Ambiguity-Aware Keyword-Enhanced Label-Aware Semantic Fusion for Text Classification

Peijun Xie[1[0009-0008-0661-742X]], Chuncheng Chi[1[0009-0008-2035-5892]], Fuxue Li[1[0000-0002-2878-6119]] , Hong Yan[1[0000-0003-1834-8956]]

[1] Yingkou Institute of Technology, Yingkou, China

xiepeijun0216@163.com

Abstract. Text classification remains challenging in real-world scenarios due to semantic ambiguity, insufficient contextual information, and unclear category boundaries. To address these issues, this paper proposes an Ambiguity-Aware Keyword-Enhanced Label-Aware Semantic Fusion Network (AAK-LASFNet). The model combines local contextual features extracted by TextRCNN with global semantic representations obtained from a frozen pre-trained language model. To improve discriminative representation learning, ambiguity estimation, label-aware attention, keyword enhancement, high-order semantic fusion, and semantic consistency regularization are integrated into a unified framework. Experiments on four benchmark datasets show that AAK-LASFNet achieves better classification accuracy than strong baseline models. Ablation and ambiguity-oriented analyses further demonstrate the effectiveness and robustness of the proposed method.

Keywords: Text Classification, Semantic Ambiguity, Label-Aware Attention, Keyword Enhancement, High-Order Semantic Fusion.

1 Introduction

Text classification is a fundamental task in natural language processing and has been widely applied to news categorization, sentiment analysis, topic detection, and information retrieval [1–4]. With the development of deep learning, neural models such as CNNs, RNNs, and TextRCNN have been widely used to capture local contextual and semantic features from textual sequences [5–8]. More recently, Transformer-based pre-trained language models, including BERT and RoBERTa, have further improved text representation learning through large-scale contextual pre-training [9–14].

Despite these advances, real-world text classification still faces several challenges. First, short texts and noisy user-generated content often provide insufficient contextual clues. Second, semantic ambiguity may cause the same word or phrase to express different meanings under different contexts. Third, category boundaries are sometimes

unclear, especially in fine-grained or multi-class classification scenarios. These issues make it difficult for models relying only on local lexical features or single-view representations to learn robust and discriminative semantics.

Large-scale pre-trained models and LLMs provide stronger global semantic modeling ability, but directly applying them to downstream classification often brings high computational cost and limited controllability. Therefore, an effective solution is to combine lightweight local semantic encoders with frozen global semantic representations while explicitly modeling ambiguity and category-related semantic information.

To this end, this paper proposes an Ambiguity-Aware Keyword-Enhanced Label-Aware Semantic Fusion Network (AAK-LASFNet). The framework integrates local contextual features, frozen LLM-based global semantics, label-aware attention, keyword cues, ambiguity estimation, and high-order semantic fusion into a unified architecture. In addition, a semantic consistency loss is introduced to align local and global semantic views and improve representation stability.

The main contributions of this paper are summarized as follows:

1.A unified ambiguity-aware semantic fusion framework is proposed for robust text classification.

2.Label-aware attention and keyword enhancement are introduced to strengthen category-related semantic representation.

3.An ambiguity estimation module and high-order fusion strategy are designed to improve classification robustness under uncertain semantic conditions.

4.Experiments on four benchmark datasets demonstrate the effectiveness of the proposed method.

The remainder of this paper is organized as follows. Section 2 reviews related work. Section 3 presents the proposed method. Section 4 reports experimental results and analysis. Section 5 concludes the paper.

2 Related Work

2.1 Neural Text Classification

Early neural text classification methods mainly relied on convolutional and recurrent architectures. CNN-based models are effective in capturing local n-gram features, while RNN-based models are suitable for modeling sequential dependencies [5–7]. TextRCNN further combines recurrent contextual modeling with convolutional feature extraction, improving semantic representation for classification tasks [8]. Attention mechanisms have also been introduced to emphasize informative tokens and improve interpretability [19–21]. However, these models mainly depend on local contextual patterns and may be insufficient for resolving semantic ambiguity in complex or noisy texts.

2.2 Pre-trained Language Models and Large Language Models

Pre-trained language models have significantly improved text classification by learning contextualized representations from large-scale corpora. ELMo, BERT, RoBERTa, ALBERT, and DeBERTa have demonstrated strong representation ability across various NLP tasks [9–14]. More recently, LLMs have shown stronger semantic understanding and reasoning capability due to large-scale pre-training [17,18]. Nevertheless, directly applying large-scale models to downstream classification tasks usually results in high computational cost. Therefore, using frozen pre-trained representations as complementary global semantic information provides a practical solution.

2.3 Semantic Fusion and Label-aware Classification

Semantic fusion methods aim to integrate heterogeneous information from multiple semantic views to improve representation robustness [23–25]. Meanwhile, label-aware classification explicitly models the interaction between text features and label semantics, helping the model capture category-relevant information [22]. Existing studies have shown the effectiveness of multi-view learning, knowledge-enhanced representation, and label-aware attention in classification tasks. However, most methods focus on a single aspect of semantic modeling and rarely integrate ambiguity estimation, label semantics, keyword cues, and local-global semantic fusion within a unified framework. This motivates the proposed AAK-LASFNet.

3 Methodology

3.1 Overall Framework

To address the challenges of semantic ambiguity and unclear category boundaries in text classification, this paper proposes an ambiguity-aware semantic fusion network (AAK-LASFNet). The proposed framework aims to construct robust and discriminative text representations by jointly modeling local contextual semantics, global semantic knowledge, and category-related semantic cues.

Given an input text x, the model first employs a TextRCNN encoder to capture local contextual features, while a large language model is utilized to obtain global semantic representations, forming a dual-view semantic representation of the text. An ambiguity estimation module is then introduced to quantify the semantic uncertainty of each sample. Meanwhile, a label-aware attention mechanism strengthens the interaction between textual features and label semantics, and a keyword enhancement module highlights salient semantic cues within the text. Finally, a high-order semantic fusion module integrates the multi-source semantic information to generate the final representation for classification.

Let the input text be represented as

$$x = (w_1, w_2, \dots, w_n) \tag{1}$$

where n denotes the length of the text and w_i is the i-th token.

A schematic overview of the proposed framework is provided in Fig. 1.

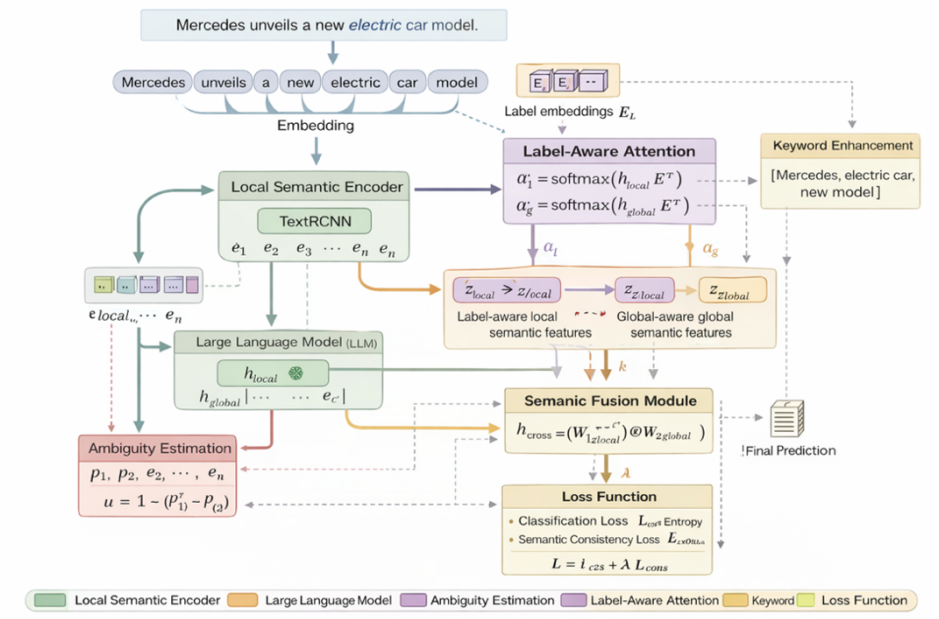

Fig. 1. AAK-LASFNet Model architecture.

3.2 Local Semantic Representation

To capture local contextual information, TextRCNN is employed as the local semantic encoder. The input tokens are first mapped into embedding vectors:

$$e_i = \text{Embedding}(w_i) \tag{2}$$

where $e_i \in \mathbb{R}^d$ denotes the embedding of the i-th token.

Then a bidirectional recurrent network is used to model contextual dependencies:

$$\overrightarrow{h_i} = \text{RNN}\left(e_i, \overrightarrow{h_{i-1}}\right) \tag{3}$$

$$\overleftarrow{h_i} = \text{RNN}\left(e_i, \overleftarrow{h_{i+1}}\right) \tag{4}$$

The hidden states from the two directions are concatenated as

$$h_i = \left[\overrightarrow{h_i}; \overleftarrow{h_i}\right] \tag{5}$$

Finally, a pooling operation is applied to aggregate the sequence-level representation:

$$h_{\text{local}} = \text{Pooling}(h_1, h_2, \ldots, h_n) \tag{6}$$

This representation mainly captures local semantic patterns and contextual features within the input text.

3.3 Global Semantic Representation

The global semantic feature is obtained from a frozen pre-trained language model. In our implementation, RoBERTa-base is selected for English texts, whereas Chinese-RoBERTa-wwm-ext is employed for Chinese texts. These encoders are not updated

during training, so the additional computational burden is limited and the extracted semantic features remain stable. For each input sequence, we use the final-layer representation at the [CLS] position as the sentence-level global semantic vector.

$$h_{\text{global}} = \text{LLM}(x) \tag{7}$$

Here, $h_{global} \in \mathbb{R}^d$ denotes the global semantic vector of the input text, which provides complementary semantic information beyond the local contextual features captured by lightweight encoders.

3.4 Label-Aware Attention

Traditional text classification models typically treat labels as discrete indices and fail to fully exploit their underlying semantic information. To explicitly model the interaction between textual representations and label semantics, a label-aware attention mechanism is introduced.

Assume that there are C categories, and each category is represented by a learnable embedding vector. All label embeddings form a matrix $E \in \mathbb{R}^{C \times d}$, where d denotes the embedding dimension. Given the local semantic representation $h_{local} \in \mathbb{R}^{1 \times d}$ and the global semantic representation $h_{global} \in \mathbb{R}^{1 \times d}$, the relevance between textual features and label semantics is computed through a similarity-based attention mechanism:

$$\alpha_l = \text{softmax}(h_{local} E^T) \tag{8}$$

$$\alpha_g = \text{softmax}\left(h_{global} E^T\right) \tag{9}$$

where $\alpha_l, \alpha_g \in \mathbb{R}^{1 \times C}$ denote the attention weights over label embeddings. These weights indicate the degree of semantic alignment between the input text and each category.

Based on the learned attention distributions, label-aware semantic representations are obtained as:

$$z_{local} = \alpha_l E \tag{10}$$

$$z_{global} = \alpha_g E \tag{11}$$

where $z_{local}, z_{global} \in \mathbb{R}^{1 \times d}$. By incorporating label semantics into the representation learning process, the model enhances category-relevant feature aggregation and improves class discrimination. These label-aware representations are subsequently integrated with other semantic features in the semantic fusion module to construct the final classification representation.

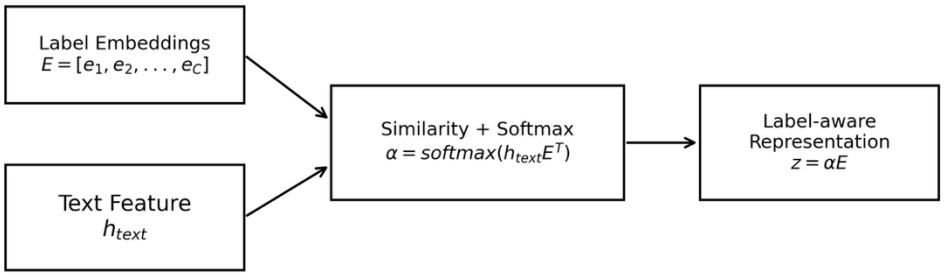

Fig. 2. Label-aware attention mechanism

As shown in Fig. 2, the label-aware attention module aligns textual representations with learnable label embeddings to enhance category-related semantic aggregation.

3.5 Keyword Enhancement Module

In textual data, salient keywords often provide strong cues for category discrimination. To explicitly emphasize such informative semantic signals, a keyword enhancement module is introduced. Specifically, keywords are first extracted from the input text using a large language model, forming a keyword set $K = (k_1, k_2, \dots, k_m)$, where m denotes the number of extracted keywords. In this work, at most five keywords are extracted for each input text.

The keyword extraction is conducted via a prompt-based strategy. The LLM is instructed as follows:

"Given the following text, extract up to five keywords that are most useful for topic classification. Return only the keywords separated by commas. Text: [input text]."

For Chinese datasets, the prompt is translated into Chinese while preserving the same extraction constraint.

To ensure the quality and consistency of extracted keywords, a post-processing step is applied. Specifically, duplicated words, punctuation marks, and stop words are removed. In addition, keywords that are not present in or semantically inconsistent with the original text are filtered out. If fewer than five keywords are obtained, zero-padding is applied after embedding to maintain a consistent representation.

The keyword sequence is then mapped into embedding vectors using the shared embedding layer and aggregated through an average pooling operation to obtain a compact semantic representation:

$$k = \text{Encoder}(K) \tag{12}$$

This keyword-aware representation captures the thematic semantics of the input text and is jointly utilized with other semantic features during the subsequent semantic fusion process, enabling the model to better focus on category-relevant information and improve representation discriminability.

3.6 Ambiguity Estimation

To explicitly model semantic ambiguity in texts with unclear category boundaries, an uncertainty estimation module is introduced to quantify the degree of classification ambiguity. The ambiguity score is derived from the predicted probability distribution of the classifier:

$$p = [p_1, p_2, \dots, p_C] \tag{13}$$

where p_1 and p_2 denote the highest and second-highest category probabilities, respectively. The semantic uncertainty is defined as:

$$u = 1 - (p_1 - p_2) \tag{14}$$

A larger value of u indicates higher prediction uncertainty, suggesting that the input text is more semantically ambiguous and difficult to classify. In such cases, the model is encouraged to rely more on global semantic information to improve representation robustness and classification reliability.

3.7 High-Order Semantic Fusion

To explicitly model the interaction between local and global semantic representations, a cross-semantic interaction mechanism is introduced:

$$h_{cross} = (W_1 z_{local}) \odot (W_2 z_{global}) \tag{15}$$

where $\odot$denotes element-wise multiplication. This operation enables fine-grained feature interaction across corresponding semantic dimensions.

Subsequently, multi-source semantic features are aggregated through concatenation:

$$h_{meta} = [z_{local}; z_{global}; h_{cross}; k; u] \tag{16}$$

where $k$represents the keyword-aware semantic representation and $u$denotes the semantic uncertainty score.

Finally, the fused representation is obtained via a nonlinear transformation using a multilayer perceptron (MLP):

$$h_{fusion} = \mathrm{ReLU}(W_f h_{meta} + b_f) \tag{17}$$

This fusion strategy jointly integrates local contextual semantics, global semantic knowledge, cross-view interaction features, keyword information, and ambiguity-aware signals, resulting in more discriminative and robust semantic representations for classification.

3.8 Loss Function

To enforce consistency between different semantic representations, a semantic consistency loss is incorporated into the training objective alongside the classification loss.

The classification loss is defined as:

$$L_{cls} = \mathrm{CrossEntropy}(y, \hat{y}) \tag{18}$$

where y denotes the ground-truth label and $\hat{y}$ represents the predicted probability distribution.

To reduce the discrepancy between local and global semantic representations, a cosine-based consistency loss is introduced:

$$L_{cons} = 1 - \cos(z_{local}, z_{global}) \tag{19}$$

The overall training objective is formulated as:

$$L = L_{cls} + \lambda L_{cons} \tag{20}$$

where λ is a balancing coefficient that controls the contribution of the consistency constraint.

By jointly optimizing these objectives, the model maintains strong classification capability while encouraging alignment between semantic views, leading to more stable and discriminative feature representations.

4 Experiments

4.1 Datasets

Experiments are conducted on four benchmark datasets, including Cnews, IFLYTEK, Today's Headline, and AG News. These datasets cover both Chinese and English text

classification scenarios, including news categorization, short-text classification, and large-scale multi-class classification. Their statistics are summarized in Table 1.

Table 1. Details of the dataset.

Dataset	Language	Classes	Training Samples	Test Samples
Cnews	Chinese	10	50,000	10,000
IFLYTEK	Chinese	119	121,000	7,000
Today's Headline	Chinese	15	382,688	40,000
AG News	English	4	120,000	7,600

4.2 Experimental Settings

All models are trained under the same experimental protocol. In AAK-LASFNet, TextRCNN is used as the local encoder, with a recurrent hidden size of 256 and word embedding dimension of 300. The global semantic encoder is frozen during training. Adam is used for optimization, with a learning rate of 1×10^{-4}, batch size of 32, and maximum epoch number of 20. The semantic consistency coefficient is set to $\lambda = 0.1$. Accuracy is adopted as the main evaluation metric, and results are averaged over five runs with different random seeds.

4.3 Overall Performance Comparison

We compare AAK-LASFNet with TextCNN, TextRNN, TextRCNN, BERT, and RoBERTa. Table 2 reports the classification accuracy on four benchmark datasets, where the best result is marked in bold.

Table 2. Overall classification performance (Accuracy).

Model	Cnews	IFLYTEK	Today's Headline	AG News
TextCNN	92.1	61.3	95.0	91.2
TextRNN	91.6	60.4	94.6	90.8
TextRCNN	93.4	62.7	96.1	92.5
BERT	95.2 ± 0.2	64.8 ± 0.3	97.3 ± 0.2	94.6 ± 0.2
RoBERTa	95.8 ± 0.1	65.2 ± 0.2	97.8 ± 0.1	95.1 ± 0.2
Proposed	**96.7 ± 0.2**	**67.3 ± 0.2**	**98.5 ± 0.1**	**96.4 ± 0.1**

As shown in Table 2, AAK-LASFNet achieves the best performance on all datasets. Compared with traditional neural baselines, the improvement indicates the benefit of incorporating global semantic knowledge, label semantics, keyword cues, and ambiguity-aware fusion. Compared with BERT and RoBERTa, the proposed model also obtains further gains, suggesting that the unified semantic fusion strategy provides additional discriminative information beyond standard pre-trained representations.

4.4 Ablation Study

The contribution of each module is analyzed through the ablation results presented in Table 3.

Table 3. Ablation study results.

Model Variant	Cnews	IFLYTEK
Base (TextRCNN + LLM)	95.8	65.4
+ Uncertainty Estimation	96.1	66.1
+ Label-aware Attention	96.3	66.5
+ Advanced Semantic Fusion	96.5	66.9
+ Semantic Consistency Loss	96.6	67.1
Full Model	**96.7**	**67.3**
w/o Keyword Enhancement	**96.4**	**67.2**

As shown in Table 3, each module contributes to the final performance. Ambiguity estimation improves the handling of uncertain samples, label-aware attention enhances category-related semantic alignment, and high-order semantic fusion strengthens the interaction between local and global representations. The semantic consistency loss further improves representation stability. In addition, removing the keyword enhancement module causes a performance drop, confirming the usefulness of keyword cues.

4.5 Analysis on Ambiguous Texts

To evaluate the robustness of the proposed model under semantic ambiguity, the test set is partitioned into ambiguous and non-ambiguous subsets. A sample is considered ambiguous when the gap between the top two predicted probabilities is below 0.1, a threshold selected empirically. The corresponding results are presented in Table 4.

Table 4. Performance on ambiguous and non-ambiguous subsets.

Model	Non-Ambiguous	Ambiguous
TextRCNN	96.4	79.5
BERT	97.2	82.1
Proposed Model	**97.8**	**87.6**

As reported in Table 4, the proposed approach exhibits clear performance gains on semantically ambiguous instances, indicating the effectiveness of the ambiguity-aware modeling and semantic fusion strategy in addressing uncertainty during classification.

4.6 Visualization and Sensitivity Analysis

To further analyze the representation quality and interpretability of the proposed framework, several visualization studies are conducted. As shown in Fig. 3, the fused representations learned by the proposed model form more compact clusters and clearer class boundaries than local or global representations alone, indicating that the semantic fusion strategy improves feature discriminability.

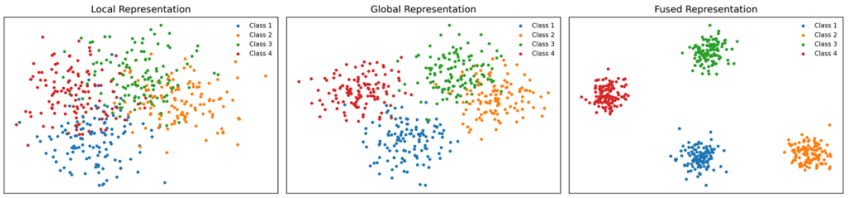

Fig. 3. t-SNE visualization of local, global, and fused semantic representations.

Fig. 4 presents representative attention and keyword visualization results. The model assigns higher weights to semantically informative tokens, and the extracted keywords are generally consistent with the main topic of the input text. This suggests that the proposed framework can capture salient category-related cues and provide a certain degree of interpretability.

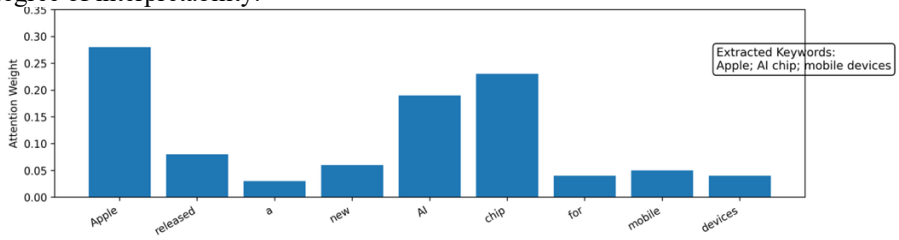

Fig. 4. Attention weights and extracted keywords (representative English sample).

In addition, Fig. 5 reports the sensitivity analysis of the semantic consistency coefficient λ. The model achieves stable performance when λ is set within a moderate range, and the best result is obtained at λ = 0.1. Fig. 6 further shows the confusion matrix on the Cnews dataset, where most predictions are concentrated along the diagonal. The remaining errors mainly occur between semantically related categories, indicating that the proposed model learns meaningful category boundaries.

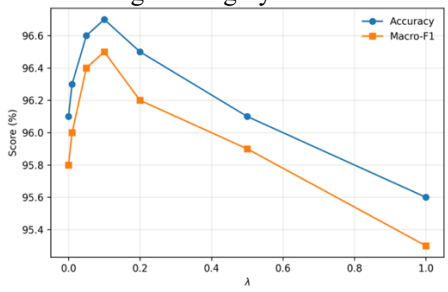

Fig. 5. Parameter sensitivity analysis.

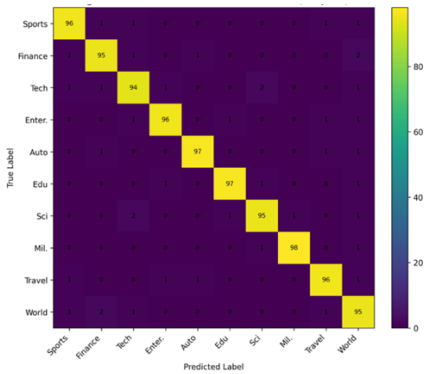

Fig. 6. Confusion matrix on the Cnews dataset.

5 Conclusion

This paper proposes AAK-LASFNet, an ambiguity-aware keyword-enhanced label-aware semantic fusion network for text classification. The model integrates local contextual semantics, frozen pre-trained global semantics, label-aware attention, keyword enhancement, ambiguity estimation, and high-order semantic fusion into a unified framework. A semantic consistency loss is further introduced to align local and global semantic views. Experiments on four benchmark datasets show that the proposed method achieves consistent improvements over strong baselines. Ablation studies and ambiguity-oriented analysis further verify the effectiveness and robustness of the proposed components. Future work will explore more efficient semantic fusion strategies and extend the framework to low-resource, long-document, and multilingual classification scenarios.

Acknowledgments. This work is partially supported by School-level Scientific Research Project of Yingkou Institute of Technology (QNL202527); Sub-project of Electrical Engineering College of Yingkou Institute of Technology (AICPTI08); School-level Scientific Research Project of Yingkou Institute of Technology (QNL202520); School-level Scientific Research Project of Yingkou Institute of Technology (FDL202504).

References

1. Minaee S, Kalchbrenner N, Cambria E, et al. Deep learning--based text classification: a comprehensive review[J]. ACM computing surveys (CSUR), 2021, 54(3): 1-40.
2. Taha K, Yoo P D, Yeun C, et al. A comprehensive survey of text classification techniques and their research applications. Computer Science Review, 2024.
3. Gasparetto A, Marcuzzo M, Zangari A, et al. A survey on text classification algorithms: From text to predictions[J]. Information, 2022, 13(2): 83.
4. Li Q, Peng H, Li J, et al. A survey on text classification: From traditional to deep learning[J]. ACM Transactions on Intelligent Systems and Technology (TIST), 2022, 13(2): 1-41.
5. Kim Y. Convolutional neural networks for sentence classification[C]//Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP). 2014: 1746-1751.
6. Zhang X, Zhao J, LeCun Y. Character-level convolutional networks for text classification[J]. Advances in neural information processing systems, 2015, 28.
7. Lai S, Xu L, Liu K, et al. Recurrent convolutional neural networks for text classification[C]//Proceedings of the AAAI conference on artificial intelligence. 2015, 29(1).
8. Yang Z, Yang D, Dyer C, et al. Hierarchical attention networks for document classification[C]//Proceedings of the 2016 conference of the North American chapter of the association for computational linguistics: human language technologies. 2016: 1480-1489.
9. Joulin A, Grave E, Bojanowski P, et al. Bag of tricks for efficient text classification[C]//Proceedings of the 15th conference of the European chapter of the association for computational linguistics: volume 2, short papers. 2017: 427-431.

10. Howard J, Ruder S. Universal language model fine-tuning for text classification[C]//Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). 2018: 328-339.
11. Devlin J, Chang M W, Lee K, et al. Bert: Pre-training of deep bidirectional transformers for language understanding[C]//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). 2019: 4171-4186.
12. Liu Y, Ott M, Goyal N, et al. Roberta: A robustly optimized bert pretraining approach[J]. arXiv preprint arXiv:1907.11692, 2019.
13. Lan Z, Chen M, Goodman S, et al. Albert: A lite bert for self-supervised learning of language representations[J]. arXiv preprint arXiv:1909.11942, 2019.
14. He P, Liu X, Gao J, et al. Deberta: Decoding-enhanced bert with disentangled attention[J]. arXiv preprint arXiv:2006.03654, 2020.
15. Ethayarajh K. How contextual are contextualized word representations? Comparing the geometry of BERT, ELMo, and GPT-2 embeddings[C]//Proceedings of the 2019 conference on empirical methods in natural language processing and the 9th international joint conference on natural language processing (EMNLP-IJCNLP). 2019: 55-65.
16. Yenduri G, Ramalingam M, Selvi G C, et al. GPT (generative pre-trained transformer)—a comprehensive review on enabling technologies, potential applications, emerging challenges, and future directions[J]. IEEE access, 2024, 12: 54608-54649.
17. Brown T, Mann B, Ryder N, et al. Language models are few-shot learners[J]. Advances in neural information processing systems, 2020, 33: 1877-1901.
18. Wei J, Tay Y, Bommasani R, et al. Emergent abilities of large language models[J]. arXiv preprint arXiv:2206.07682, 2022.
19. Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need[J]. Advances in neural information processing systems, 2017, 30.
20. Lin Z, Feng M, Santos C N, et al. A structured self-attentive sentence embedding[J]. arXiv preprint arXiv:1703.03130, 2017.
21. Wang Z, Yang B. Attention-based bidirectional long short-term memory networks for relation classification using knowledge distillation from BERT[C]//2020 IEEE Intl conf on dependable, autonomic and secure computing, intl conf on pervasive intelligence and computing, intl conf on cloud and big data computing, intl conf on cyber science and technology congress (DASC/PiCom/CBDCom/CyberSciTech). IEEE, 2020: 562-568.
22. Liu M, Liu L, Cao J, et al. Co-attention network with label embedding for text classification[J]. Neurocomputing, 2022, 471: 61-69.
23. Cheng Q, Shi W. Hierarchical multi-label text classification of tourism resources using a label-aware dual graph attention network[J]. Information Processing & Management, 2025, 62(1): 103952.
24. Ding Y, Shalaby M, Singh N S S, et al. Multi-view text classification through integrated RNN autoencoder learning of word, sentence, emotion and paragraph representations[J]. Scientific Reports, 2025.
25. Wei J, Wang X, Schuurmans D, et al. Chain-of-thought prompting elicits reasoning in large language models[J]. Advances in neural information processing systems, 2022, 35: 24824-24837.