Article

Fusion of Local and Global Context in Large Language Models for Text Classification

Ran Hao , Xin Hu , Jiasen Zheng , Chong Peng , Junjiang Lin *

Article

Fusion of Local and Global Context in Large Language Models for Text Classification

Ran Hao [1], Xin Hu [2], Jiasen Zheng [3], Chong Peng [4] and Junjiang Lin [5,*]

[1] University of North Carolina at Chapel Hill, Chapel Hill, USA

[2] Hofstra University, Hempstead, USA

[3] Northwestern University, Evanston, USA

[4] Carnegie Mellon University, Pittsburgh, USA

[5] University of Toronto, Toronto, Canada

* Correspondence: junjianglin@cs.toronto.edu

Abstract

This study addresses the problem of insufficient context capture in text classification and proposes a large language model method enhanced with contextual mechanisms. At the input layer, raw text is transformed into vector sequences that incorporate both semantic and sequential features through the combination of embedding representation and positional encoding. A context encoder based on self-attention is then introduced to capture global dependencies within the sequence. At the same time, a context gating unit is designed to achieve dynamic fusion of local and global information, which preserves fine-grained features while strengthening overall contextual consistency. Furthermore, a global context aggregation module integrates semantic information across sentences or paragraphs, enhancing the model's ability to represent long texts and implicit semantics. In the output stage, sentence-level pooling is used to generate a unified representation, followed by a classification head to complete label prediction. To validate the effectiveness of the method, comparative experiments were conducted on a public news text classification dataset. The results show that the proposed method outperforms traditional deep learning models and mainstream large-model baselines in terms of accuracy, precision, recall, and F1-score. It maintains a more stable classification performance when dealing with semantic ambiguity and topic shifts. In addition, sensitivity experiments on hidden dimension settings demonstrate that moderate model capacity significantly improves performance, while excessive complexity may introduce redundant representations and slight overfitting. This study demonstrates the practical value of context enhancement mechanisms in large language models and provides a more robust and effective solution for text classification tasks.

Keywords: context enhancement; large language models; text classification; semantic modeling

1. Introduction

In the era of information explosion and accelerated digitalization, text data has become the core carrier of social operation and knowledge dissemination [1]. From news reports and social media to academic publishing and business documents, massive amounts of text are generated every second. How to quickly and accurately extract valuable knowledge from such complex and redundant text has long been an important issue in artificial intelligence and natural language processing. Text classification, as a fundamental task, plays the key role of transforming unstructured language into structured semantics. It supports applications such as opinion analysis, content recommendation, and sentiment computation, and also plays a vital role in sensitive fields such as healthcare, finance, and education. With the rise of large-scale language models, text classification methods have made unprecedented progress. However, despite the continuous increase in model scale and parameters,

how to fully utilize contextual information and enhance the ability to capture semantic details remains a pressing challenge[2].

The core characteristic of language lies in its contextual dependence. Words and sentences do not exist in isolation but are embedded in multilayered contexts. Traditional statistical or shallow semantic models often fail to effectively capture this complexity, leading to fragmented semantics or ambiguous contexts. Even current large language models, while demonstrating strong semantic understanding and generation capabilities, may still face classification bias when dealing with long texts, cross-paragraph information, or implicit semantic reasoning[3]. In practice, this insufficiency may appear as misclassification of ironic expressions in sentiment analysis, blurred categorization of cross-domain topics in news classification, or semantic deviations in legal and medical texts due to missing context. These problems not only reduce accuracy but also pose risks in real-world decision-making. Therefore, integrating context enhancement mechanisms into large language models to build classification methods with stronger global understanding and local sensitivity holds significant research value[4].

From the perspective of technological evolution, the development of text classification reflects continuous shifts in natural language processing methodologies[5]. From early rule-based methods relying on manual features, to sequence modeling driven by deep learning, and now to end-to-end semantic learning based on large language models, the central focus has always been on "how to better capture context." The introduction of context enhancement mechanisms is not only a renewed recognition of the essence of language but also an extension of the capability boundaries of large models. By introducing explicit or implicit contextual constraints, models can establish closer associations between words, sentences, and paragraphs, thereby improving sensitivity to semantic shifts and contextual transitions[6]. The value of this mechanism goes beyond improving classification accuracy. More importantly, it strengthens model robustness and adaptability in open environments, ensuring consistent performance in diverse and dynamic text scenarios[7].

Meanwhile, the application scenarios of text classification are continuously expanding, and the demands on model capabilities are rising. In social media, texts often carry strong emotions and informal expressions, requiring models to perform fine-grained understanding under complex contexts. In cross-lingual and cross-domain applications, contextual differences become more significant, demanding models with cross-cultural and cross-context transfer capabilities. In high-risk industries such as medical diagnosis, financial risk control, and legal judgment, classification results may directly affect human safety and economic decisions [8–10]. In these scenarios, simply scaling up model parameters is not sufficient. Context enhancement mechanisms must be employed to provide more refined semantic modeling and stronger robustness. This not only improves model capability but also represents a critical step toward trustworthy, explainable, and controllable artificial intelligence.

In conclusion, the integration of contextual mechanisms into large language model text classification methods is not only a natural progression in technological development but also a crucial response to practical demands. These advancements address the fundamental challenge of contextual complexity in language, enabling large models to achieve deeper semantic understanding. Moreover, they provide reliable support for high-value application domains, making this research direction not only academically important but also highly practical in advancing the application of artificial intelligence across various societal sectors.

2. Proposed Approach

In the overall method design, the first step is to convert the input raw text into a vectorized representation suitable for processing by a large language model. Let the input text sequence be $T = \{w_1, w_2, ..., w_n\}$, where n is the length of the text. The word is mapped to a continuous space through the embedding layer, resulting in a representation matrix:

$$X = [e(w_1), e(w_2), ..., e(w_n)] \in R^{n \times d} \qquad (1)$$

Where d represents the embedding dimension. To capture the position information, the position encoding P is introduced to obtain the initial input representation:

$$H_0 = X + P \qquad (2)$$

This process provides a basis for modeling subsequent context enhancement mechanisms and ensures that semantic and sequential features can be preserved simultaneously. The overall model architecture is shown in Figure 1.
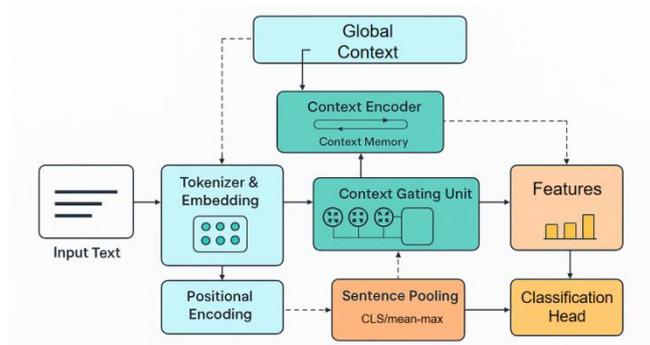

Figure 1. Overall model architecture.

In the context enhancement mechanism, the self-attention mechanism is first used to model the dependencies within the sequence [11,12]. Let the input be $H \in R^{n \times d}$, and the query, key, and value are obtained through linear mapping:

$$Q = HW_Q, K = HW_K, V = HW_V \qquad (3)$$

where $W_Q, W_K, W_V \in R^{d \times d_k}$ is the parameter matrix. The attention weight is calculated by scaling the dot product:

$$Attention(Q,K,V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \qquad (4)$$

This mechanism can dynamically allocate the dependency strength between different positions, thereby achieving weighted modeling of contextual information.

To further strengthen global semantic consistency, this method introduces a contextual gating unit after self-attention [13]. Specifically, the local semantic representation h_i interacts with the contextual representation c_i, and an enhanced representation is obtained through the gating mechanism:

$$g_i = \sigma(W_g[h_i;c_i] + b_g) \qquad (5)$$

$$h_i = g_i \otimes h_i + (1 - g_i) \otimes c_i \qquad (6)$$

where $[h_i;c_i]$ represents vector concatenation, $\sigma(\cdot)$ is the Sigmoid function, and $\otimes$ represents element-by-element multiplication. This mechanism dynamically balances local and contextual information, improving the model's ability to uniformly model both semantic details and global themes.

Finally, after obtaining the text representation sequence $H = \{\tilde{h}_1, \tilde{h}_2, ..., \tilde{h}_n\}$ after fusion context enhancement, classification prediction is required. To obtain sentence-level representation, this study uses a pooling operation:

$$z = Pool(\tilde{H}) \qquad (7)$$

where $Pool(\cdot)$ can be mean pooling or maximum pooling. It is then mapped to the category space through a fully connected layer:

$$y = \text{softmax}(W_o z + b_o) \qquad (8)$$

where $W_o \in R^{d \times C}$ and C are the number of categories. The final output vector $y \in R^C$ represents the model's predicted probability distribution for each category, thus completing the target task of text classification.

3. Performance Evaluation

3.1. Dataset

This study adopts the AG News dataset as the benchmark corpus for text classification. The dataset is composed of English news titles and short descriptions. It covers four thematic categories: World, Sports, Business, and Sci/Tech. The sources are diverse, the time span is long, and the domains differ significantly. These characteristics make it representative of general news scenarios for topic classification. Most samples consist of one to several short sentences, containing many entities and keywords. The structure is compact and the information density is high, which makes it suitable for testing semantic discrimination within a limited context.

AG News provides a standard train and test split. The training set contains 120,000 samples, and the test set contains 7,600 samples. The four classes are approximately balanced, which reduces bias caused by class imbalance. For model tuning and early stopping, 10% of the training set can be used as a validation set. Texts can be normalized to a consistent case while retaining important punctuation. This reduces irrelevant noise while preserving semantic cues. Each data entry includes two fields (title and description). They can be concatenated into a single sequence for unified encoding. To fit the inference length of large language models, the maximum sequence length can be set to a moderate range, with tail truncation applied to the few very long samples.

The choice of this dataset is based on several reasons. The topics are clearly defined, and the label system is simple. It allows evaluation of global topic recognition and also exposes challenges such as cross-domain vocabulary overlap and semantic ambiguity. At the same time, the title and description form a lightweight contextual link. This aligns naturally with the idea of context enhancement and enables validation of how explicit context modeling in short texts can improve semantic aggregation and class alignment. Due to its wide use and clear task definition, AG News provides a stable and reproducible task setting and a reliable baseline for the methodology.

3.2. Experimental Results

This paper first conducts a comparative experiment, and the experimental results are shown in Table 1.

Table 1. Comparative experimental results.

Method	Acc	Precision	Recall	F1-Score
BiLSTM[14]	91.42	91.10	90.75	90.92
MLP[15]	89.85	89.20	89.46	89.33
1DCNN[16]	92.30	92.05	91.67	91.86
Transformer [17]	94.12	94.01	93.45	93.73
Ours	95.78	95.40	95.06	95.23

The results show a stepwise improvement that is positively correlated with the model's ability to capture context. From MLP to BiLSTM and 1D CNN, then to Transformer, and finally to the method with context enhancement, all four metrics increase steadily. Compared with the baseline Transformer, adding context enhancement improves accuracy by 1.66 percentage points (94.12 to

95.78) and F1 by 1.50 percentage points (93.73 to 95.23). The error rate decreases from 5.88% to 4.22%, which is about a 28% relative reduction. This indicates that adding explicit context modeling on top of large language models not only increases overall accuracy but also significantly reduces errors in boundary and ambiguous samples.

From the perspective of balance between precision and recall, the context-enhanced method maintains better symmetry (Precision = 95.40, Recall = 95.06), while the Transformer shows a slight lack of recall (94.01 vs 93.45). This means that in cases of overlapping classes, irony, or topic shift, the improved global context representation reduces missed detections (recall rises by 1.61), without paying the cost of many false positives (precision only increases by 1.39). For text classification, this balanced state is especially important. When class boundaries are driven by context, gated fusion and global aggregation suppress the noise amplification effect of local fragments, keeping decisions consistent across categories.

A horizontal comparison with traditional structures shows that 1D CNN and BiLSTM can capture local n-gram patterns and sequential dependencies, but they have limited sensitivity to cross-sentence cues and paragraph-level shifts. This leads to F1 scores lower than Transformer by 1.87 and 2.81 percentage points, respectively. The weakness of MLP is most obvious, showing that without temporal or attention-based modeling, text representations lack robustness to contextual changes. The trend suggests that performance gains are highly related to the expansion of the "effective context window." As models evolve from local convolution and sequential memory to global attention and then to explicit context enhancement, class separability and boundary stability improve together.

From a methodological standpoint, even on top of the Transformer, which already models strong global dependencies, further context enhancement still yields significant improvements. This indicates that the bottleneck of semantic understanding extends beyond the range of attention and lies in how to effectively aggregate and route contexts at various levels in a structured and dynamic manner. The results suggest that designing the interface between global aggregation, gated fusion, and the classifier as an adjustable context channel can enhance calibration and robustness while preserving discriminative power. In practical applications, this approach ensures stable performance in texts with diverse topics, sparse expressions, or domain transfer, offering reliable semantic support for downstream tasks that are sensitive to risk.

This paper also presents an experiment on the sensitivity of the hidden layer dimension to the experimental results, and the experimental results are shown in Figure 2.
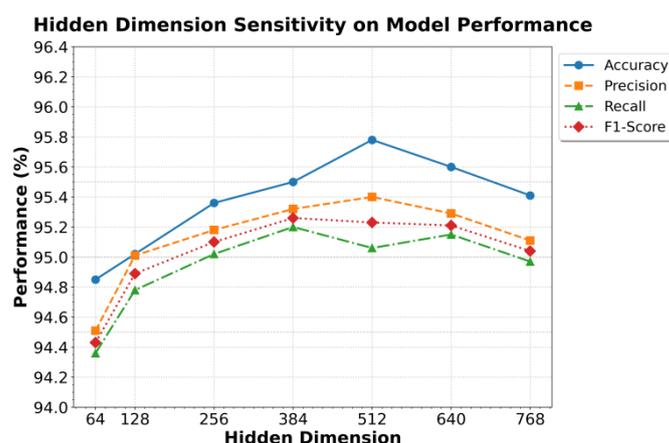

Figure 2. Sensitivity experiment of the hidden layer dimension to experimental results.

From the experimental results, it can be observed that as the hidden dimension increases, the model shows an overall trend of rising performance followed by stabilization and even a slight decline. When the hidden dimension grows from 64 to 512, accuracy, precision, recall, and F1-score all improve significantly, with the best performance reached around 512. This indicates that a larger

hidden capacity can capture richer contextual semantic features and demonstrate stronger expressive power in combining global and local information, which leads to improved performance.

However, when the hidden dimension is further increased to 640 and 768, the model shows fluctuations and decline, especially in recall and F1-score. This suggests that overly large hidden dimensions may introduce redundant representations and raise the risk of overfitting, which weakens the advantage of context enhancement in generalization scenarios. Overall, the results confirm the effectiveness of context enhancement under moderate model capacity. At the same time, they highlight the need to balance model complexity and semantic modeling ability in practical applications to achieve more stable classification performance.

4. Conclusion

This study focuses on a text classification method for large language models enhanced with contextual mechanisms. It systematically analyzes the key role of context in semantic modeling and proposes an improved framework that captures both local details and global dependencies. Comparative and sensitivity experiments show that the method achieves comprehensive improvements in accuracy, precision, recall, and F1-score. It demonstrates clear advantages in handling semantic ambiguity, topic shifts, and long-text reasoning. These findings confirm the applicability and effectiveness of context enhancement in large language models and provide new insights for text classification methods, making them closer to the real contextual characteristics of natural language.

Future research may further explore the transferability of this method across domains and languages, while extending it to meet practical application needs. For example, in industries such as healthcare, finance, and law, where semantic precision is critical, the method has potential value for risk prediction, information retrieval, and decision support. In applications such as opinion monitoring, content recommendation, and intelligent customer service, it can also improve the model's sensitivity and stability to contextual changes. Overall, this study not only advances the technical development of text classification methods but also provides more interpretable and reliable support for the application of artificial intelligence in key industries.

References

1. Edwards A, Camacho-Collados J. Language models for text classification: is in-context learning enough?(2024)[J]. arXiv preprint arXiv:2403.17661, 2024.
2. Sun X, Li X, Li J, et al. Text classification via large language models[J]. arXiv preprint arXiv:2305.08377, 2023.
3. Sen J, Pandey R, Waghela H. Context-Enhanced Contrastive Search for Improved LLM Text Generation[J]. arXiv preprint arXiv:2504.21020, 2025.
4. Abd-Elaziz M M, El-Rashidy N, Abou Elfetouh A, et al. Position-context additive transformer-based model for classifying text data on social media[J]. Scientific Reports, 2025, 15(1): 8085.
5. Karthick R. Context-Aware Topic Modeling and Intelligent Text Extraction Using Transformer-Based Architectures[J]. Available at SSRN 5275391, 2025.
6. Li C, Xie Z, Wang H. Short Text Classification Based on Enhanced Word Embedding and Hybrid Neural Networks[J]. Applied Sciences, 2025, 15(9): 5102.
7. Luo D, Zhang C, Zhang Y, et al. CrossTune: Black-box few-shot classification with label enhancement[J]. arXiv preprint arXiv:2403.12468, 2024.
8. X. Yan, J. Du, X. Li, X. Wang, X. Sun, P. Li and H. Zheng, "A Hierarchical Feature Fusion and Dynamic Collaboration Framework for Robust Small Target Detection," IEEE Access, vol. 13, pp. 123456–123467, 2025.
9. Peng, S., Zhang, X., Zhou, L., and Wang, P., "YOLO-CBD: Classroom Behavior Detection Method Based on Behavior Feature Extraction and Aggregation," Sensors, vol. 25, no. 10, p. 3073, 2025.

10. J. Wei, Y. Liu, X. Huang, X. Zhang, W. Liu and X. Yan, "Self-Supervised Graph Neural Networks for Enhanced Feature Extraction in Heterogeneous Information Networks", 2024 5th International Conference on Machine Learning and Computer Application (ICMLCA), pp. 272-276, 2024.

11. Gong, M., Deng, Y., Qi, N., Zou, Y., Xue, Z., and Zi, Y., "Structure-Learnable Adapter Fine-Tuning for Parameter-Efficient Large Language Models," arXiv preprint arXiv:2509.03057, 2025.

12. Li, X. T., Zhang, X. P., Mao, D. P., and Sun, J. H., "Adaptive robust control over high-performance VCM-FSM," 2017.

13. Li, Y., Han, S., Wang, S., Wang, M., and Meng, R., "Collaborative Evolution of Intelligent Agents in Large-Scale Microservice Systems," arXiv preprint arXiv:2508.20508, 2025.

14. Enamoto L, Santos A R A S, Maia R, et al. Multi-label legal text classification with BiLSTM and attention[J]. International Journal of Computer Applications in Technology, 2022, 68(4): 369-378.

15. Galke L, Scherp A. Bag-of-words vs. graph vs. sequence in text classification: Questioning the necessity of text-graphs and the surprising strength of a wide MLP[J]. arXiv preprint arXiv:2109.03777, 2021.

16. Cacciari I, Ranfagni A. Hands-On Fundamentals of 1D Convolutional Neural Networks—A Tutorial for Beginner Users[J]. Applied Sciences (2076-3417), 2024, 14(18).

17. Karl F, Scherp A. Transformers are short text classifiers: a study of inductive short text classifiers on benchmarks and real-world datasets. Corr abs/2211.16878 (2022)[J]. URL: https://doi. org/10.48550/arXiv, 2211.