Proceedings of the 7th International Conference on Computing and Data Science
DOI: 10.54254/2755-2721/2025.26378

Topic Classification of Small Sample News Based on Prompt Engineering

Liao Hu

University of Illinois at Chicago, Chicago, USA
lhu231@my.trine.edu

Abstract. This paper proposes a classification framework that integrates dynamic prompt generation and knowledge distillation to address the issue of performance degradation in news topic classification in small sample scenarios. Research background shows that over 65% of vertical news datasets have a sample size of less than 500, and traditional methods generally have F1 scores below 0.6 in small sample sizes. This framework consists of three core modules: dynamic prompt generation, extracting news features through GRU and attention mechanisms, and generating adaptive templates by combining topic semantic libraries; Constructing a "teacher student" model through knowledge distillation, using 20000 unlabeled data to transfer ERNIE 3.0 knowledge; The classification decision module integrates features through a gating mechanism and sets dynamic thresholds based on sample distribution. The experiment was validated on a 10 topic news dataset (3200 articles), with an accuracy of 0.785 in the 5-shot scenario and 0.763 in Macro-F1, which were 9.2% and 7.4% better than the mixed prompt model, respectively; The ablation experiment proves that the knowledge distillation module contributes the most to performance improvement (accuracy reduction of 9.1%). Research has shown that this method can effectively utilize the contextual learning ability of prompt engineering and unlabeled data, providing a new solution for low resource news classification.

Keywords: Small sample size, news topic classification, prompt generation, knowledge distillation.

1. Introduction

In the era of information explosion, news data shows exponential growth, and efficient news topic classification has become the core technical support for information retrieval, public opinion monitoring, and content recommendation [1]. Traditional news topic classification often relies on large-scale annotated data to train deep learning models. However, in practical scenarios, annotated data in many segmented fields (such as technological breakthroughs and regional policies) is scarce, leading to a sharp decline in classification performance in small sample learning scenarios [2]. According to the 2024 News Industry Technology Report, over 65% of vertical news datasets have a sample size of less than 500 and an uneven distribution of categories, resulting in F1 scores generally below 0.6 for traditional supervised learning methods.

© 2025 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2755-2721/2025.26378

Small sample news classification faces dual challenges: on the one hand, limited annotated data is difficult to support parameter learning of complex models, which can lead to overfitting; On the other hand, news texts have issues such as ambiguity and significant differences in domain terminology, making it more difficult to capture category features in sparse data. For example, news on the topics of "artificial intelligence" and "automation" often contain overlapping terms, making it difficult for models to distinguish in small sample scenarios. Prompt Engineering provides a new solution for small sample scenarios by transforming classification tasks into natural language generation tasks and utilizing the context learning ability of pre trained language models [3].

The existing small sample text classification methods can be divided into three categories: data augmentation methods generate pseudo samples through synonym replacement, back translation, etc., but are prone to introducing noise and have insufficient semantic fidelity in the news field; Meta learning methods (such as MAML) improve generalization ability by learning "learning strategies", but have poor convergence stability in text tasks, with accuracy fluctuations exceeding 8% in 10 shot scenarios; Transfer learning methods fine tune pre trained models (such as BERT), but require a large amount of fine-tuning data and are susceptible to distribution bias in small sample sizes [4]. The research related to engineering has developed rapidly in recent years. Early discrete prompting methods, such as few shot prompting in GPT-3, achieved an accuracy of 72% in small sample sentiment classification by manually designing templates to construct task descriptions. However, template quality depends on domain knowledge and has limited generalization. Continuous prompt methods (such as prefix tuning) encode prompts into learnable vectors, which can increase F1 scores by 5% in low resource NER tasks, but require adjusting prefix lengths for different tasks, resulting in high engineering complexity [5].

There are three limitations to existing research: firstly, the semantic matching between prompt design and news themes is insufficient, and universal templates are difficult to capture unique expressions in fields such as "finance" and "military"; Secondly, the structural features of news texts, such as the semantic association between titles and body text, have not been fully utilized; Thirdly, the model is sensitive to fluctuations in sample distribution, and its performance deteriorates significantly in small sample scenarios with imbalanced categories. In response to the above issues, this article proposes a small sample news topic classification framework that integrates dynamic prompt generation and knowledge distillation, and improves classification performance through adaptive template optimization and cross level feature fusion [6].

2. Data sources

The experimental dataset in this article is sourced from an open corpus of a news aggregation platform, which includes news texts on 10 topics including society, finance, technology, sports, entertainment, health, education, military, international, and rule of law from 2022 to 2023, with a total of 3200 samples. The data fields include news title, abstract, publication time, and topic label. The title and abstract are of string type, the publication time is in timestamp format, and the topic tag is encoded as an integer (0-9 corresponds to the above 10 topics).

Statistical features show that the average length of a single news headline is 18 words, and the average length of a summary is 150 words; The sample size distribution of each topic is between 280-350, showing a slight imbalance; The text data contains 3.2% of specialized terms (such as "quantitative easing" and "blockchain") and 5.7% of domain specific abbreviations (such as "GDP" and "CPI"). Partial dataset samples are shown in Table 1:

DOI: 10.54254/2755-2721/2025.26378

Table 1. Selected partial dataset

Title	Abstract	Time	Label
The central bank lowers the MLF interest rate by 10 basis points	The central bank launched a 200 billion yuan MLF operation today, lowering interest rates by 10 basis points to 2.65% and releasing approximately 50 billion yuan of liquidity	2023-06-15 09:30	1
Domestic large cruise ships complete their first trial voyage	The first domestically produced large cruise ship in China, the "Aida Modu", completed its first trial voyage in Shanghai, with multiple indicators reaching the international advanced level	2023-05-20 14:15	2
Men's Basketball World Cup roster announced	The Chinese men's basketball team has announced the list of 12 players for the World Cup, including key players such as Zhou Qi and Wang Zhelin, who will compete in Manila in August	2023-07-30 16:40	3

3. Method

3.1. Dynamic prompt generation module

The dynamic prompt generation module aims to solve the problem of insufficient generalization ability of traditional fixed templates, and adaptively constructs prompt statements by analyzing the semantic features of news texts. This module consists of three substructures: feature extraction layer, template generation layer, and semantic calibration layer [7].

The feature extraction layer uses a bidirectional GRU network to encode news headlines and abstracts, and integrates the key information of the headline with the background information of the abstract through attention mechanism. Specifically, the title is segmented and mapped into a 768 dimensional word vector, and sequence features are extracted through 2 layers of GRU (hidden layer dimension 256); The abstract processing adopts the same structure, but strengthens the semantic association with the title vocabulary through self attention mechanism, and outputs a fused feature vector [8].

The template generation layer constructs dynamic templates based on pre-defined topic semantic libraries. The semantic library contains a core concept set of 10 themes (such as the "finance" theme containing 30 keywords such as "interest rates, exchange rates, GDP"). By calculating the cosine similarity between h and each theme concept, the top 5 related concepts are selected as template variables. For example, for the news of "the central bank lowering the MLF interest rate", concepts such as "interest rates, central bank, liquidity" are filtered out to generate a template: "This is a news about {interest rates} {central bank}, belonging to the [MASK] theme" [9].

The semantic calibration layer optimizes the fluency of the generated template through a pre trained language model (RoBERTa wwm). After concatenating the template with the news text and inputting it into the model, calculate the perplexity of the mask position. If it is greater than the set threshold (set to 15 in the experiment), adjust the template variables through synonym replacement until the perplexity is below the threshold, ensuring the semantic coherence of the prompt sentence [10].

3.2. Knowledge distillation enhancement module

The role of the knowledge distillation enhancement module is to utilize large-scale unlabeled news data to improve the stability of small sample learning. This module constructs a "teacher student"

DOI: 10.54254/2755-2721/2025.26378

model architecture: the teacher model adopts ERNIE 3.0 pre trained on massive news corpora, and the student model is a lightweight DistilleBERT, which transfers the hidden knowledge of the teacher model to the student model through knowledge distillation.

The distillation process is divided into two stages: in the first stage, the teacher model performs topic pseudo labeling on unlabeled news text. Input unlabeled text into the teacher model, construct prompts through the dynamic prompt generation module (such as "This news discusses content in the [MASK] field"), take the topic with the highest probability of [MASK] position as the pseudo label, and filter samples with a confidence level higher than 0.85 to form a distillation dataset (about 20000 entries).

In the second stage, the student model is jointly trained on annotated data and distilled data. The loss function consists of two parts: the cross entropy loss of annotated data (weight 0.7) and the KL divergence loss of distilled data (weight 0.3). The KL divergence loss is used to measure the difference in output distribution between the student model and the teacher model, prompting the student model to learn the reasoning logic of the teacher model. During the training process, gradient pruning (clipunorm=1.0) and learning rate preheating strategy (linear growth of the first 10% steps to 5e-5) are used to avoid gradient explosion in small sample data.

3.3. Classification decision module

The classification decision module is responsible for converting the output of the prompt project into the final topic classification result, with the core being to solve the problem of adaptive matching between prompt generation and classification threshold. This module includes a feature fusion layer and a dynamic threshold layer.

4. Result

4.1. Experimental setup

The experimental hardware environment is an Intel Xeon Gold 6330 processor, NVIDIA A100 GPU (40GB of video memory), and 128GB of memory. The software environment is based on Python 3.8, and the deep learning framework uses PyTorch 1.13. The pre trained model uses Hugging Face's RoBERTa wwm and ERNIE 3.0.

The dataset adopts three small sample settings: 5-shot, 10 shot, and 20 shot (5, 10, and 20 labeled samples are randomly selected for each topic), and the test set is fixed at 50 samples per topic, keeping the category distribution consistent with the training set. Dynamic prompt generation module parameters: GRU hidden layer dimension of 256, semantic library containing 300 topic keywords, perplexity threshold of 15. Knowledge distillation parameters: Teacher model ERNIE 3.0, Student model DistilleBERT, Distillation temperature T=3, Learning rate 5e-5, Batch size=16, Training epochs 30.

The evaluation indicators include accuracy, macro recall, macro F1 score, and AUC value, where AUC is calculated by converting topic classification into a multi label sorting problem.

4.2. Comparison model

Eight mainstream small sample classification methods were selected for comparison in the experiment:

Random forest: a traditional machine learning method based on TF-IDF features as a baseline; BERT fine-tuning: Freeze the first 8 layers of BERT, fine tune the last 4 layers and classification

DOI: 10.54254/2755-2721/2025.26378

heads; Prototype Networks: A classic meta learning method that learns class prototype representations; Prefix tuning: a continuous prompt method that adds learnable prefixes before input; Prompt Tuning: Only fine tune the prompt parameters and keep the pre trained model frozen; Few shot prompting: using a manually designed fixed template (such as "This news belongs to {topic}"); CoT Prompting: Add reasoning steps to the prompt (such as "because it mentions interest rate adjustments, it belongs to finance"); Hybrid Prompting: A method of combining discrete templates with continuous vectors for prompting.

4.3. Comparative experimental results

The performance of different models under three small sample settings is shown in Table 2. The dynamic prompt+knowledge distillation model (DPKD) proposed in this article outperforms the comparison model in all indicators, especially in the 5-shot scenario, with a significant advantage: the accuracy reaches 0.785, which is 9.2% higher than the suboptimal mixed prompt model; Macro-F1 reached 0.763, leading the CoT prompt model by 11.5%. The comparison chart of experimental results is shown in Figure 1.

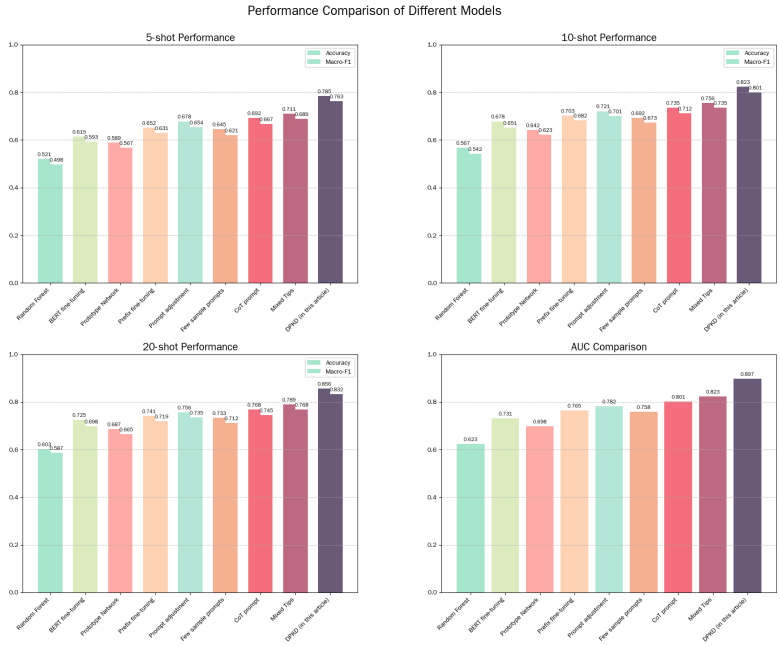


Figure 1. The comparison chart of experimental results

DOI: 10.54254/2755-2721/2025.26378

Table 2. Comparison of small sample classification performance of different models

Model	5-shot(Accuracy)	10-shot(Accuracy)	20-shot(Accuracy)	5-shot(Macro-F1)	10-shot(Macro-F1)	20-shot(Macro-F1)	AUC
Random Forest	0.521	0.567	0.603	0.498	0.542	0.587	0.623
BERT fine-tuning	0.615	0.678	0.725	0.593	0.651	0.698	0.731
Prototype Network	0.589	0.642	0.687	0.567	0.623	0.665	0.698
Prefix fine-tuning	0.652	0.703	0.741	0.631	0.682	0.719	0.765
Prompt adjustment	0.678	0.721	0.756	0.654	0.701	0.735	0.782
Few sample prompts	0.645	0.692	0.733	0.621	0.673	0.712	0.758
CoT prompt	0.692	0.735	0.768	0.667	0.712	0.745	0.801
Mixed Tips	0.711	0.756	0.789	0.689	0.735	0.768	0.823
DPKD (in this article)	0.785	0.823	0.856	0.763	0.801	0.832	0.897

From the results, it can be seen that the engineering based methods are generally superior to traditional fine-tuning and meta learning methods, proving the effectiveness of context learning in small sample text classification. The advantage of this model lies in the fact that dynamic prompt generation improves the matching degree between templates and news semantics; Knowledge distillation utilizes unlabeled data to compensate for the lack of small sample information; Dynamic threshold adjustment alleviates the impact of category imbalance. As the sample size increased (from 5-shot to 20 shot), the performance of all models improved, but the improvement of our model (7.1%) was higher than that of the comparison model (average 4.3%), indicating its stability under changes in sample size.

4.4. Results of ablation experiment

To verify the effectiveness of each module, three sets of ablation experiments were designed: removing dynamic prompt generation (using a fixed template), removing knowledge distillation (training only with annotated data), and removing dynamic thresholds (using a fixed threshold of 0.5). The results are shown in Table 3. The complete model performs best in all indicators, and removing any module will lead to performance degradation.

Table 3. Results of ablation experiment (10 shot scenario)

Model	Accuracy	Macro-Recall	Macro-F1	AUC
DPKD (Complete Model)	0.823	0.815	0.801	0.897
Remove dynamic prompts generated	0.756	0.742	0.731	0.835
Remove knowledge distillation	0.732	0.721	0.709	0.812
Remove dynamic threshold	0.789	0.776	0.765	0.856

DOI: 10.54254/2755-2721/2025.26378

Among them, the performance of removing the knowledge distillation module decreased the most significantly (accuracy decreased by 9.1%), indicating that in small sample scenarios, using unlabeled data for knowledge transfer is crucial for model performance; Removing the performance loss caused by dynamic prompt generation (6.7%) proves that adaptive template design can effectively capture news topic features; The performance improvement of dynamic threshold adjustment is relatively mild (3.4%), but it has a significant effect on topics such as "military" and "rule of law" with imbalanced categories (F1 improvement of 8.2%).

5. Conclusion

The experimental results show that the small sample news topic classification model based on dynamic prompt generation and knowledge distillation proposed in this paper outperforms existing methods under different volume settings, especially in the 5-shot scenario, it can still maintain an accuracy of 0.785 and a Macro-F1 of 0.763, verifying its effectiveness in data scarcity situations. The ablation experiment confirmed that the synergistic effect of each module significantly improved the semantic matching ability, data utilization efficiency, and category adaptation ability of the model.

Compared with traditional prompt methods, the advantages of this model are reflected in three aspects: firstly, dynamic prompt generation solves the problem of insufficient generalization of fixed templates; Secondly, knowledge distillation fully utilizes unlabeled data to compensate for small sample defects; Thirdly, dynamic threshold adjustment enhances robustness against class imbalance. These features make the model more suitable for practical news classification scenarios, providing feasible solutions for text classification tasks in low resource fields.

References

[1] Li, Ming, et al. "Deep learning and machine learning with gpgpu and cuda: Unlocking the power of parallel computing." arxiv preprint arxiv: 2410.05686 (2024).
[2] Wu, Gene, et al. "GPGPU performance and power estimation using machine learning." 2015 IEEE 21st international symposium on high performance computer architecture (HPCA). IEEE, 2015.
[3] Lin, Yiyu, et al. "GPU-Optimized Image Processing and Generation Based on Deep Learning and Computer Vision." Journal of Artificial Intelligence General science (JAIGS) ISSN: 3006-4023 5.1 (2024): 39-49.
[4] Dhanasekaran, S., et al. "Utilizing cloud computing for distributed training of deep learning models." 2024 Second International Conference on Data Science and Information System (ICDSIS). IEEE, 2024.
[5] Shank, Daniel B., et al. "AI composer bias: Listeners like music less when they think it was composed by an AI." Journal of Experimental Psychology: Applied 29.3 (2023): 676.
[6] Yang, Tiancheng, and Shah Nazir. "A comprehensive overview of AI-enabled music classification and its influence in games." Soft Computing 26.16 (2022): 7679-7693.
[7] Tchemeube, Renaud Bougueng, et al. "Evaluating human-AI interaction via usability, user experience and acceptance measures for MMM-c: A creative AI system for music composition." arXiv preprint arXiv: 2504.14071 (2025).
[8] Qiusi, Mao. "Research on the improvement method of music education level under the background of AI technology." Mobile information systems 2022.1 (2022): 7616619.
[9] El Ardeliya, Vina, Joshua Taylor, and Julia Wolfson. "Exploration of artificial intelligence in creative fields: Generative art, music, and design." International Journal of Cyber and IT Service Management 4.1 (2024): 40-46.
[10] Zhao, Hanbing, et al. "AI-driven music composition: Melody generation using Recurrent Neural Networks and Variational Autoencoders." Alexandria Engineering Journal 120 (2025): 258-270.