the cue engineering template transformation as " aspect: plot; sentiment polarity: positive" and "aspect: special effects; sentiment polarity: negative".

Table 1. Design and Transformation of Prompt Text

	Template	**Example**
Context	The { Aspect 1 } is { Opinion 1 },the { Aspect 2 } is { Opinion 2 } and...	The plot of this movie is very exciting, but the special effects are a bit fake.
Question	You are an aspect-based sentiment analysis model. Please analyze the Aspect in the sentence and their Sentiment polarity.	You are an aspect-based sentiment analysis model. Please analyze the aspect in the sentence and their sentiment polarity.
Answer	aspect :{ Aspect 1 }; sentiment polarity: { Polarity 1 }; aspect :{ Aspect 2 }; sentiment polarity: { Polarity 2 };...	aspect: plot; sentiment polarity: positive; aspect: special effects; sentiment polarity: negative

Finally, in order to evaluate the effect of prompt engineering on data quality improvement, the information gain metric $\mathcal{G}$ is introduced, which is used to measure the degree of uncertainty reduction of the dataset after prompt engineering optimization, and is derived by comparing the entropy of the original dataset $H(Y)$ with the conditional entropy of the optimized dataset after the prompt engineering $H(Y|\Gamma(x_i))$. $H(Y)$The formulas of and $H(Y|\Gamma(x_i))$ are shown in (2-4) (2-5).

$$H(Y) = -\sum_{y \in Y} P(y) log P(y) \,\#(2-4)$$

$$H\big(Y|\Gamma(x_i)\big) = -\sum_{y \in Y} P(y|\Gamma(x_i)) log P(y|\Gamma(x_i)) \,\#(2-5)$$

where $P(y)$ and $P(y|\Gamma(x_i))$ are the probability distributions of the output y in the original dataset and given the cue $\Gamma(x_i)$, respectively.

After calculating the above two entropies, the difference can be taken to get the information gain of a single sample x_i , that is, the degree of uncertainty reduction after the optimization of the cue engineering. Then take the average value of the information gain of all the samples to get the information gain of the whole dataset $\mathcal{G}$, and the formula of the information gain index $\mathcal{G}$ is shown in (2-6).

$$\mathcal{G} = \frac{1}{N} \sum_{i=1}^{N} \Big(H(Y) - H\big(Y|\Gamma(x_i)\big)\Big) \,\#(2-6)$$

In summary, by designing prompt engineering for aspectual sentiment analysis data optimization, the original dataset is reconstructed into structurally clear, accurately labeled, high-quality training data, which lays a solid foundation for subsequent model fine-tuning.

2.2 Aspect Sentiment Analysis Based on Model Fine-Tuning

Along with the rapid development of artificial intelligence, aspectual sentiment analysis for text has made great progress in the past decade or so, especially after the proposal of the Transformer architecture, a variety of large-scale language models are rapidly emerging, which further promotes the development of textual aspectual sentiment analysis. For example, very large models with parameters exceeding 100B, such as GPT-3[9]and BLOOM[10], have both demonstrated excellent capabilities in the field of aspectual sentiment analysis to understand and recognize the sentiment of specific aspects of a text. However, despite the exciting advances and practical applications of large language models, most of the well-known language models, such as

OpenAI's GPT-4[11]and Google's PaLM2[12], are still closed-source models, and this restriction of access to the parameters of the full model poses a challenge to the study of large language models. In order to break the monopoly of closed-source models and bring more openness, Meta-AI released the Llama open-source model, and the recently released Llama-3 model further challenged the status of closed-source models with its inference efficiency[13]. While Llama-3 excels as a general-purpose large language model in natural language understanding tasks, it still has limitations in fine-grained aspectual sentiment analysis tasks. Aspect sentiment analysis requires the model to not only understand the global semantics of a text, but also pinpoint specific aspects and infer their sentiment polarity. However, the pre-training data for generic models covers a wide range of domains and lacks targeted modeling of domain-specific fine-grained sentiment. Therefore, adapting Llama-3 to aspectual sentiment analysis tasks through fine-tuning is a key step to improve its performance. Fine-tuning can adjust the distribution of the model's parameters to better fit the characteristics of the target task, and enhance the model's sensitivity to aspectual terminology and accuracy of sentiment classification[14]. In this section, in order to further improve the performance of the Llama-3 model in the domain of aspectual sentiment analysis, it is fine-tuned using a cue-engineered optimized sentiment dataset.

Traditional Full-Parameter Fine-Tuning requires updating all the parameters of the model, which is very expensive in terms of memory usage and computational cost for models such as Llama-3-8B, which contains billions of parameters. Full-parameter fine-tuning can also lead to Catastrophic Forgetting, which causes the model to lose its original generalized knowledge when adapting to new tasks. In contrast, parameter-efficient fine-tuning methods become a better choice for improving model aspects of sentiment analysis. In this section, we introduce Low-Rank Adaptation (LoRA)[15], which introduces trainable bypass matrices in the model weights through Low-Rank Decomposition to achieve task adaptation with minimal parameter updates.The structure of LoRA module is shown in Figure 2.

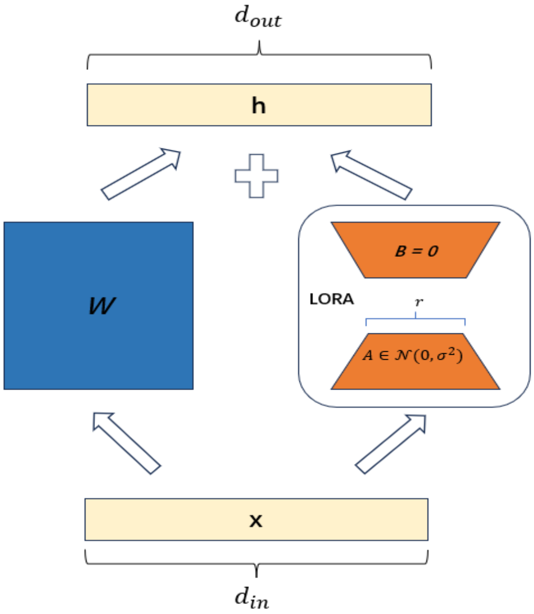

Figure 2. Structure diagram of LoRa

Specifically, for a given linear layer weight matrix$W \in \mathbb{R}^{d_{in} \times d_{out}}$ of the pretrained model,d_{in} denotes the dimension of the feature vectors input to the linear layer, andd_{out} denotes the dimension of the feature vectors output from the linear layer. LoRA decomposes its parameter update quantityΔW into the product of two low-rank matrices$B$ and$A$. Its formula is shown in (2-7).

$$\Delta W = B \cdot A\#(2-7)$$

where the low-rank matrix,$B \in \mathbb{R}^{d_{in} \times r}$, is used to map the input from the high-dimensional space (d_{in} dimensions) to the low-dimensional space (r

dimensions); and the low-rank matrix, $A \in \mathbb{R}^{r\times d_{out}}$, is used to map the representation in the low-dimensional space back to the high-dimensional space (d_{out} dimensions). r is the low-rank factor, which is used to control the dimension of the low-rank decomposition. Here, Gaussian initialization and zero initialization are applied to matrix A and matrix B, respectively, so that $B \cdot A$ is initially zero to prevent additional noise to the model at the beginning of training.

When fine-tuning is performed, the original weights of the model W are frozen and only the low-rank matrices A and B are trained. At this point the output calculation of this layer is changed and its formula is shown in (2-8).

$$h = Wx + \alpha \cdot (B \cdot A)x \#(2-8)$$

where α is a scaling factor to balance the contribution of pre-trained knowledge with new task adaptation.

After introducing LoRA for fine-tuning, the amount of parameter updates of the training model is reduced significantly, and the number of its training parameters is reduced from the original $d_{in} \times d_{out}$ tor $\times (d_{in} + d_{out})$. For the Llama-3 model selected in this paper, this study embeds the LoRA module into its attention mechanism. Adding LoRA bypasses to the query and key projection matrices of the multi-head attention module, respectively, enables the model training to capture task-relevant feature interaction patterns without significantly increasing the computational overhead. This is shown in Figure 3.

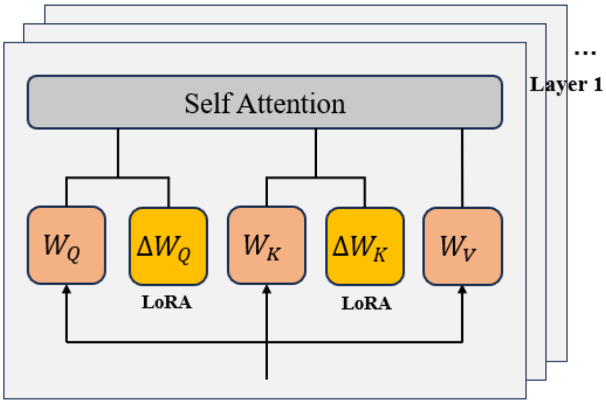

Figure 3. Embedding position of the LoRA module in the Llama-3 model

During the training process, the last fully-connected layer of the model outputs the original logarithm set $z = [z_1, z_2, \dots, z_C]$, where C is the number of categories, $C = 3$,because in this paper, the sentiment polarity is labeled into three categories, which are positive, negative, and neutral. Next, the model will convert the original logarithmic set z to a probability distribution $p = [p_1, p_2, \dots, p_C]$ by Softmax function before outputting the answers, and its conversion formula is shown in Equation (2-9).

$$p_C = \frac{e^{z_C}}{\sum_{j=1}^{C} e^{z_j}} \#(2-9)$$

After that, the LoRA parameters are optimized using the cross-entropy loss function, which is used to measure the difference between the probability distribution predicted by the model and the true label distribution, and its formula is shown in (2-10).

$$\mathcal{L} = -\sum_{c=1}^{C} y_c log p_c \#(2-10)$$

where y_c denotes the true label of the sample in category c and p_c is the probability that the model predicts that the sample belongs to category c .

Finally, while keeping the original parameters of the Llama-3 model fixed, the low-rank matrix A and the matrix B are updated by the back-propagation algorithm. This not only preserves the model's original general-purpose language comprehension ability, but also significantly improves the model's performance in the field of textual aspect sentiment analysis.

3 Results

In order to better evaluate the optimized textual aspect sentiment analysis approach based on the

Llama-3 model, experiment-specific details, including the dataset, experimental setup, evaluation metrics, comparison experiments, and experimental results of the ablation experiments, are listed in this section.

3.1 Introduction to Data Sets

The dataset selected for the experiments in this paper is Semeval-2016 Task 5: Aspect-Based Sentiment Analysis[16], which is an important dataset in the field of aspectual sentiment analysis, and is mainly used to evaluate and advance the development of aspectual sentiment analysis techniques. The text in the dataset is labeled with fine-grained annotations, including aspect categories, sentiment polarity, and other key information. It covers data from a number of different domains, such as digital cameras, cell phones, restaurants, laptops, and other domains. The dataset provides researchers with a unified standard and data base that helps to compare the performance of different models and algorithms on this task.

Given that this dataset is too large, and using an excessive amount of data to fine-tune the model with lower parameters may lead to the problem of model overfitting, thus affecting the training effect of the model. Therefore, in this paper, 6750 of these text data are screened for experiments, of which 80% are used as a high-quality dataset for training the model after template optimization, 5% are used as a validation set for verifying the experimental effect, and 15% are used as a test set for evaluating the final performance of the model.

3.2 Experimental Environment

The training of large language model has high requirements on the configuration of the experimental environment, this paper builds a set of stable and efficient experimental environment for this purpose, the specific experimental environment is shown in Table 2.

Table 2. Experimental environment table

name (of a thing)	**releases**
operating system	Ubuntu 20.04.6 LTS
processing unit	Intel® Core™ i7-14700kf
GPU	NVIDIA® GeForce RTX™ 4070 Ti SUPER
Deep Learning Framework	Pytorch 2.5.1
programming language	Python 3.10

For the training configuration, the experiments use the SFTTrainer trainer from the trl package with the AdamW_8bit optimizer, the maximum sentence length of the model input is set to 128tokens, the low-rank factor$r$ is set to 16, the learning rate is set to $2e^{-4}$, and the number of training rounds epoch is set to 60. These settings were chosen to ensure stable training, and Figures 4 provide detailed information on the training loss, from which it can be seen that the training loss is trending down and eventually stabilizing, highlighting the effectiveness of the selected configuration.

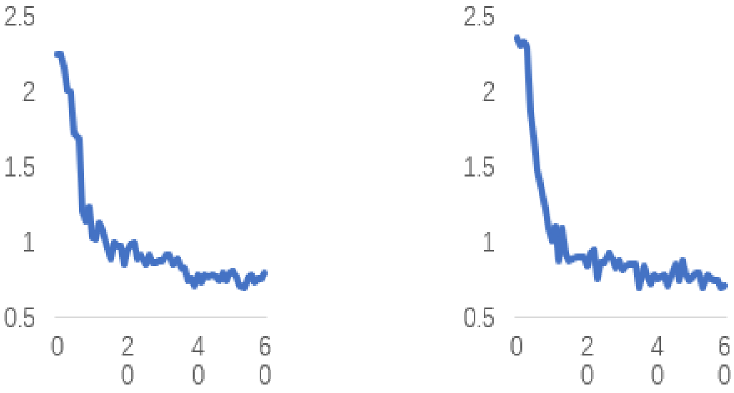

Figure 4. Training Loss Curve

3.3 Assessment of Indicators

In natural language processing tasks, evaluation metrics are a core tool for measuring model performance. Through quantitative metrics, we can objectively compare the advantages and disadvantages of different models, analyze the strengths and weaknesses of models in specific tasks, and verify the rationality of experimental design. For the task of aspectual sentiment analysis, it is especially important to choose appropriate evaluation metrics because it involves multi-class fine-grained classification. In this paper, Accuracy and F1-Score are chosen as the main evaluation indexes to fully reflect the comprehensive performance of the model in the aspectual sentiment analysis task.

Accuracy is the most intuitive index of classification performance, indicating the ratio of the number of samples correctly predicted by the model to the total number of samples. Its calculation formula is shown in Equation (3-1).

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \#(3-1)$$

Among them,TP (True Positive) is the true example, i.e., the number of samples in which the actual positive emotions were identified as positive;TN (True Negative) is the true negative example, i.e., the number of samples in which the actual negative emotions were identified as negative;FP (False Positive) is the false positive example, i.e., the number of samples in which the actual negative emotions were identified as positive;FN (False Negative) is the number of false negative examples, i.e., the number of samples in which actual positive emotions were identified as negative.

The F1 score is a more comprehensive reflection of the model's ability to categorize each fine-grained category, and is derived by calculating the reconciled average of Precision and Recall. Precision is the proportion of samples predicted by the model to be positive that are actually positive, reflecting the reliability of the model's prediction of positive categories; Recall is the proportion of samples that are actually positive that are correctly predicted, reflecting the model's ability to cover positive categories. The formulas for precision rate, recall rate and F1 score are shown in Equations (3-2) (3-3) (3-4), respectively.

$$Precision = \frac{TP}{TP + FP} \#(3-2)$$

$$Recall = \frac{TP}{TP + FN} \#(3-3)$$

$$F1 = \frac{2 \times Precision \times Recall}{Precision + Recall} \#(3-4)$$

3.4 Comparison Experiments

In order to verify the effectiveness of the optimized textual aspect sentiment analysis method based on Llama-3 model proposed in this paper, this chapter conducts comparison experiments between the optimized Llama-3 model and the mainstream large language model in the same period. The details of the comparison models are as follows:

Llama-2-7B[17]: a second-generation open-source model released by Meta in 2023 with a parameter scale of 7B, improved based on the Llama-1 architecture, whose training data covers multi-domain text and supports zero- and few-sample learning.

Mistral-7B[18]: an efficient open-source model released by Mistral AI in 2023 with a parameter scale of 7B, which employs a sliding-window attention mechanism and sparse activation strategy to significantly reduce the inference explicit memory footprint.

Gemma-7B[19]: a lightweight open-source model released by Google DeepMind in 2024, with a parameter scale of 7B, optimized based on the Transformer architecture, using dynamic mask training with efficient sparse activation strategy, supporting multilingual tasks and academic research scenarios.

GPT-3.5: OpenAI's closed-source model released in 2022, based on the improved Transformer decoder architecture, which excels in the areas of generative tasks and less-sample learning, participates in the comparison as a performance benchmark.

All models were evaluated on the same test set, and the experimental environment was kept consistent with the preprocessing process. The experimental results are shown in Table 3.

Table 3. Results of comparative experiments

Model	**Accuracy(%)**	**Precision(%)**	**Recall(%)**	**F1-Score (%)**
Llama-2-7B	88.73	87.28	85.12	86.19
Mistral-7B	88.93	87.64	85.82	86.72
Gemma-7B	89.43	88.21	86.67	87.43
GPT-3.5	91.50	90.12	88.89	89.50
Ours	**92.29**	**91.52**	**89.47**	**90.48**

The experimental results show that the optimized Llama-3 model significantly outperforms other open-source models in all four metrics, with an accuracy improvement of 3.56%, 3.36%, and 2.86%, and an F1 score improvement of 4.29%, 3.76%, and 3.05%, respectively, compared to Llama-2-7B, Mistral-7B, and Gemma-7B. This indicates that the cue engineering optimization and LoRA fine-tuning strategies proposed in this paper effectively improve the model's fine-grained understanding of aspectual sentiment analysis. Compared with the closed-source model GPT-3.5, the optimized Llama-3 model is 0.79% and 0.98% higher in accuracy and F1 score, respectively, which verifies that the open-source model can outperform the larger-scale closed-source model in specific tasks through targeted optimization. Meanwhile, the optimized Llama-3 model has a balanced performance in precision rate and recall rate, which are 91.52% and 89.47%, respectively, indicating that it can both effectively reduce misclassification (FP low) and cover more real positive examples (FN low).

3.5 Ablation Experiments

Since the textual aspect sentiment analysis method based on Llama-3 model optimization proposed in this chapter contains two improvements, cue engineering optimization and LoRA fine-tuning, in order to verify more deeply the independent contribution of these two improvements to model performance enhancement and their synergistic effects, ablation experiments are designed in this section to compare and analyze the following four model configurations, respectively:

Original Llama-3-8B model (Baseline): a pre-trained model without any optimization, used directly for aspectual sentiment analysis tasks.

Llama-3-8B model (Baseline + Prompt) with only prompted engineering optimization introduced: inference using the prompted engineering-optimized dataset, but without LoRA fine-tuning.

Llama-3-8B model with LoRA fine-tuning only (Baseline + LoRA): LoRA fine-tuning using the original dataset, no prompt engineering optimization.

Combining prompt-engineering optimization with LoRA fine-tuning for the Llama-3-8B model (Ours): the complete method, using both the optimized dataset and LoRA fine-tuning.

The dataset and experimental environment used in the experiments remain the same as the previous experiments. The experimental results are shown in Table 4.

Table 4. Results of ablation experiments

Method	**Accuracy(%)**	**Precision(%)**	**Recall(%)**	**F1-Score (%)**
Baseline	89.53	88.12	86.89	87.50
Baseline + Prompt	91.30 **(+1.77)**	89.75 **(+1.63)**	88.47 **(+1.58)**	89.11 **(+1.61)**
Baseline + LoRA	91.80 **(+2.27)**	90.02 **(+1.90)**	88.92 **(+2.03)**	89.47 **(+1.97)**
Ours	92.29 **(+2.76)**	91.52 **(+3.40)**	89.47 **(+2.58)**	90.48 **(+2.98)**

Analyzing the experimental results, the accuracy and F1 score of Baseline + Prompt are improved by 1.77% and 1.61% respectively compared to Baseline, which is due to the optimized dataset generated by the designed prompt engineering template reduces the noise, which makes the model more accurately capture the association between aspectual terminology and sentiment polarity, which proves the effectiveness of the designed prompt engineering in this paper. The accuracy and F1 score of Baseline + LoRA are improved by 2.27% and 1.97%, respectively, compared to Baseline, which proves the effectiveness of introducing the LoRA low-rank adapter to adjust the model weights in this paper. Finally, the model that jointly uses these two methods shows further improvement in accuracy and F1 score, especially the F1 score is significantly improved due to the joint optimization which enhances the robustness of the model, proving that there is a synergistic effect between the high-quality data optimized by the prompt engineering and the task-adaptation parameters fine-tuned by LoRA.

4 Conclusions

This paper proposes a textual aspectual sentiment analysis method based on the optimization of the Llama-3 model, and the design of the prompt engineering template solves the problems of insufficient data annotation and noise interference in the aspectual sentiment analysis domain of the existing large language models. In addition, the introduction of the trainable bypass matrix LoRA in the model weights can effectively reduce the computational overhead during the model training process, so that the model can improve the performance of the model in the field of textual aspectual sentiment analysis while retaining the original general-purpose language comprehension ability. The experimental results show that the optimized Llama-3 model outperforms some current mainstream large language models in the textual aspect sentiment analysis task, proving the effectiveness of the improvement of this method.

Conflicts of Interest: The authors declare no conflict of interest.

Funding: 2023 Basic research business cost project for universities of Hebei Province (Basic Research Innovation Team Project of natural science)(No.2023JCTD07)

References

1. Aljanabi, M., Ghazi, M., Ali, A. H., & Abed, S. A. (2023). ChatGpt: open possibilities. Iraqi journal for computer science and mathematics, 4(1), 7.
2. Min, B., Ross, H., Sulem, E., Veyseh, A. P. B., Nguyen, T. H., Sainz, O., ... & Roth, D. (2023). Recent advances in natural language processing via large pre-trained language models: A survey. ACM Computing Surveys, 56(2), 1-40.
3. Brauwers, G., & Frasincar, F. (2022). A survey on aspect-based sentiment classification. ACM Computing Surveys, 55 (4), 1-37.
4. Zhao, H., Chen, H., Yang, F., Liu, N., Deng, H., Cai, H., ... & Du, M. (2024). Explainability for large language models: A survey. *ACM Transactions on Intelligent Systems and Technology*, *15*(2), 1-38.
5. Liu, Y., Tao, S., Meng, W., Wang, J., Ma, W., Chen, Y., ... & Jiang, Y. (2024, April). Interpretable online log analysis using large language models with prompt strategies. In Proceedings of the 32nd IEEE/ACM International Conference on Program Comprehension (pp. 35-46).
6. Battle, R., & Gollapudi, T. (2024). The unreasonable effectiveness of eccentric automatic prompts. arXiv preprint arXiv:2402.10949.
7. Gunasekar, S., Zhang, Y., Aneja, J., Mendes, C. C. T., Del Giorno, A., Gopi, S., ... & Li, Y. (2023). Textbooks are all you need. *arXiv preprint arXiv:2306.11644*.
8. Zhang, W., Li, X., Deng, Y., Bing, L., & Lam, W. (2022). A survey on aspect-based sentiment analysis: Tasks, methods, and challenges. IEEE Transactions on Knowledge and Data Engineering, 35(11), 11019-11038.
9. Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., ... & Amodei, D. (2020). Language models are few-shot learners. Advances in neural information processing systems, 33, 1877-1901.
10. Le Scao, T., Fan, A., Akiki, C., Pavlick, E., Ilić, S., Hesslow, D., ... & Al-Shaibani, M. S. (2023). Bloom: A 176b-parameter open-access multilingual language model.
11. Achiam, J., Adler, S., Agarwal, S., Ahmad, L., Akkaya, I., Aleman, F. L., ... & McGrew, B. (2023). Gpt-4 technical report. arXiv preprint arXiv:2303.08774.
12. Anil, R., Dai, A. M., Firat, O., Johnson, M., Lepikhin, D., Passos, A., ... & Wu, Y. (2023). Palm 2 technical report. *arXiv preprint arXiv:2305.10403*.
13. Grattafiori, A., Dubey, A., Jauhri, A., Pandey,

A., Kadian, A., Al-Dahle, A., ... & Vasic, P. (2024). The llama 3 herd of models. *arXiv preprint arXiv:2407.21783*.

14. Han, Z., Gao, C., Liu, J., Zhang, J., & Zhang, S. Q. (2024). Parameter-efficient fine-tuning for large models: A comprehensive survey. *arXiv preprint arXiv:2403.14608*.
15. Zhang, L., Zhang, L., Shi, S., Chu, X., & Li, B. (2023). Lora-fa: Memory-efficient low-rank adaptation for large language models fine-tuning. arXiv preprint arXiv:2308.03303.
16. Pontiki, M., Galanis, D., Papageorgiou, H., Androutsopoulos, I., Manandhar, S., Al-Smadi, M., ... & Eryiğit, G. (2016, January). Semeval-2016 task 5: Aspect based sentiment analysis. In International workshop on semantic evaluation (pp. 19-30).
17. Touvron, H., Martin, L., Stone, K., Albert, P., Almahairi, A., Babaei, Y., ... & Scialom, T. (2023). Llama 2: Open foundation and fine-tuned chat models. arXiv preprint arXiv:2307.09288.
18. Jiang, A.Q., Sablayrolles, A., Mensch, A., Bamford, C., Chaplot, D.S., Casas, D.D., Bressand, F., Lengyel, G., Lample, G., Saulnier, L., Lavaud, L.R., Lachaux, M., Stock, P., Scao, T.L., Lavril, T., Wang, T., Lacroix, T., & Sayed, W.E. (2023). Mistral 7B. *ArXiv, abs/2310.06825*.
19. Team, G., Mesnard, T., Hardin, C., Dadashi, R., Bhupatiraju, S., Pathak, S., ... & Kenealy, K. (2024). Gemma: Open models based on gemini research and technology. *arXiv preprint arXiv:2403.08295*.