Available Online: 19 May 2026 DOI: 10.54254/2977-3903/2026.33698

Research on a sentiment analysis method based on sentiment chain-of-thought

Xiaohui Fan

FanNorth China University of Technology, Beijing, China

1791174842@qq.com

Abstract. With the rapid development of the Internet and social media, the multidimensional subjective sentiments embedded in massive volumes of user-generated content have become a critical data foundation for business intelligence, public opinion monitoring, and public decision-making, thereby driving the rapid advancement of sentiment analysis technologies. Traditional models typically treat sentiment analysis as an end-to-end pattern matching or classification task, lacking explicit modeling of the intrinsic logic underlying sentiment formation. As a result, their robustness and interpretability remain limited when handling challenging scenarios such as implicit sentiments and complex sentence structures. In particular, when confronted with cases that require deep semantic understanding and multi-step reasoning to accurately determine sentiment polarity, existing methods often prove inadequate. To address these issues, this paper proposes a sentiment analysis method based on a sentiment chain-of-thought framework combined with LoRA fine-tuning. The proposed approach leverages large language models to generate high-quality reasoning data that conform to a five-step protocol, transforming implicit sentiment judgments into explicit, structured chains of thought encompassing entity localization, contextual analysis, and transition detection. In addition, quantized low-rank adaptation is employed to conduct full-sequence joint modeling training, which significantly reduces computational and memory overhead while minimizing performance degradation, thereby enabling cost-efficient training of large models. Experimental results demonstrate that the proposed method outperforms traditional deep learning baselines and generic chain-of-thought fine-tuning approaches on the Laptop and Restaurant datasets of SemEval-2014 Task 4, validating its effectiveness.

Keywords: sentiment analysis, large language models, chain of thought, prompt templates, LoRA fine-tuning

1. Introduction

Sentiment Analysis, as one of the core tasks in natural language processing, aims to automatically identify and extract the subjective sentiment orientations embedded in text [1]. Although methods based on pre-trained language models (PLMs) have achieved remarkable progress under the paradigm of large-scale annotated data [2], notable limitations remain in low-resource scenarios—particularly when dealing with domain-specific data characterized by complex contexts, implicit sentiments, and multiple logical transitions [3].

Early discriminative models (e.g., BERT, RoBERTa) typically adopt a paradigm in which classification labels are directly mapped from the [CLS] token [4]. This end-to-end black-box inference is highly prone to shortcut learning in low-data regimes, where models tend to rely on superficial statistical cues such as "good" or "bad," while neglecting deeper logical dependencies [5]. For instance, when processing sentences with adversative relations, such models often fail to capture the shift in semantic focus due to the absence of explicit reasoning steps, leading to misinterpretations of complex rhetorical patterns such as concessive or contrastive structures.

Subsequently, general-purpose generative Large Language Models (LLMs) have demonstrated potential in commonsense reasoning [6]. However, in small-sample tasks within specialized domains, these models frequently exhibit hallucinations or unstable instruction-following behavior if not properly adapted [7]. They may waver between conflicting sentiment cues, producing inconsistent outputs. Although full fine-tuning can mitigate these issues, it is highly susceptible to catastrophic forgetting in low-resource settings, potentially degrading the model's general linguistic competence. Moreover, the substantial computational cost makes it impractical for deployment in resource-constrained environments [8].

Therefore, a key challenge lies in how to inject interpretable, structured, and explicit reasoning capabilities into large models with minimal data cost, without compromising their general-purpose abilities. This motivates a paradigm shift from conventional sequence classification to a sequence generation framework that integrates chain-of-thought reasoning with conclusion derivation.

To address the aforementioned challenges, this study proposes the Emo-CoT-LoRA joint optimization method. Its central idea is to reconstruct sentiment analysis from an implicit "black-box classification" paradigm into an explicit "white-box reasoning" framework. The approach tackles the problem of missing logical structure in low-resource scenarios through a dual synergy between data and model design. On the data side, an automated Emotional Chain-of-Thought (Emo-CoT) mechanism is introduced, which compels the model to generate a complete reasoning trajectory—including entity localization, contextual analysis, and logical transition detection—prior to polarity classification. This process externalizes implicit reasoning, effectively extends the attention span, and mitigates overfitting to surface-level features. On the model side, a parameter-efficient fine-tuning strategy is adopted: by freezing the backbone network's general capabilities, domain-specific logical reasoning patterns are injected solely through low-rank adaptation. This deep integration of explicit reasoning guidance and lightweight parameter adaptation not only ensures accurate modeling of complex contexts and long-range dependencies under extremely limited data conditions, but also significantly enhances interpretability and robustness while avoiding catastrophic forgetting.

2. Methodology

The proposed method aims to transform general-purpose language understanding capabilities into domain-specific logical reasoning abilities for sentiment analysis. It comprises three core components: a structured reasoning protocol, chain-of-thought construction, and parameter-efficient fine-tuning. First, to address the difficulty of capturing implicit logic in fine-grained sentiment analysis, this study defines an Emotional Chain-of-Thought (Emo-CoT) protocol that incorporates entity identification, sentiment word recognition, contextual analysis, transition detection, and conclusion derivation, thereby providing explicit logical constraints for the reasoning process. Second, a "teacher–student" distillation strategy is employed to construct the training corpus. A high-performance large language model serves as the teacher, strictly adhering to the aforementioned protocol to perform inferential expansion on raw samples and generate high-quality intermediate reasoning trajectories, thus externalizing the implicit decision-making process. Finally, at the model training level, Qwen2-7B is adopted as the base model, and quantized Low-Rank Adaptation (LoRA) is introduced for fine-tuning. This mechanism freezes the backbone parameters of the pre-trained model and injects trainable low-rank adapters into transformation matrices, enabling the structured chain-of-thought patterns to be internalized as parameter priors at minimal computational cost. This architecture not only preserves the general semantic representation capacity of the base model, but also equips it with specialized skills for interpretable reasoning along predefined logical pathways, thereby laying a solid foundation for sentiment classification in complex contexts.

2.1. Construction of emotional chain-of-thought

To address the core challenges in low-resource scenarios—namely the scarcity of annotated data, the absence of explicit reasoning signals, and the limited instruction-following capability of lightweight models under complex prompts—this section proposes a data augmentation strategy based on cognitive decomposition. Without relying on large-scale external corpora, the approach constructs a high-quality dataset that simultaneously alleviates data scarcity and improves instruction alignment. Its central idea is to reconstruct traditional sentiment classification tasks from an implicit black-box mapping into an explicit white-box reasoning process. Through structured prompt templates, a high-capacity teacher model is guided to deeply process a small set of labeled samples and transform them into triplets containing complete reasoning trajectories. This process is not intended merely to expand the absolute quantity of data; rather, by injecting explicit intermediate reasoning steps, it significantly increases the informational density of each sample. In doing so, it compensates for the lack of reasoning signals while providing rigorous instruction-following demonstrations for the subsequent student model.

Table 1. Template for emotional chain-of-thought construction

Prompt	Please strictly follow the Five-Step Reasoning Protocol below to analyze the input review:1.Entity Identification: Identify the evaluation targets (Aspects).2.Sentiment Word Recognition: Extract sentiment words and their initial polarity.3.Context Analysis: Analyze the impact of degree adverbs, negation words, and punctuation,).4.Logic Transition: Detect contrastive conjunctions and determine the semantic focus.5.Polarity Determination: Synthesize the above information to derive the final conclusion.
Rationale	Input:"The food was delicious, but the service was incredibly slow."1.Entities: food (Positive), service (Negative) 2.Context: Detected 'incredibly', negative intensity amplified3.Transition: Detected 'but'; semantic focus shifts to the latter clause.4.Determination: The severe defect in service dominates the overall experience.5.Lable: Negative
Input	The atmosphere is cozy and warm, however, the prices are outrageously high for such small portions.

To ensure that the generated chains of thought possess both logical depth and strict adherence to instructions, this method designs a rigorous five-step reasoning protocol as the structural backbone of data construction. The protocol decomposes vague sentiment judgments into five progressively structured cognitive stages: entity identification to clarify evaluation targets, sentiment word recognition to establish the initial tone, contextual analysis to handle implicit modifiers such as negation, irony, and degree adverbs, logic transition detection to capture shifts in semantic focus, and final polarity determination through synthesis of all preceding cues. By embedding few-shot demonstrations covering typical scenarios—including straightforward statements, contrastive conflicts, and ironic expressions—within the prompt template, this mechanism not only guides the model to emulate expert-level reasoning pathways and avoid logical leaps or hallucinations, but also, more critically, enforces structured compliance with multi-step instructions through standardized output formats. This design, in which the protocol itself functions as an instruction, ensures that every generated chain of thought strictly follows causal reasoning logic, thereby establishing a robust paradigm for the model to accurately understand and execute similarly complex instructions during subsequent fine-tuning. Details are presented in Table 1.

This structured data construction strategy transforms the original small-sample dataset into an Emo-CoT-enhanced corpus containing rigorous logical chains. Complex sentiment analysis heuristics—such as "post-transition dominance" and "irony reversal"—are explicitly encoded as textual supervisory signals, while abstract instruction requirements are concretized into learnable reasoning trajectories. These high-information-density augmented samples are subsequently used to guide parameter-efficient fine-tuning, enabling the model, under extremely limited data conditions, to rapidly internalize complex domain-specific reasoning patterns and compensate for missing signals. At the same time, by imitating high-quality chain-of-thought formats, the model significantly improves its ability to follow and execute multi-step reasoning instructions. This approach preserves the low-resource nature of few-shot learning while, through the introduction of explicit reasoning chains, fundamentally shifts the model from shortcut learning based on superficial statistical features to a robust reasoning paradigm grounded in deep logical inference and strict instruction adherence.

2.2. LoRA implementation procedure

In the fine-tuning of Large Language Models (LLMs), full-parameter updates not only incur substantial computational and memory overhead, but also tend to induce catastrophic forgetting in low-resource scenarios, thereby undermining the general linguistic competence and knowledge distribution acquired during pre-training. To address this challenge, this study adopts Low-Rank Adaptation (LoRA) to establish a parameter-efficient fine-tuning paradigm. The core idea is to freeze the weights of the pre-trained backbone network and approximate weight updates by injecting a small number of trainable low-rank decomposition matrices. In this way, the model can be precisely adapted to sentiment chain-of-thought reasoning with minimal resource consumption. The overall architecture is illustrated in Figure 1.

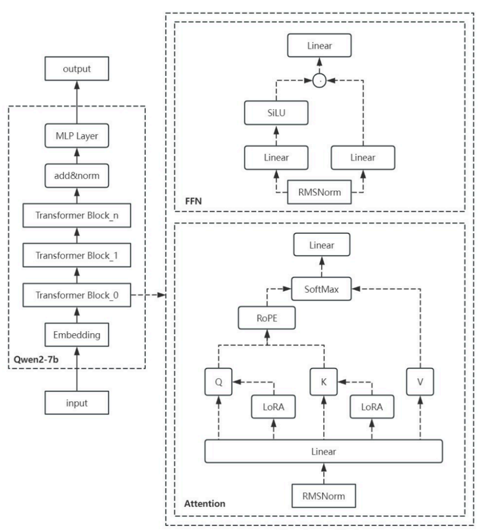

Figure 1. Model architecture

Conventional fine-tuning requires updating all model parameters θ , leading to memory usage that scales linearly with model size. LoRA is based on the assumption that the update to model weights ΔW exhibits an intrinsic low-rank structure. For a pre-trained weight matrix $W_0 \in R^{d \times k}$, the correction term in the forward propagation process can be decomposed into the product of two low-rank matrices $B \in R^{d \times r}$ and $A \in R^{r \times k}$, where the rank $r \ll min(d,k)$.

The modified forward propagation is given in Equation (1):

$$h = W_0 x + \Delta W x = W_0 x + \frac{\alpha}{r}(BA)x \tag{1}$$

Where x denotes the input vector; W_0 is the frozen pre-trained weight matrix, which in this study is further quantized into 4-bit NormalFloat (NF4) format to minimize memory usage; matrix A is randomly initialized using a Gaussian distribution, while B is initialized as a zero matrix, ensuring that the model's behavior at the initial stage of training is equivalent to that of the original pre-trained model; α is a learnable scaling factor used to regulate the contribution of the low-rank update.

To maximize the representational capacity of LoRA, the proposed method injects low-rank adapters in parallel into key linear projection layers of the model. The target module set includes the query, key, value, and output projections in the self-attention mechanism, as well as the gate and up/down projection layers in the Feed-Forward Network (FFN). By applying LoRA across these modules simultaneously, the model is able to capture the fine-grained semantic features required for sentiment reasoning while maintaining an extremely low parameter footprint.

During training, the optimization objective is to minimize the negative log-likelihood loss function, as shown in Equation (2):

$$\mathscr{L}(\theta_{\text{LoRA}}) = -\sum_{i=1}^{T} \log P(y_i y_{<i}, x; \theta_{\text{quant}}, \theta_{\text{LoRA}}) \tag{2}$$

Where θ_{quant} denotes the frozen 4-bit quantized parameters, and θ_{LoRA} represents the low-rank matrix parameters to be optimized.

After training, the low-rank weight matrices BA are merged back into the quantized or dequantized backbone weights to produce the final inference model, as expressed in Equation (3):

$$W_{\text{final}} = \text{Dequantize}(W_0) + \frac{\alpha}{r} BA \tag{3}$$

This merging operation eliminates any additional computational overhead during inference, allowing the final model to retain its original inference speed while acquiring the multi-step sentiment reasoning capability specified by Emo-CoT. Overall, this process realizes a methodological shift from resource-intensive full-parameter fine-tuning to a low-resource, high-efficiency parameter adaptation paradigm.

For completeness, the optimal separating hyperplane of a Support Vector Machine (SVM) is defined by a kernel function, as shown in Equation (4):

$$f(x) = \sum_{i=0}^{N} \alpha_i t_i K(x, y_i) + \beta \tag{4}$$

where $x = [x_1, x_2, \ldots, x_n]^T$, with n denoting the number of features; $y_i = [y_{i1}, y_{i2}, \ldots, y_{in}]^T$; N is the total number of samples in the training set; t_i is the class label of the training sample, taking values of 1 and -1 for positive and negative samples, respectively; α_i and β are parameters learned during SVM training; and $K(x, y_i)$ denotes the kernel function. The kernel function must satisfy the Mercer condition, as expressed in Equation (5):

$$K(x, y_i) = b(x) b(y_i) \tag{5}$$

where $b(x)$ and $b(y_i)$ denote the mappings from the original feature space of x and y_i into a higher-dimensional space, respectively.

3. Results

3.1. Datasets

This study evaluates the proposed model on two core subsets of the SemEval-2014 Task 4 benchmark dataset [9]: Laptop and Restaurant. These datasets represent reviews in the domains of consumer electronics and the catering industry, respectively, and exhibit substantial differences in domain characteristics and semantic complexity.

To eliminate potential bias arising from inconsistent annotation schemes across studies and to ensure comparability of experimental results, this work follows standard preprocessing practices by removing samples labeled as Conflict. Only instances with a single, clearly defined sentiment polarity (positive, negative, or neutral) are retained as valid data.

The detailed statistics are presented in Table 2.

Table 2. Dataset statistics

Dataset	Split	Positive	Neutral	Negative
Laptop	Train	976	455	851
	Test	337	167	128
Restaurant	Train	2,164	637	807
	Test	727	196	196

The Laptop dataset primarily consists of user reviews on laptop hardware features, performance, and design. After preprocessing, the training set contains 2,282 samples (976 positive, 455 neutral, and 851 negative), while the test set includes 632 samples. This dataset is characterized by domain-specific terminology and a higher prevalence of implicit sentiment expressions.

The Restaurant dataset reflects consumer evaluations of food quality, service, and dining environment. It is larger in scale, with 3,608 training samples (2,164 positive, 637 neutral, and 807 negative) and 1,119 test samples. Compared with the Laptop dataset, its language is more colloquial, while sentiment expressions tend to be more explicit.

All samples are annotated under a three-class sentiment labeling scheme. Each review is further associated with explicit aspect terms and their corresponding sentiment polarity, providing high-quality supervision for fine-grained sentiment analysis.

3.2. Baseline models

To comprehensively evaluate the effectiveness of the proposed Emo-CoT-based LoRA fine-tuning approach, this study constructs a multi-level baseline framework, covering traditional pre-trained models, prompt-based paradigms, and various large language model fine-tuning strategies. The selected baselines are as follows:

ATAE-LSTM [10]: A seminal attention-based model for aspect-based sentiment analysis, which concatenates aspect embeddings with contextual word embeddings as input to an LSTM, and employs an attention mechanism to model aspect–context interactions.

DGEDT [11]: Constructs explicit relationships between aspect and opinion words using dependency graphs, and employs a dual-channel Transformer to encode syntactic structures and semantic information, enhancing long-range dependency modeling.

DualGCN [12]: Utilizes dual graph convolutional networks to model aspect-specific context and inter-aspect relationships, capturing fine-grained sentiment features through the joint learning of syntactic and semantic graphs.

BERT-SPC [13]: A representative approach based on BERT sentence-pair classification, where aspect terms and review text are jointly encoded to leverage deep contextual interactions for sentiment prediction.

BERT-ADA [14]: Introduces adversarial training for domain adaptation, mitigating distributional discrepancies between source and target domains through joint optimization of a domain discriminator and a sentiment classifier.

Qwen2-7B (Zero-shot): Directly applies the untuned Qwen2-7B model for zero-shot inference, aiming to evaluate the intrinsic sentiment analysis capability and general knowledge boundary of the base model.

Qwen2-7B (LoRA): A parameter-efficient fine-tuning variant using LoRA, where pre-trained weights are frozen and only low-rank matrices are trained for domain adaptation under limited labeled data.

Qwen2-7B (CoT): Employs chain-of-thought prompting to guide the model in generating explicit reasoning paths, using the structured template proposed in this study. This baseline serves to verify the superiority of the task-specific Emo-CoT structure over generic reasoning patterns.

3.3. Experimental setup and evaluation metrics

To ensure the logical rigor and generalization capability of the Emo-CoT data, GPT-4 is employed as the teacher model. Owing to its strong performance in general language understanding and complex reasoning tasks, GPT-4 can generate coherent sentiment reasoning processes that align well with human intuition. During data generation, structured prompts are used to guide output, without applying any additional instruction fine-tuning to GPT-4. This design leverages the model's inherent reasoning ability while avoiding domain-specific overfitting, thereby ensuring diversity and logical robustness in the generated training data.

The student model is Qwen2-7B, which demonstrates strong instruction-following capability at a comparable parameter scale. To address the computational bottlenecks of full fine-tuning, LoRA is adopted for parameter-efficient adaptation. Specifically, pre-trained weights are loaded in 4-bit NF4 precision to preserve statistical distributions while minimizing memory usage. The backbone parameters remain frozen throughout training, and only the injected low-rank adapters are updated via gradient descent.

For the key hyperparameters of LoRA, the rank r is set to 8 and the scaling factor α to 32. According to the LoRA scaling mechanism, the effective update magnitude applied to the weights is amplified by a factor of 4, which enhances the model's capacity to capture complex logical patterns and accelerates convergence. Under this configuration, the number of trainable parameters accounts for only approximately 0.5% of the total model parameters. Detailed settings are shown in Table 3.

Table 3. LoRA hyperparameter configuration

Parameter	Value
Rank (r)	8
Scaling Factor (α)	32
Quantization	4-bit
Learning Rate	1e-4

This study adopts two standard evaluation metrics in sentiment analysis: Accuracy (Acc) and Macro-F1 Score.

Accuracy measures the proportion of correctly predicted samples over the entire test set, reflecting overall classification performance. Its computation is given in Equation (6):

$$\mathrm{Acc} = \frac{TP+TN}{TP+TN+FP+FN} \tag{6}$$

The Macro-F1 Score addresses class imbalance by computing the arithmetic mean of F1 scores across all sentiment categories, thereby better reflecting performance on minority classes. For each class $c \in C$ (where $|C| = 3$), precision (P_c) and recall (R_c) are first computed, as shown in Equations (7) and (8):

$$P_c = \frac{TP_c}{TP_c+FP_c} \tag{7}$$

$$R_c = \frac{TP_c}{TP_c+FN_c} \tag{8}$$

Where TP_c, FP_c, FN_c denote the numbers of true positives, false positives, and false negatives for class c , respectively. The F1 score for class c is then calculated as in Equation (9):

$$F1_c = \frac{2 \cdot P_c \cdot R_c}{P_c+R_c} \tag{9}$$

Finally, the Macro-F1 Score is defined as the average of F1 scores across all classes, as shown in Equation (10):

$$\text{Macro-F1} = \frac{1}{|C|} \sum_{c \in C} F1_c \tag{10}$$

3.4. Experimental results and analysis

The experimental results are presented in Table 4. The untuned Qwen2-7B (Zero-shot) already demonstrates strong general semantic understanding, achieving F1 scores of 78.96% on the Restaurant dataset and 74.70% on the Laptop dataset. This indicates that large language models, by virtue of extensive pre-training, possess a baseline capability to capture implicit sentiment tendencies. However, zero-shot inference remains limited when dealing with domain-specific and contextually complex scenarios. After introducing LoRA-based parameter-efficient fine-tuning, model performance improves significantly, with F1 scores increasing to 80.65% and 76.46%, respectively. This gain confirms that, while preserving the general capabilities of the backbone network, injecting domain-specific distributional features via low-rank adaptation can effectively mitigate overfitting in low-resource settings and avoid catastrophic forgetting, thereby enhancing task-specific adaptability.

Table 4. Sentiment analysis results

Dataset	Restaurant		Laptop	
Evaluation metrics	Acc	F1	Acc	F1
ATAE-LSTM	77.2	-	68.5	-
DGEDT	83.9	75.1	76.8	72.3
DualGCN	84.27	78.08	78.48	74.74
BERT-SPC	84.46	79.52	78.3	74.2
BERT-ADA	87.89	-	80.23	-
Qwen2-7B (Zero-shot)	83.1	78.96	80.4	74.7
Qwen2-7B((LoRA)	84.81	80.65	81.22	76.46

Table 4. Continued

Qwen2-7B(CoT)	86.17	82.3	82.56	78.82
Qwen2-7B+ours	89.42	86.14	84.26	80.54

Building upon fine-tuning, the incorporation of reasoning mechanisms leads to stepwise performance improvements. Qwen2-7B (CoT), guided by general chain-of-thought prompting, performs step-by-step reasoning and achieves F1 improvements of 1.65% and 2.36% over the LoRA-only model on the Restaurant and Laptop datasets, respectively. This demonstrates that explicit reasoning paths help the model organize logical structures and reduce intuitive misjudgments. However, the gains from generic CoT remain limited, primarily because it lacks task-specific structural constraints. The generated reasoning trajectories tend to be loosely organized and fail to consistently capture critical logical elements in sentiment analysis, such as contrast and negation.

The proposed method achieves the best performance on both datasets. Specifically, it attains an accuracy of 89.42% and an F1 score of 86.14% on the Restaurant dataset, and an accuracy of 84.26% with an F1 score of 80.54% on the Laptop dataset. Compared with the generic CoT baseline, the proposed method improves F1 by 3.84% and 1.72%, respectively. Relative to the traditional DualGCN baseline, the improvements reach 8.06% and 5.80%, respectively.

By introducing a sentiment-aware interactive reasoning mechanism, the model is required to generate a complete reasoning trajectory prior to polarity determination. This structured constraint transforms implicit logical judgments into explicit chains of thought, enabling precise modeling of long-range dependencies and overcoming the limitations of traditional models that rely on surface-level statistical features. Combined with the LoRA strategy, the model efficiently learns domain-specific reasoning patterns while retaining general semantic understanding. This integration not only addresses the issue of missing logical structure in low-resource scenarios, but also significantly enhances robustness in handling complex sentence structures.

Overall, the experimental results provide strong evidence that the proposed chain-of-thought-based LoRA fine-tuning method not only leverages the parameter efficiency of large models, but also substantially improves the accuracy and interpretability of sentiment analysis through explicit logical guidance.

To systematically disentangle the contributions of each core component in the Emo-CoT framework and to validate the effectiveness of parameter-efficient fine-tuning and structured reasoning in fine-grained sentiment analysis, this study further conducts ablation experiments under controlled conditions. By keeping data splits, evaluation metrics, and hyperparameter settings consistent, three variant models are constructed. The results are reported in Table 5.

Table 5. Ablation study results

Dataset	Restaurant		Laptop	
Metric	Acc	F1	Acc	F1
Full Model	89.42	86.14	84.26	80.54
w/o LoRA	86.17	82.30	82.56	78.82
w/o Emo-CoT	86.85	83.45	81.10	76.12
w/o Logic-Constraint	87.90	84.58	83.45	78.95

(1) w/o LoRA: This variant removes the low-rank adaptation module, using only the pre-trained base model combined with the proposed Emo-CoT prompts. The aim is to verify the necessity of domain adaptation. Results show that even with carefully designed Emo-CoT prompts, performance declines without

fine-tuning. On the Restaurant dataset, F1 drops from 86.14% to 82.30%, and on the Laptop dataset, from 80.54% to 78.82%. Although large models possess strong general reasoning capabilities, they struggle to capture domain-specific conventions without distributional alignment. The LoRA module effectively corrects prediction biases, demonstrating the indispensability of parameter-efficient fine-tuning in low-resource scenarios.

(2) w/o Emo-CoT: This variant retains LoRA fine-tuning but replaces the structured Emo-CoT protocol with generic chain-of-thought prompting. The results show that the proposed sentiment-specific reasoning framework yields more robust improvements. The full model outperforms the generic CoT baseline by 2.69% and 4.42% in F1 on the Restaurant and Laptop datasets, respectively. While generic CoT encourages stepwise reasoning, its trajectories lack structured enforcement of critical task-specific logic, such as contrast handling, negation processing, and aspect alignment.

(3) w/o Logic-Constraint: This variant retains both LoRA and chain-of-thought reasoning, but removes the enforced logical constraints, allowing free-form reasoning generation. Experimental results indicate a performance drop of approximately 1.5%–2.0% in F1. This demonstrates that stepwise reasoning alone is insufficient for handling the complexity of fine-grained sentiment analysis; strong logical constraints tailored to the task are essential. Without such constraints, the model tends to generate generalized descriptive text rather than rigorous logical reasoning, leading to logical inconsistencies or hallucinations when processing complex sentences.

The ablation results confirm that the explicitly designed logical nodes in Emo-CoT successfully transform implicit black-box decisions into interpretable white-box reasoning processes. This ensures the rigor of the model's decision-making and maintains higher robustness in complex contextual scenarios.

4. Conclusion

This chapter focuses on aspect-based sentiment analysis and systematically develops the Emo-CoT framework, completing a full pipeline from theoretical formulation to empirical evaluation. To address the limitations of existing methods in handling complex semantic and logical structures, the proposed approach formalizes a task paradigm integrated with chain-of-thought reasoning. It further designs a model architecture incorporating domain knowledge-enhanced templates and aspect alignment mechanisms, aiming to guide large language models through an explicit reasoning process of "first identify, then reason, and finally determine."

Extensive experimental evaluations demonstrate that Emo-CoT consistently outperforms traditional graph neural network baselines and generative models across multiple key metrics. In particular, it exhibits stronger robustness and generalization capability in challenging scenarios involving multi-opinion conflicts and long-range dependencies. The ablation study further quantifies the contribution of the chain-of-thought guidance strategy, confirming the necessity of explicit reasoning pathways for performance improvement. In addition, qualitative case analyses provide deeper insights into the model's internal decision-making process. By dynamically adjusting attention weights and identifying logical relations such as contrast and concession, the proposed method effectively mitigates misclassification risks caused by static syntactic parsing and simplistic sentiment aggregation strategies.

References

[1] Liu, B. (2012). Sentiment analysis and opinion mining. *Synthesis Lectures on Human Language Technologies, 5*(1), 1–167. https: //doi.org/10.2200/s00416ed1v01y201204hlt016

[2] Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. *In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (pp. 4171–4186)*. Association for Computational Linguistics. https: //doi.org/10.18653/v1/N19-1423

[3] Brown, T., Mann, B., & Ryder, N. (2020). Language models are few-shot learners. *In Advances in Neural Information Processing Systems (Vol. 33*, pp. 1877–1901). Curran Associates.

[4] Sun, C., Huang, L., & Qiu, X. (2019). Utilizing BERT for aspect-based sentiment analysis via constructing auxiliary sentence. *In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1* (Long and Short Papers) (pp. 380–385). Association for Computational Linguistics.

[5] Geirhos, R., Jacobsen, J. H., & Michaelis, C. (2020). Shortcut learning in deep neural networks. *Nature Machine Intelligence, 2*(11), 665–673. https: //doi.org/10.1038/s42256-020-00257-z

[6] Wei, J., Tay, Y., & Bommasani, R. (2022). Emergent abilities of large language models. arXiv. https: //doi.org/10.48550/arXiv.2206.07682

[7] Ji, Z., Lee, N., & Frieske, R. (2023). Survey of hallucination in natural language generation. *ACM Computing Surveys, 55*(12), Article 248, 1–38. https: //doi.org/10.1145/3571730

[8] Loke, B. S. Y., Quadri, F., & Vivanco, G. (2025). Overcoming catastrophic forgetting in neural networks. arXiv. https: //doi.org/10.48550/arXiv.2507.10485

[9] Pontiki, M., Galanis, D., & Pavlopoulos, J. (2014). SemEval-2014 task 4: Aspect based sentiment analysis. *In Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014)* (pp. 27–35). Association for Computational Linguistics.

[10] Wang, Y., Huang, M., Zhu, X., & Zhao, L. (2016). Attention-based LSTM for aspect-level sentiment classification. *In Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing* (pp. 606–615). Association for Computational Linguistics.

[11] Tang, Y., Huang, H., & Wang, X. (2020). Dependency graph enhanced dual-transformer network for aspect-based sentiment analysis. *IEEE/ACM Transactions on Audio, Speech, and Language Processing, 28*, 779–791. https: //doi.org/10.1109/TASLP.2020.2966751

[12] Li, R., Chen, H., & Feng, F. (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)* (pp. 6319–6329). Association for Computational Linguistics.

[13] Song, Y., Wang, J., & Jiang, T. (2019). Attentional encoder network for targeted sentiment classification. arXiv. https: //doi.org/10.48550/arXiv.1902.09314

[14] Rietzler, A., Stabinger, S., Opitz, P., & Schütze, H. (2020). Adapt or get left behind: Domain adaptation through BERT language model finetuning for aspect-target sentiment classification. *In Proceedings of the 12th Language Resources and Evaluation Conference* (pp. 4933–4941). European Language Resources Association.