Transactions on Computational and Scientific Methods | Vo. 4, No. 10, 2024
ISSN: 2998-8780
https://pspress.org/index.php/tcsm
Pinnacle Science Press

Knowledge Graph-Augmented Semantic Fusion for Large Language Model-Based Long-Text Classification

Zhizhen Zhu[1*]

[1]Columbia University, New York, USA

*Corresponding author: zz2978@columbia.edu

Abstract: This paper proposes a text classification method that integrates large language models with knowledge graphs to address the problems of semantic redundancy, insufficient contextual dependency, and limited external knowledge utilization in long-text classification. The method first employs the contextual modeling ability of large language models to obtain global semantic representations of texts, and uses multi-layer self-attention mechanisms to capture cross-paragraph dependencies, thereby alleviating the semantic dispersion and redundancy issues found in traditional approaches. It then introduces a knowledge graph to structurally model entities and relations in the text, and applies embedding and alignment strategies to deeply fuse knowledge graph information with text representations, allowing the model to capture semantic information while gaining explicit knowledge constraints, which enhances the completeness and accuracy of semantic understanding. Furthermore, a series of sensitivity experiments is conducted from multiple perspectives, including hyperparameters, environmental conditions, and data characteristics, to systematically analyze the stability and robustness of the method and verify its applicability and consistency in complex contexts. Experimental results show that the proposed method outperforms comparison approaches on multiple metrics such as AUC, ACC, F1-Score, and Precision, demonstrating the advantages of combining large language models with knowledge graphs in long-text classification. In summary, this study not only achieves an organic integration of semantic modeling and knowledge modeling at the theoretical level but also provides a feasible technical pathway for effective classification of texts.

Keywords: text classification; knowledge graph; large language model; semantic fusion

1. Introduction

In the context of information explosion and the rapid growth of multi-source heterogeneous data, text classification has become a core task in natural language processing and shows increasing importance. With the emergence of massive textual data from social media, news reports, academic literature, and government or enterprise archives, accurately identifying and organizing information in complex contexts has become a key prerequisite for knowledge discovery and intelligent applications. However, traditional text classification methods often rely on shallow features or static semantic representations, which are limited in capturing hidden relationships and deep semantic structures. The rise of large language models brings new momentum to text classification. Their strong semantic modeling and generative abilities help address challenges in long-text processing, contextual association, and semantic disambiguation. Yet redundancy, ambiguity, and the lack of cross-domain knowledge in texts still restrict further improvements in classification performance[1].

Against this background, the introduction of knowledge graphs provides stronger external knowledge support for text classification. Knowledge graphs are built around entities, relations, and attributes, forming multi-level semantic networks. This allows language models to acquire more precise background knowledge and semantic constraints during training and inference. Such structured knowledge not only compensates for the limitations of language models in factual reasoning and domain expertise but also enhances their ability to understand implicit semantics and cross-sentence dependencies. By deeply integrating textual features with knowledge graph representations, classification models can achieve multi-granularity understanding from lexical to conceptual levels, thus maintaining robustness and generalization in complex contexts[2].

Moreover, the combination of large language models and knowledge graphs is not a simple technical overlay but rather a complement between two modeling paradigms. Large language models learn semantic distributions in a data-driven way and show strong contextual modeling capabilities. Knowledge graphs, on the other hand, provide explicit symbolic representations that support logical reasoning and semantic constraints[3]. Their integration enables text classification to combine surface-level linguistic flexibility with deeper knowledge-level reasoning. This advantage is not only valuable for general text processing but also shows strong application potential in fields such as finance, healthcare, law, and education, helping achieve more precise domain-specific classification and knowledge management[4].

At the same time, knowledge-enhanced large language model classification methods are also significant for addressing problems of information asymmetry and semantic drift. In dynamic environments, text data often evolves with changing semantics and the emergence of new knowledge. Traditional classifiers tend to suffer from performance degradation when facing new domains and concepts. The dynamic update mechanism of knowledge graphs, together with the transfer capabilities of language models, enables classification systems to adapt continuously and update themselves. This feature ensures the timeliness and stability of classification and lays the foundation for building artificial intelligence systems with lifelong learning and intelligent evolution.

In conclusion, research on text classification that integrates knowledge graphs with large language models holds both theoretical and practical value. From an academic perspective, this direction advances the deep integration of semantic modeling and knowledge modeling, expanding the boundaries of natural language processing research. From an application perspective, it provides strong support for large-scale information organization, intelligent retrieval, automated decision-making, and cross-domain knowledge services. As related technologies continue to develop, this research is expected to become a key step in driving intelligent systems toward knowledge-driven and reasoning-enhanced development, with significant implications for building more efficient and trustworthy artificial intelligence ecosystems.

2. Related work

In the development of text classification research, early methods mainly relied on shallow features and statistical modeling, such as bag-of-words models, TF-IDF weighting, and traditional machine learning classifiers. These approaches achieved certain results in small-scale and well-structured text tasks. However, when dealing with texts, cross-domain corpora, and semantically complex scenarios, they often suffer from sparse features, semantic loss, and insufficient context dependency. With the rise of deep learning, convolutional neural networks and recurrent neural networks became mainstream choices[5]. These models could capture local patterns and contextual information in sequence modeling, which helped alleviate the shortcomings of traditional methods. Nevertheless, they still showed clear limitations in handling long-distance dependencies and maintaining global semantic consistency.

In recent years, the emergence of large language models has offered a new solution for text classification. Models based on the self-attention mechanism show significant advantages in semantic representation, contextual modeling, and capturing cross-paragraph dependencies. They have greatly improved accuracy

and generalization. Through large-scale pre-training and fine-tuning strategies for downstream tasks, such models demonstrate strong adaptability across multiple domains. However, despite their advantages in semantic representation, they are still affected by limitations in training data when handling knowledge-intensive texts and domain-specific tasks. This often leads to factual errors or insufficient domain expertise, which restricts their practical application[6].

To address this limitation, knowledge graphs have gradually become an important support in text classification research. By explicitly modeling entities and relations, knowledge graphs provide structured external knowledge that enhances reasoning ability and interpretability. In existing studies, knowledge graphs have been applied to entity recognition, relation extraction, and knowledge-enhanced classification tasks, and have shown performance improvements. By aligning and integrating textual representations with knowledge representations, models can use background knowledge as constraints while capturing contextual semantics. This improves the understanding of ambiguous texts and the modeling of complex semantics. This approach is particularly suitable for tasks that require precise knowledge support, such as financial news, medical literature, and legal documents[7].

On this basis, the integration of large language models with knowledge graphs has become a new research focus. Current explorations indicate that such integration enables unified modeling of semantic understanding and knowledge reasoning in an end-to-end manner. This not only improves classification accuracy but also provides advantages in interpretability and robustness. Methods in this direction introduce knowledge retrieval, graph neural networks, or multimodal fusion mechanisms, allowing models to continuously update and use external knowledge in dynamic environments. This helps overcome the performance bottlenecks caused by limited single-source corpora. Overall, related studies have laid a theoretical and methodological foundation for knowledge-driven intelligent text classification. They also point to directions for building more complex cross-modal and multi-domain intelligent systems.

Recent studies also indicate that knowledge-structured reasoning and verification-oriented LLM workflows can strengthen the reliability of knowledge-enhanced classification. Explainable causal discovery based on financial knowledge graphs demonstrates how heterogeneous entities, relations, and interpretable evidence paths can be organized through graph-structured domain knowledge, offering a useful perspective for constraining text representations and reducing semantic ambiguity in classification decisions[8]. Trustworthy LLM-agent-oriented enterprise ETL orchestration further shows that language-model decision processes can benefit from explicit execution-state modeling and neural verification, which supports traceable evidence, stable reasoning, and controllable output validation in long-text classification pipelines[9]. These studies support the construction of a classification framework in which knowledge graph semantics, entity-relation structures, and reliability-oriented reasoning jointly improve interpretability and robustness.

3. Proposed Approach

This study proposes a method that centers on large language models and incorporates knowledge graphs to enhance text classification. Let the input text sequence be:

$$X = \{x_1, \ldots, x_n\}$$

where x_i denotes the i-th token or subword. Through an embedding layer, discrete symbols are mapped into continuous representations, yielding the initial matrix:

$$H^{(0)} = Embed(X) \in \mathbb{R}^{n \times d}$$

where d is the embedding dimension. Then, multi-layer self-attention is employed to capture long-range dependencies. The self-attention is defined as:

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

where

$$Q = H^{(l)}W_Q,\ K = H^{(l)}W_K,\ V = H^{(l)}W_V$$

and the updated representation is obtained via residual connection and layer normalization:

$$H^{(l+1)} = LayerNorm\left(H^{(l)} + Attention(Q,K,V)\right)$$

This process ensures that the model captures both local and global semantic information. The model architecture is shown in Figure 1.

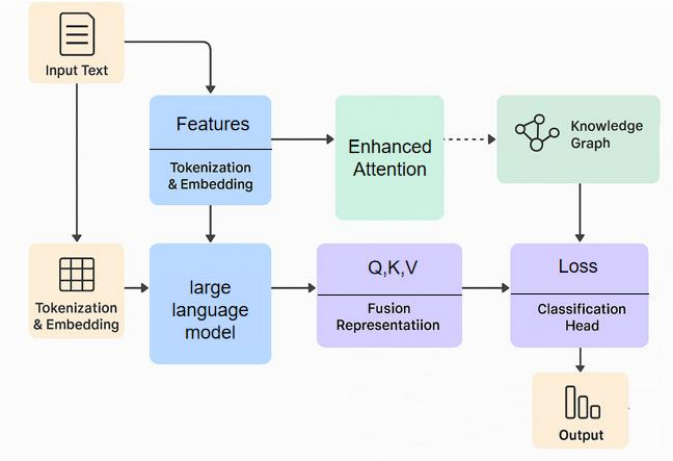

Figure 1. Overall model architecture

To incorporate external knowledge, a knowledge graph with entity set e and relation set r is embedded. Each triple (h,r,t) is represented as:

$$f(h,r,t) = \|e_h + r - e_t\|^2$$

where e_h, e_t are entity embeddings and r is a relation embedding. The knowledge embedding is concatenated with the text representation and projected into a unified space:

$$Z = Concat\left(H^{(L)}, E\right)W_z$$

where $H^{(L)}$ is the output of the last language model layer, E is the knowledge embedding matrix, and W_z is the projection parameter matrix. This design ensures the deep fusion of textual semantics and knowledge representations.

In the classification stage, the fused representation is pooled to obtain a global semantic vector:

$$h = Pooling(Z)$$

which is mapped to the label space through a fully connected layer, yielding the final probability distribution:

$$\hat{y} = softmax(W\,h + b)$$

Where W and b are the classification layer parameters. Through this method, the large language model can capture contextual dependencies while leveraging explicit knowledge in the knowledge graph, achieving more accurate and robust text classification.

4. Experiment result

4.1 dataset

This study uses the DBpedia Ontology Classification dataset as the benchmark corpus. The dataset is derived from the titles and abstracts of encyclopedia entries and is annotated according to the top-level categories of the DBpedia ontology. It covers 14 general categories with a balanced distribution. The official split usually consists of 560,000 training samples and 70,000 test samples, with 40,000 and 5,000 samples for each category, respectively. The texts are primarily in English and cover a wide range of topics, including institutions, locations, works, events, characters, organizations, and infrastructure. The dataset shows good generality and transferability, making it suitable as a foundation for general text classification tasks.

The basic structure of each sample consists of two parts, a short title and a brief abstract. The original text retains identifiers (URIs) linked to DBpedia resources, which enables natural alignment with external knowledge graphs. The category labels are taken from the top-level classes of the ontology, which have clear semantic boundaries and low ambiguity. Common preprocessing steps include removing control characters and extra spaces, unifying case, normalizing punctuation, minimizing HTML traces, and limiting the maximum length of very long samples. For knowledge enhancement, entity recognition and linking can be applied after cleaning, allowing retrieval of entity attributes and relations from the knowledge graph and construction of lightweight adjacency subgraphs for each sample.

The choice of this dataset is based on its large scale, clear categories, standardized annotations, and strong connection with knowledge graphs, which make it convenient for reproduction and comparison. With the support of DBpedia's ontology and entity index, structured knowledge can be introduced without altering the original text distribution. This improves semantic consistency and interpretability. At the same time, the dataset's open availability and stable official split reduce the noise and partition differences that may affect results, ensuring comparability and generalizability of methodological research in general scenarios.

4.2 Experimental Results

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

Table 1. Comparative experimental results

Method	AUC	ACC	F1-Score	Precision
Teleclass[10]	0.882	0.826	0.812	0.835
Dylas[11]	0.901	0.842	0.831	0.844
ULMFiT[12]	0.927	0.869	0.861	0.874
Tasneef[13]	0.938	0.883	0.874	0.889
Ours	0.961	0.908	0.901	0.915

From the comparison results in Table 1, it can be seen that different methods show significant differences across the four evaluation metrics. Traditional approaches such as Teleclass and Dylas provide a certain level of classification ability, but their overall performance remains limited. They are especially weak in capturing global dependencies of texts and complex semantic structures. This indicates that relying only on shallow features or limited context modeling methods is insufficient to handle semantic redundancy and cross-sentence associations. Stronger models are therefore required to improve overall performance.

At the deep model level, ULMFiT and Tasneef demonstrate clear advantages, with notable improvements in both AUC and ACC. This shows that pre-trained language representations and transfer learning strategies can effectively enhance the generalization ability of models. Such methods not only learn semantic distributions from large-scale corpora but also adapt to target data through task-specific fine-tuning. As a result, they achieve high classification accuracy in complex textual contexts. This confirms that pre-training and fine-tuning mechanisms bring structural breakthroughs for text classification.

It is worth noting that knowledge-enhanced methods further drive performance improvements. Compared with other approaches, the proposed framework achieves the best results across all four metrics, with particularly strong performance in F1-Score and Precision. This indicates that the model can maintain recall while effectively reducing false positives. Its advantage lies in combining the contextual modeling capacity of large language models with the explicit semantic constraints of knowledge graphs. This balance between linguistic flexibility and knowledge accuracy enables more robust text understanding and classification.

Overall, the comparison highlights the importance of large language models and knowledge-enhanced mechanisms in text classification. Traditional methods remain of reference value but show increasing limitations when applied to texts, domain-specific texts, and complex semantic relations. Pre-training methods compensate for deficiencies in semantic representation, while knowledge-driven fusion strategies push classification performance to higher levels. This trend confirms the necessity of dual modeling with semantics and knowledge and provides a solid theoretical and practical foundation for future research.

This paper also presents an experiment on the sensitivity of batch size to ACC, and the experimental results are shown in Figure 2.

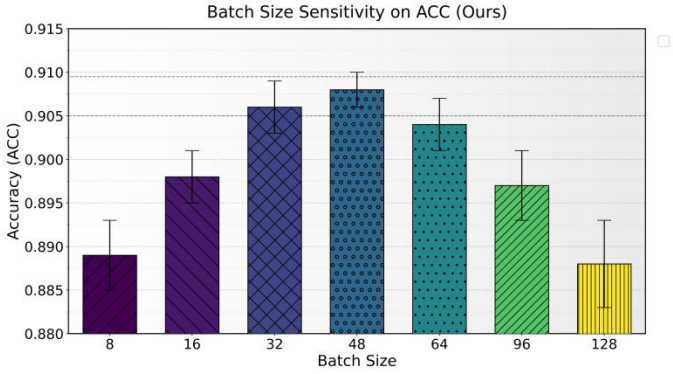

Figure 2. Experiment on the sensitivity of batch size to ACC

From the results in Figure 2, it can be observed that different batch sizes have a clear impact on model classification accuracy. With smaller batch sizes, such as 8 and 16, the ACC values are relatively low. This indicates that under these settings, the model struggles to capture the global semantics of texts stably. The gradients fluctuate more during training, which limits the generalization ability of the model. As the batch size increases, the performance improves, and the ACC values rise. This reflects the positive effect of batch size on training stability and convergence speed.

When the batch size increases to 32 and 48, the model reaches a higher level of ACC with more stable overall performance. This trend suggests that within an appropriate range of batch sizes, the model can capture global semantics while reducing the interference of redundant noise. In this case, the knowledge-graph-enhanced contextual modeling mechanism is used more effectively. The model not only maintains high accuracy but also achieves stability and robustness, showing good overall performance.

However, when the batch size further increases to 64, 96, and even 128, the ACC values decline to some extent. The reason may be that overly large batches make gradients too smooth during parameter updates, which reduces sensitivity to fine-grained semantic features and local variations. In knowledge-enhanced

classification tasks, this reduction in sensitivity directly affects the recognition of implicit relations and key entities in texts, thereby weakening overall classification performance.

In summary, the effect of batch size on model classification accuracy shows a typical "middle-optimal" trend. Both very small and very large batch sizes cause negative effects, while moderate batch sizes maximize the advantages of combining large language models with knowledge graphs. In practice, this finding suggests that batch size should be chosen carefully based on corpus scale and task characteristics. This ensures that training efficiency is maintained while the model can fully utilize external knowledge to improve the understanding and classification of complex semantics.

This paper also gives the impact of the labeling noise ratio on the experimental results, and the experimental results are shown in Figure 3.

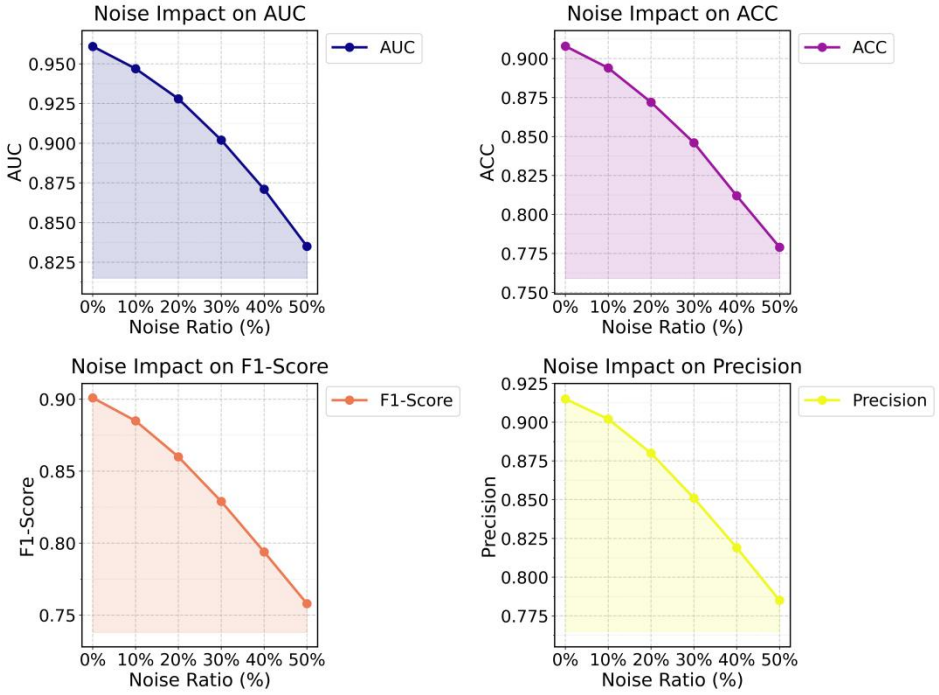

Figure 3. The impact of annotation noise ratio on experimental results

Figure 3 shows that as the proportion of annotation noise increases, all four metrics show a continuous downward trend. AUC remains high when the noise ratio is zero, indicating that the model effectively captures the global semantics and cross-sentence dependencies of the text using clean data. However, as the noise ratio increases, the AUC decreases, indicating that the model's ability to distinguish between positive and negative categories gradually weakens, which is directly related to the interference of noise on semantic representation.

The ACC metric also shows a significant decline. This indicates that as the number of noisy samples increases, the original key information in the text is obscured, making the model more likely to make incorrect predictions during classification. Although the knowledge augmentation mechanism can mitigate this interference to a certain extent, excessive noise can weaken the model's stability in text classification, leading to a decrease in overall accuracy. This result emphasizes the importance of data quality for classification tasks.

The declining trend in F1-Score is more revealing, reflecting the dual loss of recall and precision. When the noise ratio is low, the model can maintain a high balance of performance. However, as the noise ratio increases, recall is severely affected, the model makes more mistakes in identifying the true category, and

precision also decreases due to misclassification. This indicates that noise not only affects global performance but also weakens the model's performance in fine-grained classification scenarios.

The decrease in Precision further reveals the impact of noise on the model's judgment boundaries. As the noise ratio increases, the model's recognition accuracy for positive examples decreases, indicating that knowledge-augmented constraints are unable to fully offset interference when faced with semantic ambiguity or mislabeling. This phenomenon demonstrates that when building classification frameworks based on the integration of large language models and knowledge graphs, particular attention must be paid to the accuracy and consistency of data annotations. Otherwise, even if the model is well designed, performance will be severely limited.

This paper also presents an experiment on the sensitivity of weight decay to F1-Score, and the experimental results are shown in Figure 4.

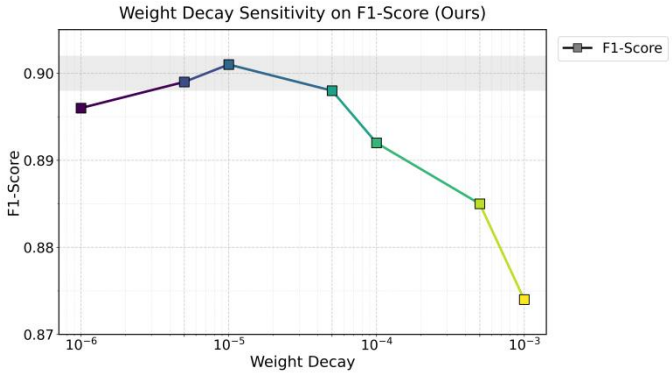

Figure 4. Experiment on the sensitivity of weight decay to F1-Score

The results in Figure 4 show that changes in weight decay have a significant impact on the model's F1-Score. When weight decay is low, the model maintains stable performance, with a high F1-Score. This indicates that the regularization strength is moderate, helping the model effectively capture the global semantics of the text corpus and avoid overfitting. As weight decay gradually increases, the F1-Score first increases and reaches a peak, demonstrating that moderate regularization can enhance the model's generalization ability and thus improve classification performance.

As weight decay continues to increase, the F1-Score begins to decline. This suggests that excessive regularization inhibits the model's ability to learn fine-grained semantic features, weakening some key information in complex text classification. In particular, in the knowledge augmentation framework, excessive weight constraints can limit the model's effective utilization of external knowledge, leading to reduced prediction accuracy. This phenomenon reveals the trade-off between regularization strength and semantic modeling capability.

Further analysis reveals that the optimal F1-Score value occurs within a moderate range of weight decay values, consistent with the model's need to avoid overfitting while maintaining sufficient expressiveness in complex tasks. Appropriate regularization can help the model maintain good stability across diverse text scenarios while enhancing its ability to filter redundant information. Therefore, weight decay sensitivity experiments can effectively reveal differences in model performance under varying regularization strengths.

Overall, these experimental results demonstrate that weight decay is a crucial hyperparameter choice for text classification tasks. Too low a value can lead to insufficient generalization on complex data, while too high a value can weaken the model's representational capabilities. By setting weight decay within an appropriate range, the advantages of integrating a large language model with a knowledge graph can be maximized, achieving higher classification accuracy and greater robustness. This finding guides subsequent parameter tuning for diverse data distributions and domain tasks.

This paper also gives the impact of knowledge graph coverage on experimental results, and the experimental results are shown in Figure 5.

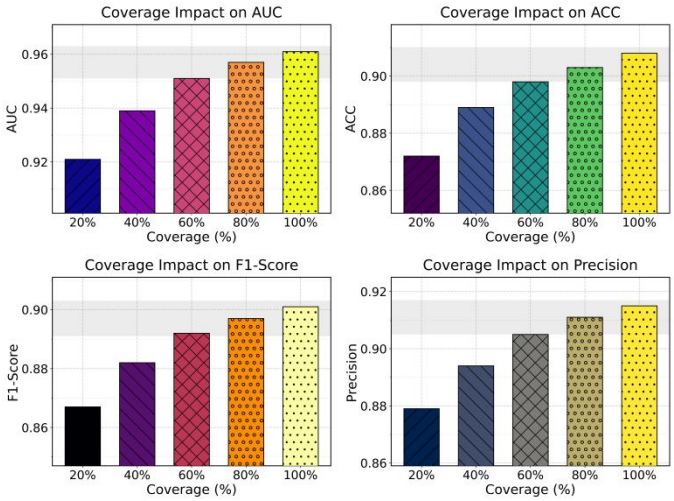

Figure 5. The impact of knowledge graph coverage on experimental results

From the results in Figure 5, it can be observed that as the coverage of the knowledge graph increases, the model shows a steady upward trend in AUC, ACC, F1-Score, and Precision. When the coverage is only 20 percent, the model performance remains at a relatively low level. This indicates that with insufficient knowledge graph information, the model cannot fully utilize external semantics to enhance long-text semantic modeling, which limits classification effectiveness. As the coverage expands and more entities and relations are introduced, the model's ability to supplement and constrain text semantics is significantly strengthened.

The upward trends in AUC and ACC are particularly clear. This shows that broader knowledge coverage improves the stability and accuracy of category discrimination. When the coverage reaches more than 60 percent, the model exhibits obvious advantages. This suggests that the knowledge graph plays a key role in addressing potential semantic omissions and ambiguities in texts. This phenomenon highlights the necessity of knowledge enhancement in text classification tasks.

The rising trend of F1-Score further confirms the positive effect of knowledge coverage on balanced performance. With low coverage, the model struggles to maintain both recall and precision. As coverage increases, the model achieves higher recall while keeping precision at a strong level. This means that the knowledge graph provides more comprehensive semantic information, allowing the model to identify true classes in complex texts more accurately and reduce both false positives and false negatives.

The improvement in Precision also reflects the importance of knowledge coverage. With higher coverage, the accuracy of recognizing positive cases increases significantly. This indicates that the knowledge graph helps narrow the candidate space and reduce semantic noise. Overall, the experimental results clearly show that the effect of knowledge graph coverage on text classification is both gradual and significant. This confirms that combining large language models with high-coverage knowledge representation can effectively improve classification performance and robustness.

5. Conclusion

This study addresses the challenges of semantic redundancy, insufficient contextual dependency, and limited use of external knowledge in long-text classification. A classification framework is proposed that integrates large language models with knowledge graphs. The framework combines the strong semantic modeling ability of language models with the explicit semantic constraints of knowledge graphs, achieving multi-level

enhancement of text representation. Through this design, the model not only captures global semantics and cross-paragraph dependencies in texts but also leverages the structured information of knowledge graphs to fill semantic gaps. This significantly improves classification accuracy and robustness. At the theoretical level, the method demonstrates the advantages of dual modeling with semantics and knowledge, providing a new solution for long-text processing.

At the methodological level, this study incorporates dynamic embeddings, attention mechanisms, and knowledge alignment strategies to achieve deep integration of text semantics and knowledge graph information. The model flexibly models context while using external knowledge to reduce ambiguity and errors in reasoning. This design not only overcomes the limitations of traditional deep learning methods in long-text scenarios but also enhances adaptability to complex semantics and domain-specific texts. Experimental design and sensitivity analysis further confirm the influence of hyperparameters, environmental conditions, and data characteristics on model performance. The results show that the framework maintains robustness and transferability under different conditions, laying the foundation for large-scale applications.

From an application perspective, the outcomes of this study have important implications for multiple real-world scenarios. In finance, the framework can help identify key information in news or reports, supporting risk assessment and decision-making. In knowledge-intensive fields such as healthcare and law, it can enhance the understanding of domain texts with the aid of knowledge graphs, improving the accuracy of automated processing. In education and public administration, it can also serve as a core tool for information organization and retrieval, enabling efficient extraction of key content from massive data. Therefore, the method not only holds research value in academia but also shows broad potential in industry and society.

Looking ahead, the approach can be extended to cross-modal and cross-domain knowledge-enhanced tasks. For example, integrating images, audio, or multimodal data with knowledge graphs could support classification and retrieval in complex multimodal scenarios. The dynamic updating of knowledge graphs can also be combined with the continual learning ability of large language models to achieve adaptive and intelligent systems. As the scale and quality of knowledge graphs improve, the framework will further demonstrate potential in interpretability, security, and cross-lingual processing. These directions will not only advance the evolution of intelligent systems but also provide new approaches and practical paths for building trustworthy artificial intelligence ecosystems.

References

[1] Liu Y, Zhang K, Huang Z, et al. Enhancing hierarchical text classification through knowledge graph integration[C]//Findings of the association for computational linguistics: ACL 2023. 2023: 5797-5810.

[2] Hu L, Liu Z, Zhao Z, et al. A survey of knowledge enhanced pre-trained language models[J]. IEEE Transactions on Knowledge and Data Engineering, 2023, 36(4): 1413-1430.

[3] Chen Y, Röder D, Erker J J, et al. Retrieval-Augmented Knowledge Integration into Language Models: A Survey[C]//Proceedings of the 1st Workshop on Towards Knowledgeable Language Models (KnowLLM 2024). 2024: 45-63.

[4] Liu J, Yang L. Knowledge-enhanced prompt learning for few-shot text classification[J]. Big Data and Cognitive Computing, 2024, 8(4): 43.

[5] Zhang C, Zhu H, Peng X, et al. Hierarchical information matters: Text classification via tree based graph neural network[J]. arXiv preprint arXiv:2110.02047, 2021.

[6] Wang K, Ding Y, Han S C. Graph neural networks for text classification: A survey[J]. Artificial intelligence review, 2024, 57(8): 190.

[7] Onan A. Hierarchical graph-based text classification framework with contextual node embedding and BERT-based dynamic fusion[J]. Journal of king saud university-computer and information sciences, 2023, 35(7): 101610.

[8] Chen L, Liu S. Explainable Cross-Market Causal Discovery via Financial Knowledge Graphs[J], 2024, 3(1).

[9] Wang W, Cheng J. Trustworthy LLM-Agent-Oriented Enterprise ETL Orchestration via Neural Execution Verification[J], 2024, 3(4).

[10] Paletto L, Basile V, Esposito R. Label augmentation for zero-shot hierarchical text classification[C]//Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). 2024: 7440-7456.

[11] Ji K, Lian Y, Gao J, et al. Hierarchical verbalizer for few-shot hierarchical text classification[C]//Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). 2023: 2918-2933.

[12] Joseph S, Joshi H. ULMFiT: Universal language model fine-tuning for text classification[J]. International Journal of Advanced Medical Sciences and Technology (IJAMST), 2024, 4.

[13] Louail M, Hamdi-Cherif C K M, Hamdi-Cherif A. Tasneef: a fast and effective hybrid representation approach for Arabic text classification[J]. IEEE Access, 2024.