Artificial Intelligence and Computing Innovations

Intelligence Science Publishing

https://www.ispublishing.com

ISSN: 3070-4944

Vo.4, No.2, 2024

Intelligent Domain Text Classification through Fine-Grained Semantic Representation Learning

Lanming Ma[1], Eric Lin[2]*, Zhengxin Zhang[3]

[1]New York University, New York, USA

[2]University of Rochester, New York, USA

[3]Cornell University, Ithaca, USA

**Corresponding author: Ericlin.1996@gmail.com*

Abstract: This study focuses on fine-grained discriminative feature extraction for domain text classification and proposes a BERT-based feature enhancement method to address the limitations of traditional models in domain adaptation and fine-grained modeling. The approach first uses a pretrained language model to generate contextual representations of input texts, and then introduces a multi-head attention mechanism to capture both global dependencies and local semantic details, leading to more discriminative feature representations. A contrastive learning constraint is further designed to optimize the separability of positive and negative samples in the semantic space, enabling the model to maintain strong discriminative ability even under blurred category boundaries and subtle semantic differences. In addition, regularization and joint optimization strategies are employed to ensure feature separability while preserving stability and generalization. To evaluate the effectiveness of the method, sensitivity experiments were conducted on key hyperparameters, including learning rate, weight decay, maximum sequence length, contrastive learning temperature coefficient, number of attention heads, and hidden dimension, followed by comparative analyses of model performance under different conditions. Experimental results show that the proposed method achieves superior performance on accuracy, precision, recall, and F1-score, demonstrating higher-level fine-grained discrimination in complex domain text classification tasks. The study not only presents innovation in method design but also shows robustness and effectiveness through multi-dimensional experimental validation, providing new technical insights and practical directions for domain text classification.

Keywords: Domain text classification; fine-grained discrimination; BERT; contrastive learning

1. Introduction

In the context of rapid advances in information technology and digitalization, massive amounts of text data are generated across domains, from news reports and medical records to financial analyses and legal judgments. The scale and speed of this generation far exceed those of the past[1]. How to classify and organize such large text resources effectively has become a pressing issue for both academia and industry. Domain-specific texts are often highly specialized and complex. Traditional classification methods struggle to capture fine-grained semantic distinctions and discriminative features. Therefore, fine-grained text classification in specific domains is not only a vital branch of natural language processing but also a core support for intelligent information management and knowledge discovery[2].

In recent years, the rise of deep learning has brought new opportunities for text classification. Pretrained language models have shown a strong ability to learn semantic representations from large corpora and transfer them to downstream tasks. However, general-purpose models often face insufficient semantic transfer and weak domain adaptation when applied across domains. This makes it difficult to accurately recognize subtle features in domain texts. Financial, medical, or legal documents usually contain many technical terms, long-tail concepts, and strong contextual dependencies. Without mechanisms for targeted feature extraction, models tend to perform poorly in fine-grained classification tasks. This highlights the importance of extracting fine-grained discriminative features tailored to domain texts[3].

Against this background, it is crucial to combine the strengths of language models with the unique characteristics of domain texts[4]. By modeling fine-grained discriminative features, models can better perceive subtle semantic differences. This enhances classification ability in scenarios where category boundaries are blurred or classes are highly similar. Such improvements not only raise classification accuracy but also provide more reliable foundations for downstream applications. For example, in medical text classification, subtle differences in language may relate to disease severity. In financial risk control, small variations in terminology may point to different levels of risk. In legal document analysis, specific wording can influence the applicability of laws. Thus, fine-grained feature extraction has broad and profound practical value[5].

From the perspective of academic research, domain-oriented fine-grained text classification enriches the theoretical framework of natural language processing. It also offers new approaches to knowledge modeling in specific fields. This line of research drives classification tasks from coarse to fine levels and from generalization to specialization. It better aligns with the demands of real applications. At the same time, it fosters cross-disciplinary integration and builds bridges between linguistics, computer science, and domain applications. By constructing frameworks that balance generality and specialization, research can advance language intelligence toward higher levels, enabling models to go beyond surface language understanding and capture the deeper logic and rules of domain knowledge[6].

In terms of application value, fine-grained discriminative feature extraction based on BERT can significantly improve classification outcomes. It also has strong potential to impact applications such as intelligent retrieval, opinion mining, risk warning, and decision support. The results of this research can enhance automation and intelligence in information processing, reduce the burden of manual analysis, and support more precise data-driven decision-making[7]. In the era of the digital economy and intelligent society, this research represents not only technological progress but also a key driver of digital transformation across industries. By exploring methods for fine-grained feature extraction in domain text classification, it is possible to address challenges of language understanding in complex contexts and build solid foundations for future intelligent information systems.

2. Related work

Research on related work mainly focuses on three areas: text classification, pretrained language models, and fine-grained feature modeling. In traditional methods, early text classification relied on bag-

of-words models, TF-IDF vectors, or rule-based feature engineering[8]. These methods could capture some statistical characteristics of text, but their performance was limited by the capacity of handcrafted features and could not effectively represent complex semantic relationships. With the development of deep learning, convolutional neural networks and recurrent neural networks were introduced into text classification. They were able to capture local context and sequential dependencies and significantly improved classification performance. However, these methods still showed limitations when dealing with cross-domain texts or fine-grained distinctions, as they could not fully represent the deep semantic structures in text[9].

In recent years, the emergence of pretrained language models has brought breakthroughs to text classification. By pretraining on large-scale corpora, such models can learn general semantic representations and then transfer them to downstream tasks through fine-tuning. Yet general pretrained models often face semantic shift and domain mismatch when applied to specific fields, leading to a decline in classification performance[10]. To address this issue, researchers proposed various strategies, such as retraining on domain corpora or adding task-specific constraints during pretraining to improve adaptability. These methods partially alleviated the limitations of general models in domain text classification. However, challenges remain in fine-grained discrimination, especially in situations with blurred category boundaries and subtle semantic differences[11].

Research on fine-grained feature extraction has provided new directions. Existing work focuses on enhancing the model ' s sensitivity to semantic details to better distinguish similar categories. Some studies use multi-level attention mechanisms to highlight key semantic units and strengthen the capture of local features. Other studies attempt to integrate external knowledge graphs or domain lexicons to compensate for the limitations of pure text modeling and achieve more refined feature representations in specific domains. In addition, methods such as multi-task learning and contrastive learning are widely applied to fine-grained classification. These approaches improve discriminative power through joint optimization or semantic contrast. They show promising results in experiments, but the complexity and specialization of domain texts still limit their generalization ability in real-world applications[12].

Overall, existing research has made significant progress in general text classification, domain adaptation, and fine-grained discriminative feature modeling, but limitations remain. Traditional methods cannot meet the demands of complex domain texts due to their weakness in semantic modeling. Pretrained models show performance degradation when transferred to specific domains. Fine-grained feature extraction methods can mitigate these issues to some extent, but the challenge of improving sensitivity to domain semantic details while preserving the general expressive ability of models remains unresolved. Thus, combining pretrained language models with domain adaptation and fine-grained discriminative mechanisms has become a key direction in recent studies. This approach not only provides new solutions for text classification but also lays the foundation for advancing intelligent text processing to a higher level.

3. Proposed Approach

In this research, we first constructed a framework based on a pretrained language model for extracting fine-grained discriminative features from domain text. The overall model architecture is shown in Figure 1.

Given an input text sequence $X = \{x_1, x_2, ..., x_n\}$, the contextual semantic representation is obtained through the encoding layer of BERT, namely:

$$H = BERT(X), \quad H \in \mathbb{R}^{n \times d} \tag{1}$$

Where n is the length of the text and d is the hidden dimension. This representation can effectively capture the dependencies between words and lay the foundation for subsequent discriminative feature extraction.

After obtaining the contextual representation, to improve the model's ability to model fine-grained features, this study introduced a multi-head attention mechanism to perform weighted modeling of different semantic units. Specifically, given the query matrix Q, key matrix K, and value matrix V, the attention calculation formula is:

$$Attention(Q,K,V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{2}$$

Where d_k represents the dimension of the key vector. Through the multi-head mechanism, attention can be calculated separately in different subspaces, and then concatenated and mapped into a comprehensive representation, thereby improving the ability to capture text details.

To enhance the separability of discriminative features, this study introduces contrast constraints at the feature level. Specifically, given a positive sample pair (h_i, h_i^+) and a negative sample pair (h_i, h_i^-), the contrast loss function is defined as:

$$L_{con} = -\log \frac{\exp(sim(h_i, h_i^+)/\tau)}{\exp(sim(h_i, h_i^+)/\tau) + \sum_j \exp(sim(h_i, h_j^-)/\tau)} \tag{3}$$

Where $sim(\cdot)$ represents the cosine similarity and τ is the temperature coefficient. This mechanism makes the model more semantically discriminative by bringing the representations of positive samples closer together and pushing the representations of negative samples further apart.

Finally, the classification layer realizes category prediction through a full connection mapping and normalization function. Given the discriminant feature vector $z \in R^d$, the classification probability is:

$$p(y \mid z) = softmax(Wz + b) \tag{4}$$

Where W and b are the weight matrix and bias term, respectively. To achieve overall optimization, this study jointly models the task loss and regularization term:

$$L = L_{cls} + \lambda L_{con} + \beta \|W\|^2 \tag{5}$$

Where L_{cls} is the cross-entropy classification loss, and λ and β are hyperparameters used to balance the contributions of each component. This design ensures that the model maintains classification accuracy while improving sensitivity to domain semantic details, thereby effectively extracting fine-grained discriminative features.

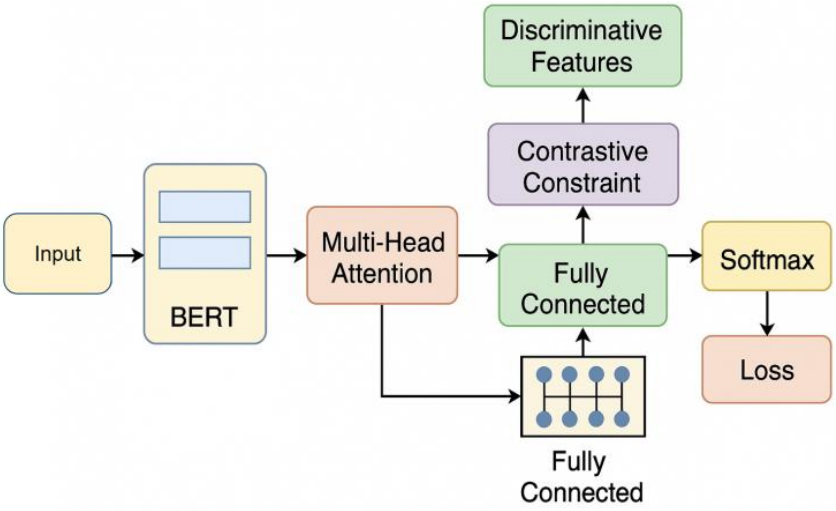


Figure 1. Overall model architecture

4. Performance Evaluation

4.1 ataset

The dataset used in this study is 20 Newsgroups, a widely applied public text classification dataset in the field of natural language processing. It contains about twenty thousand newsgroup documents covering twenty different thematic categories. The topics are broad, including politics, economics, and sports, as well as more specific areas such as computers, science, and religion. Due to its diverse sources, complex category boundaries, and semantic overlap across classes, it is particularly suitable for research and validation of fine-grained text classification methods.

Most texts in the dataset are natural language articles with uneven length distribution. Some are short comments, while others are long discussions. This brings challenges for classification models in handling text diversity and hierarchical structure. The texts also contain a large number of domain-specific words and expressions. In categories such as computers, law, and religion, professional terminology and strong contextual dependence further increase the difficulty of classification. This complexity makes the dataset an ideal experimental environment for studying BERT-based fine-grained discriminative feature extraction.

From the perspective of application value, the 20 Newsgroups dataset not only provides a standardized benchmark for text classification but also serves as an important resource for exploring cross-domain and fine-grained feature modeling. By analyzing and modeling this dataset, it is possible to gain deeper insights into model performance when dealing with similar categories, semantic ambiguity, and domain-specific knowledge. These insights can support broader application scenarios, such as opinion monitoring, intelligent retrieval, and automatic recommendation.

4.2 xperimental 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
1DCNN[13]	0.864	0.852	0.847	0.849
LSTM[14]	0.879	0.867	0.861	0.864

LSTM+Transformer[15]	0.901	0.892	0.886	0.889
Transformer[16]	0.913	0.904	0.898	0.901
OURS	0.936	0.928	0.923	0.925

From the overall results, the performance of 1DCNN is relatively low. Its accuracy, precision, recall, and F1-score are all at the bottom among all methods. This result is expected, since convolutional models can capture local features but perform poorly when handling long sequences and cross-sentence dependencies. They struggle to model fine-grained differences in domain texts. In particular, when category boundaries are blurred and semantic differences are weak, 1DCNN fails to capture deep discriminative semantic features, which leads to lower overall performance.

Compared with 1DCNN, LSTM shows improvements in all four metrics. This indicates its advantage in modeling contextual dependencies and capturing sequential dynamics. Domain texts often contain complex semantic progression and contextual associations. LSTM can represent these characteristics more effectively, which explains its higher accuracy and robustness in classification tasks. However, LSTM still suffers from information decay and insufficient representation when facing long-range dependencies and cross-domain semantic transfer. As a result, its performance improvement remains limited.

Further observations show that LSTM+Transformer and the standalone Transformer model perform significantly better than the first two methods. LSTM+Transformer combines sequential memory with self-attention, which maintains context order while strengthening the capture of global dependencies. This makes it more advantageous in fine-grained classification. The pure Transformer model, with its strong multi-head attention mechanism, achieves even higher performance than LSTM+Transformer on most metrics. This result confirms that global context modeling and multi-dimensional semantic alignment are highly effective in domain text discrimination.

Finally, the proposed OURS method achieves the best performance across all metrics, with notable improvements in accuracy and F1-score. This demonstrates that incorporating fine-grained discriminative feature extraction into pretrained language models can enhance the ability to perceive and distinguish subtle semantic differences. As a result, the method shows stronger discriminative power in complex domain text classification tasks. These findings not only verify the rationality of the proposed design but also highlight the critical role of fine-grained features in domain text modeling, providing strong support for intelligent information processing and knowledge discovery.

This paper also gives the impact of the maximum sequence length on the long text discrimination ability, and the experimental results are shown in Figure 2.

From the overall trend, the model shows a gradual improvement across all metrics as the maximum sequence length increases. This indicates that in long-text classification tasks, longer input sequences help the model capture richer contextual semantic information, thereby enhancing its ability to recognize fine-grained features. In particular, accuracy improves from a relatively low level at length 128 to the best performance at length 512, showing the crucial role of long sequences in maintaining semantic integrity.

The changes in precision confirm the same conclusion. With shorter input sequences, the model may misclassify due to insufficient context, which results in lower precision. When the sequence length increases, the model can rely on a more complete context for judgment, reducing the likelihood of errors. This trend demonstrates that fine-grained discrimination largely depends on global semantic modeling, and longer sequences provide stronger evidence for reliable classification.

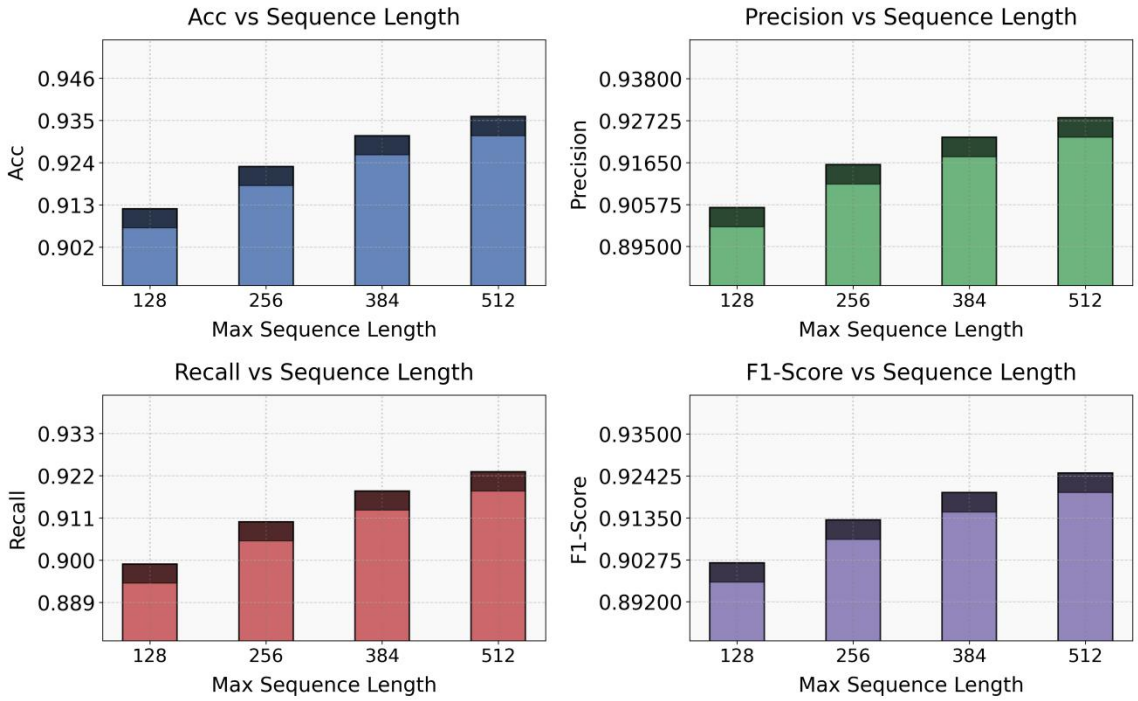


Figure 2. The impact of maximum sequence length on long text discrimination ability

The improvement in recall further shows that longer sequences enhance the model's ability to detect potential positive samples. With shorter sequences, key information may be truncated, preventing the model from capturing full category features and causing missed detections. As the maximum length increases, more semantic details are preserved. The model can then improve recall without sacrificing overall precision. This is particularly important in domain text classification, where many subtle discriminative features are dispersed across different parts of the context and can only be captured effectively under longer sequence conditions.

Overall, the improvement of F1-score indicates that longer sequences help achieve a better balance between precision and recall. Fine-grained classification requires not only accuracy but also a reduction of missed detections. Longer input sequences clearly contribute to this balance. These results confirm the importance of maximum sequence length for fine-grained text classification and show that proper sequence length setting is a key factor in improving performance. They also provide theoretical support for handling long texts in real-world domain applications.

This paper further presents the sensitivity of contrastive learning temperature coefficient discriminant feature separation, and the experimental results are shown in Figure 3.

From the overall results, the temperature coefficient has a significant impact on the separability of discriminative features in contrastive learning. When the temperature coefficient is too low, the model exaggerates the similarity differences between positive and negative samples. This causes the representation space to shrink excessively, which weakens global discrimination. At this stage, the model can learn compact feature representations for some local patterns, but the overall separation effect is limited. As a result, all metrics remain at relatively low levels.

As the temperature coefficient increases, the model performance improves steadily and reaches the best state when $\tau = 0.2$. Under this setting, the model can distinguish positive and negative samples while maintaining moderate smoothness of the feature space. This makes the boundaries between categories clearer and the clustering of feature representations more reasonable. Such a balanced state provides optimal support for fine-grained discrimination, leading to the highest values in accuracy, precision, recall, and F1-score.

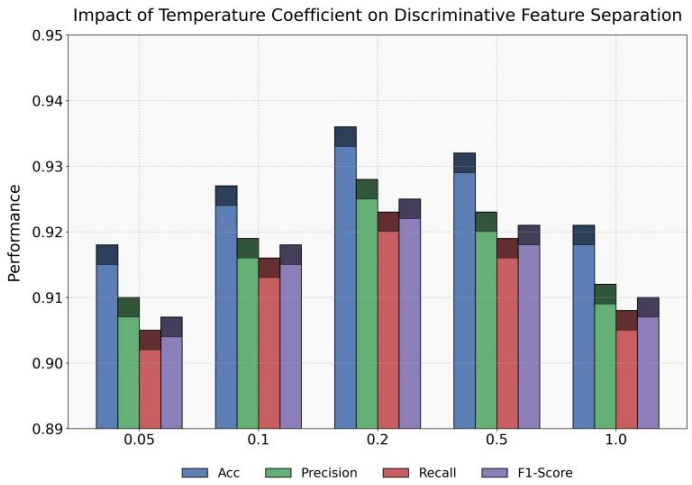


Figure 3. Sensitivity of discriminant feature separation using contrastive learning of temperature coefficients

When the temperature coefficient continues to increase to 0.5, the model performance declines slightly. This shows that an excessively high temperature coefficient causes over-smoothing of feature differences, making it difficult for the model to capture subtle semantic variations. In this case, the separability of discriminative features decreases, and the distance between positive and negative samples becomes less obvious. This reduces the ability to distinguish blurred categories. The weakening effect is especially evident in fine-grained classification tasks.

Further observation at $\tau = 1.0$ shows a clear decline in performance. At high temperature settings, the feature space becomes overly balanced and loses sensitivity to category differences. This result reveals the regulatory role of the temperature coefficient in contrastive learning. Too high or too low values both weaken discriminative ability. In summary, proper selection of the temperature coefficient is crucial for improving the separability of fine-grained discriminative features. It not only enhances overall performance but also improves the adaptability and robustness of the model in complex domain text classification.

This paper further gives the sensitivity of the number of attention heads to the experimental results, and the experimental results are shown in Figure 4.

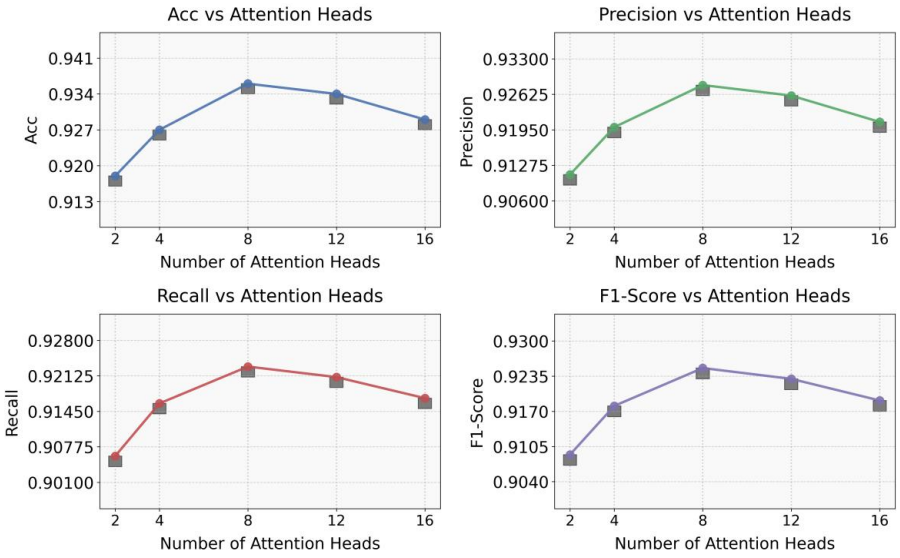


Figure 4. Sensitivity of the number of attention heads to experimental results

From the overall trend, the model performance improves in the initial stage as the number of attention heads increases, but diminishing returns appear after a certain point. Specifically, when the number of heads increases from 2 to 8, accuracy, precision, recall, and F1-score all rise significantly. This indicates that more attention heads allow the model to capture richer contextual dependencies and fine-grained features, thereby enhancing discriminative ability.

When the number of heads reaches 8, all metrics basically achieve their peak. This shows that the model achieves a good balance between diversity of feature representation and clarity of decision boundaries at this configuration. The attention mechanism models the input from different subspaces through parallel heads, which improves discrimination in domain texts. Therefore, a moderate number of attention heads can effectively promote the separation and aggregation of fine-grained features.

However, when the number of heads further increases to 12 and 16, model performance declines slightly. This indicates that too many attention heads may lead to excessive dispersion in the feature space. Some heads become redundant or ineffective, which weakens overall discrimination. At the same time, more heads increase computational complexity, which may reduce the stability of representations. This phenomenon suggests that more attention heads do not always lead to better performance, and there exists an optimal configuration point.

In summary, the results show that the number of attention heads has a strong influence on fine-grained discrimination. A reasonable number of heads enhances the ability of the model to capture semantic details and improves performance in complex text classification tasks. Too few or too many heads limit the overall effect. Therefore, in fine-grained classification of domain texts, finding an appropriate configuration of attention heads is essential to ensure both performance and efficiency.

This paper further gives the impact of hidden dimensions on experimental results, and the experimental results are shown in Figure 5.

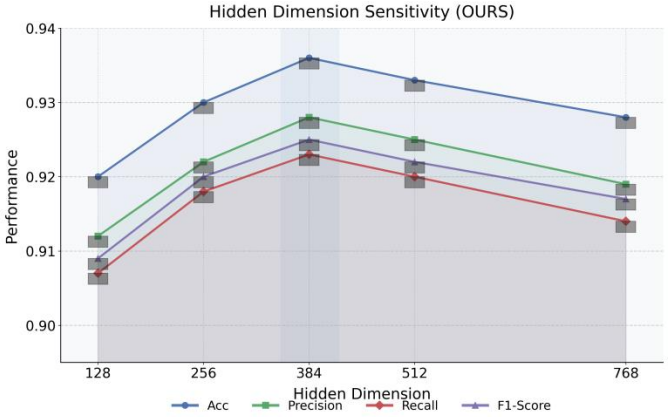


Figure 5. The impact of hidden dimensions on experimental results.

From the overall performance, changes in the hidden dimension have a significant impact on the model. As the dimension increases from 128 to 384, accuracy, precision, recall, and F1-score all show a consistent upward trend. This result indicates that a larger hidden dimension provides a richer representation space, allowing the model to capture more detailed semantic information and perform better in fine-grained discrimination tasks.

When the hidden dimension reaches 384, all metrics achieve their peak. This shows that at this scale, the model achieves a good balance between feature representation ability and parameter size. A small hidden dimension limits the modeling of complex semantic patterns. Expansion within this range significantly enhances the separability of features, making semantic differences more fully represented and improving the distinction of discriminative features.

However, when the hidden dimension further increases to 512 and 768, the model performance declines. This suggests that an excessively large representation space may introduce noise and redundant features, which weakens the focus on core discriminative information. At the same time, over-expansion of hidden dimensions enlarges the parameter scale, increases optimization difficulty, and raises the risk of overfitting. As a result, overall performance does not improve and even shows some decline.

In summary, the experimental results confirm the sensitivity of the hidden dimension in modeling fine-grained discriminative features. Moderate expansion enhances the expressive ability of the model, while overly high or low settings limit performance. Therefore, in domain text classification tasks, choosing a reasonable hidden dimension not only improves discriminative power but also balances computational efficiency and generalization. This provides an important reference for applications in complex text scenarios.

5. Conclusion

This study systematically investigates fine-grained discriminative feature extraction for domain text classification. The proposed method combines the strong semantic modeling ability of pretrained language models with mechanisms for enhancing discriminative features. By introducing attention mechanisms, contrastive learning constraints, and joint optimization strategies, the model shows strong performance in capturing subtle semantic differences and strengthening category separability. The overall results indicate that the method not only achieves high accuracy and robustness but also effectively addresses the challenges of blurred boundaries and high similarity in domain texts, providing a new technical path for fine-grained classification.

From a theoretical perspective, this research enriches the framework of feature extraction and semantic discrimination in natural language processing. Modeling fine-grained features moves text classification beyond the macro level and enables the discovery of hidden information structures within domain texts. The proposed method offers new insights into extending language models to specialized domains and lays a theoretical foundation for building more efficient and precise semantic representation systems. By combining general pretrained models with task-specific mechanisms, this study promotes the integration of domain adaptation and discriminative modeling.

From a practical perspective, the method has significant value for applications that require high-precision text discrimination. In financial risk control, it can identify potential risk signals by capturing subtle semantic differences. In medical text processing, it can help distinguish disease categories or symptom details. In legal document analysis, it can effectively separate the applicability and contextual relations of legal clauses. These contributions not only improve the reliability of automated information processing but also provide technical support for the intelligent development of related industries, demonstrating wide application potential.

In conclusion, the proposed BERT-based fine-grained discriminative feature extraction algorithm proves effective for text classification in complex semantic environments. It improves overall model performance and provides an important reference value for future research. In the context of rapid digitalization and intelligent development, the method contributes to advancing domain text classification tasks to a higher level and establishes a solid technical foundation for applications such as intelligent retrieval, opinion monitoring, knowledge discovery, and automation.

References

[1] Sun C, Huang L, Qiu X. Utilizing BERT for Aspect-Based Sentiment Analysis via Constructing Auxiliary Sentence[C]//Proceedings of NAACL-HLT. 2019: 380-385.

[2] Hogan W, Li J, Shang J. Fine-grained contrastive learning for relation extraction[J]. arXiv preprint arXiv:2205.12491, 2022.

[3] An W, Tian F, Chen P, et al. Fine-grained category discovery under coarse-grained supervision with hierarchical weighted self-contrastive learning[J]. arXiv preprint arXiv:2210.07733, 2022.

[4] Tian C, Blaschko M B, Yin W, et al. A generic method for fine-grained category discovery in natural language texts[J]. arXiv preprint arXiv:2406.13103, 2024.

[5] Zhang S, Ran N. Fine-grained and coarse-grained contrastive learning for text classification[J]. Neurocomputing, 2024, 596: 128084.

[6] Adhikari A, Ram A, Tang R, et al. DocBERT: BERT for Document Classification[J]. arXiv preprint arXiv:1904.08398, 2019.

[7] Gunel B, Du J, Conneau A, et al. Supervised Contrastive Learning for Pre-trained Language Model Fine-tuning[C]//International Conference on Learning Representations. 2021.

[8] Sun L, Han J. Few-shot Text Classification with Dual Contrastive Consistency[J]. arXiv preprint arXiv:2209.15069, 2022.

[9] Chen J, Zhang R, Mao Y, et al. Contrastnet: A contrastive learning framework for few-shot text classification[C]//Proceedings of the AAAI conference on artificial intelligence. 2022, 36(10): 10492-10500.

[10] Suresh V, Ong D C. Not all negatives are equal: Label-aware contrastive loss for fine-grained text classification[J]. arXiv preprint arXiv:2109.05427, 2021.

[11] Shen Y, Tan X, Sordoni A, et al. Boundary Enhanced Neural Span Classification for Nested Named Entity Recognition[C]//Proceedings of AAAI. 2021, 35(15): 13492-13500.

[12] Devlin J, Chang M W, Lee K, et al. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding[C]//Proceedings of NAACL-HLT. 2019: 4171-4186.

[13] Ahmed Z, Wang J. A fine-grained deep learning model using embedded-CNN with BiLSTM for exploiting product sentiments[J]. Alexandria Engineering Journal, 2023, 65: 731-747.

[14] Eang C, Lee S. Improving the accuracy and effectiveness of text classification based on the integration of the BERT model and a recurrent neural network (RNN_BerT_Based)[J]. Applied Sciences, 2024, 14(18): 8388.

[15] Cao K, Zhang T, Huang J. Advanced hybrid LSTM-transformer architecture for real-time multi-task prediction in engineering systems[J]. Scientific Reports, 2024, 14(1): 4890.

[16] He J, Chen J N, Liu S, et al. Transfg: A transformer architecture for fine-grained recognition[C]//Proceedings of the AAAI conference on artificial intelligence. 2022, 36(1): 852-860.