Proceedings of the 5th International Conference on Signal Processing and Machine Learning
DOI: 10.54254/2755-2721/108/2025.20922

Research on News Text Classification Method Based on Hybrid Neural Networks

Meijie Zhao[1,a,*], Guozhu Liu[1,b]

[1]School of Information Science and Technology, Qingdao University of Science and Technology, Qingdao, Shandong, 266061, China
a. 15053417022@163.com, b. lgz_0228@163.com
**corresponding author*

Abstract: With the rapid advancement of Internet technologies, news websites and online media platforms have become crucial components of the news dissemination landscape. Online news has also become one of the primary sources for people to obtain information. To meet the needs of numerous internet news readers and ensure that content providers can efficiently deliver personalized news recommendations, the effective control and utilization of internet news information have become essential. In this context, a news text classification method based on hybrid neural networks has emerged. Addressing the limitations of E-TF-IDF in neglecting positional relationships, this paper proposes a function that incorporates positional variables, denoted as P, based on readers' habits and the normal distribution. This study integrates the advantages of Bert, CNN, and BiLSTM. First, the preprocessed data is fed through the E-TF-IDF-P algorithm into the Bert model. Then, local features are extracted via TextCNN, and global features are captured via BiLSTM. These features are fused and input into the softmax layer for text classification.

Keywords: Neural Networks, Text Classification, Convolution, Natural Language Processing.

1. Introduction

With the continuous advancement of internet technology, news websites and online media platforms have become essential components of news dissemination and the primary channels for people to access information. While these platforms release massive amounts of news daily to meet users' diverse needs, they also pose higher demands for personalized recommendation and efficient management of news content. To achieve accurate news delivery and efficient management, it is crucial to leverage advanced technologies to deeply mine and analyze news data. As a key process, news text classification plays a vital role in improving the accuracy of news recommendations and user satisfaction. Traditional news text classification methods often rely on manually defined features, which are time-consuming, labor-intensive, and struggle to handle the diversity and complexity of news content. Therefore, exploring more efficient and intelligent news text classification methods is of paramount importance.

In recent years, the rise of deep learning technologies has provided new approaches for news text classification [1]. Hybrid neural networks, which combine multiple deep learning models, can harness the strengths of each model to improve classification accuracy and efficiency. However, existing hybrid neural network models for news text classification still have limitations, such as incomplete

DOI: 10.54254/2755-2721/108/2025.20922

feature extraction and high model complexity. Addressing these issues through improvement and optimization to build a more suitable hybrid neural network model for news text classification holds significant research and practical value. Zhou Chunting [2] and colleagues constructed a C-LSTM model based on CNN and RNN approaches. This model effectively classifies complex sentences and related words, enhancing the understanding of intricate linguistic contexts. By leveraging CNN technology, valuable words are extracted and transformed into more precise sentences, thus better describing complex sentences.

Based on existing research, this paper proposes a news text classification method based on hybrid neural networks. This method combines the advantages of the E-TF-IDF-P algorithm, the Bert model [3], TextCNN, and BiLSTM, aiming to improve feature extraction and model structure for higher classification accuracy and efficiency. By integrating deep learning and machine learning techniques, the approach utilizes the feature extraction capabilities of deep learning and the stability of traditional machine learning to achieve more robust and efficient classification performance. This method is expected to provide a more efficient and intelligent solution for news text classification, supporting personalized recommendations and efficient management for news platforms. Additionally, this research offers new insights and references for the application of deep learning technologies in natural language processing.

2. Related Work

TextCNN was first proposed by Yoon Kim [4] to address the issue that traditional CNNs cannot be directly applied to text classification. TextCNN is a convolutional neural network (CNN) technique based on a “one-dimensional image” [5]. It can effectively detect parts of the text content and classify them accordingly. The core structure of TextCNN consists of an embedding layer, a convolutional layer, a pooling layer, and a fully connected layer. The embedding layer converts words in the text into high-dimensional vectors for processing by the neural network. By projecting several n-gram features through convolutional kernels of different sizes, a sliding window approach is used to capture more information. To improve model accuracy, pooling is applied to some key information, thereby reducing model complexity. After processing through the entire network, the model can be used for text classification. The TextCNN model demonstrates efficiency and good performance in text classification tasks, with advantages such as simple network architecture, fewer parameters, fast computation, minimal time consumption, and low resource requirements. TextCNN is an efficient and flexible text classification model that effectively captures local features in the text and transforms them into global representations, enabling more accurate and efficient classification. Additionally, it has excellent scalability and can be easily applied to large datasets and complex classification tasks.

BERT was first introduced by the Google AI Language team in the paper “BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding” [6]. BERT stands for Bidirectional Encoder Representations from Transformers, representing a novel method for text representation. It features the ability to integrate different contextual structures, achieving efficient bidirectional encoding of unlabeled data. BERT utilizes two modes: Masked Language Modeling (MLM) and Next Sentence Prediction (NSP), which overcome the limitations of single-mode approaches, significantly enhancing the efficiency of fine-tuning tasks. BERT is regarded as the first model to describe complex situations with minimal changes, offering high processing efficiency for complex word-level tasks. The BERT model architecture uses a comprehensive multi-layer bidirectional Transformer encoding technique. Its output clearly refers to a token sequence, thereby better capturing information between pairs or groups of tokens. This technology, leveraging WordPiece embeddings, supports 11 different natural language processing tasks and provides more accurate predictions and efficient solutions. Notable improvements include a 7.7% increase in the GLUE score, a 4.6% increase in MultiNLI accuracy, an F1 score of 93.2 for the SQuAD v1.1 question-answering task (an absolute improvement

DOI: 10.54254/2755-2721/108/2025.20922

of 1.5 points), and an F1 score of 83.1 for the SQuAD v2.0 task (an absolute improvement of 5.1 points). BERT's introduction demonstrates the importance of bidirectional pre-training for language representations and reduces the need for carefully designed task-specific architectures. It has significantly advanced the field of natural language processing (NLP) and provided critical insights and references for subsequent research.

As research into combined models deepens, their applications are expanding across various domains. Gao Kaiyue et al. [7] proposed a PCC-BiLSTM-GRU-Attention hybrid prediction model, which shows high accuracy and stability in multivariate time-series prediction tasks. This model improves prediction performance by integrating Pearson Correlation Coefficient (PCC) for feature selection, BiLSTM for bidirectional sequential feature extraction, and GRU fused with an attention mechanism. Shen Qing et al. [8] proposed a BERT-TextCNN model enhanced with LDA (Latent Dirichlet Allocation) to address the classification of live-streaming bullet comments in e-commerce. Experiments verified the effectiveness of this model.

3. Main Research Content

3.1. E-TF-IDF Algorithm

TF-IDF (Term Frequency–Inverse Document Frequency) is a statistical method used to evaluate the influence of a term within an entire document. The formula is as follows:

$$D_{\mathrm{x}} = \frac{\sum_{j=1}^{N} (n_{w_i} - n_{w_i}^{j})^{2}}{N}$$

$$P_{\mathrm{x}} = \sqrt{2\pi}\mathrm{ne}^{\frac{\mathrm{x\text{-}m}^{2}}{2\mathrm{n}^{2}}}$$

$$E - TF - IDF - P = \sum_{\mathrm{x}=1}^{\mathrm{x}} \mathrm{tf} * \mathrm{idf} * D_{\mathrm{x}} * P_{\mathrm{x}}$$

Where: D_x represents the term frequency of word W_i in category; N is the total number of categories in the document; n_{w_i} represents the number of documents in the entire corpus that contain word W_i; $n_{w_i}^{j}$ represents the number of documents in category j that contain word W_i. In this study, m is the mean, n is the variance, and P is the positional density function.

3.2. Model Description

BERT can capture rich semantic information, providing powerful feature representations for downstream tasks. By using a bidirectional Transformer, BERT simultaneously considers contextual information from both directions, enhancing the model's understanding of text. BiLSTM can handle global sequential data and performs well in natural language processing tasks. It captures long-term dependencies effectively, making it suitable for tasks that require memory. However, when processing long texts, BiLSTM can suffer from issues such as gradient vanishing or explosion. TextCNN [9], by employing convolutional kernels of different sizes, can automatically extract features of various lengths from the text, thereby capturing local semantic information. This makes TextCNN highly efficient in processing textual data. The combined model integrating BERT, TextCNN, and BiLSTM is shown in Figure 1. The processing flow mainly consists of the following three parts:

DOI: 10.54254/2755-2721/108/2025.20922

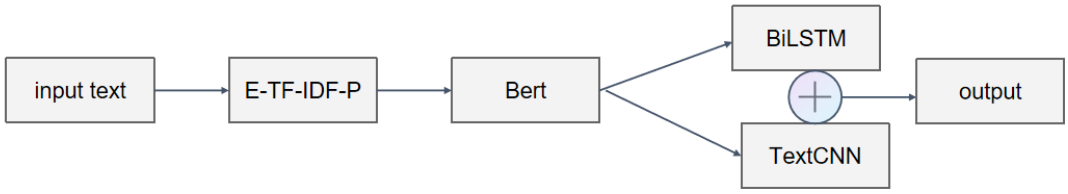

Figure 1: The specific process

3.2.1. BERT Model

The output of the E-TF-IDF-P algorithm is ranked according to specific scores, and high-scoring words are selected to form a new vector, which serves as the input to the Transformer in the BERT model. Each Transformer encoder unit consists of a multi-head attention mechanism, Layer Normalization, a feedforward network, and another Layer Normalization stacked together (as shown in Figure 2). In each self-attention calculation, three intermediate weight matrices — w_Q, w_k, W_v — are used to linearly transform the input X into three new tensors: query, key, and value. Specifically, the input matrix X is multiplied by w_Q, w_k, W_v to generate Q, K, V. The relevance between words in X is then calculated by multiplying Q and K^T matrices, as shown in Equation (3.2.1.1):

$$Attention\quad Q,K,V\ = \text{softmax}\left(\frac{QK^{T}}{\sqrt{d_k}}\right)V \tag{3.2.1.1}$$

Where d_k is the dimensionality of the vectors Q and K.

The BERT model consists of 6 encoders in the encoding component (Figure 3), while the decoding component also has the same number of decoders. Each encoder performs the same function and only requires a specific signal to complete the processing.

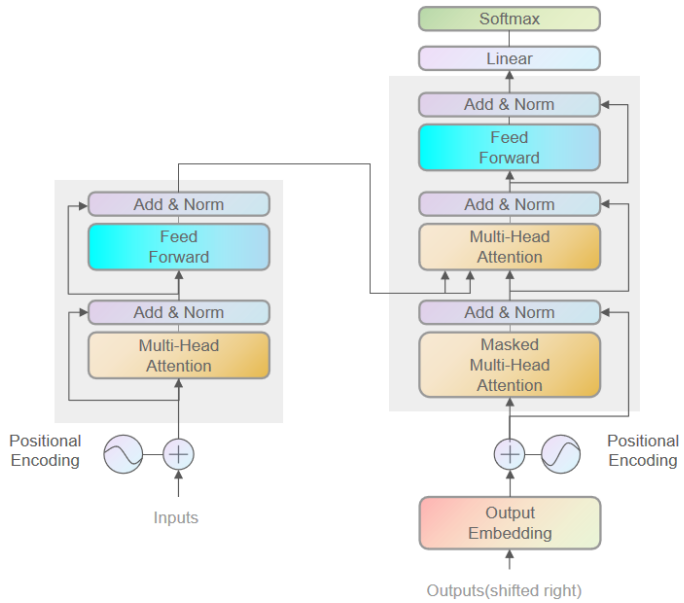

Figure 2: Model Structure

DOI: 10.54254/2755-2721/108/2025.20922

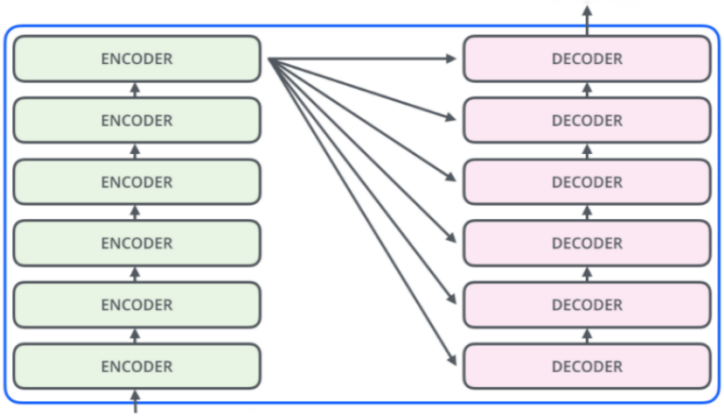

Figure 3: Encoding Component

The BERT model has two pre-training tasks: Masked Language Modeling (MLM) and Next Sentence Prediction (NSP). MLM randomly masks some words and predicts them based on their contextual positions. In BERT, 15% of WordPiece tokens are randomly masked. Among these tokens: 80% are replaced with the [MASK] token, 10% are replaced with a random word, 10% remain unchanged. In the NSP task, 50% of sentence pairs in the corpus are consecutive sentences (positive samples), while the other 50% are random sentence pairs (negative samples).

Add & Norm: The residual network is an effective topological structure that uses residual connections to simplify complex models. This reduces model complexity, enhances model stability, and improves the overall model architecture. The core idea of the Transformer is to combine the outputs of the self-attention mechanism and the feedforward neural network with the original input to produce a residual connection. To achieve this, Layer Normalization is applied as a mathematical preprocessing step, converting the original input into a format that the model can process effectively, thereby enhancing the representation of model results. Transformers use layer normalization to standardize the covariates across multiple neural networks, improving model performance by reducing covariate shift and enhancing accuracy. In a Transformer, layer normalization is typically applied after the residual connection.

3.2.2. TextCNN Model

The TextCNN model consists of an input layer, convolutional layer, pooling layer, fully connected layer, and output layer. The output of the Transformer encoder in the BERT model provides word vectors rich in semantic features as input to TextCNN. To improve model accuracy, max pooling is used. The pooled features from different channels are concatenated and fed into the fully connected layer, which further combines these feature vectors into a vector representation of the classification categories.

3.2.3. BiLSTM Model

The BiLSTM model (Figure 4) is composed of two LSTM networks: one processes the input sequence forward, and the other processes it backward. To obtain accurate BiLSTM results, multiple iterations are required, with precise calculations at each step. After six iterations of the LSTM algorithm, two output vectors are produced: one positive and one negative. By combining these two values through iterative calculations, the final BiLSTM output is obtained.

DOI: 10.54254/2755-2721/108/2025.20922

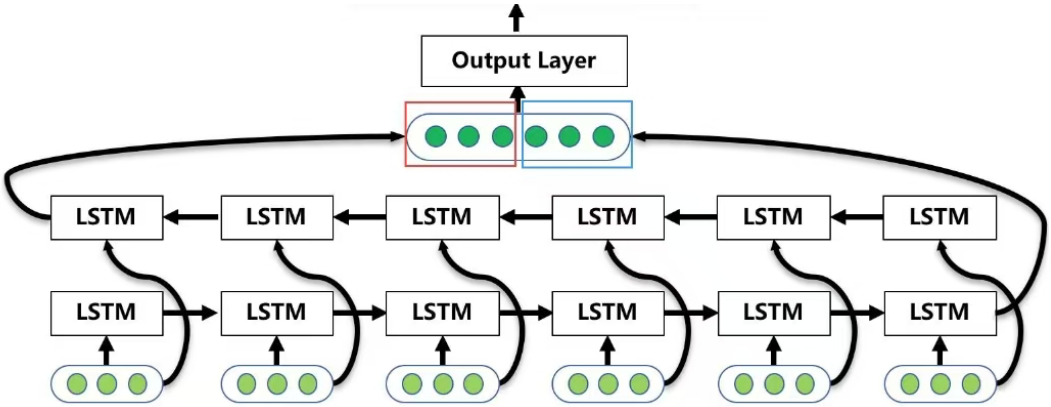

Figure 4: BiLSTM Model

4. Experimental Setup and Results Analysis

4.1. Experimental Dataset

The experimental data used in this study come from the NetEase News dataset and the THUCNews dataset compiled by Tsinghua University. These datasets cover various types of news data, and the ratio of training to validation sets is set at 8:2. The THUCNews dataset, published by the Tsinghua University Natural Language Processing and Social Humanities Computing Laboratory, includes 14 categories such as finance, real estate, and technology. A subset of approximately 65,000 data samples was selected for this study. The NetEase News dataset includes news data from five categories — domestic, international, military, aviation, and technology — comprising approximately 30,000 data samples.

4.2. Comparative Experiments

To verify the advantages of the proposed model in news text classification, the following three models were used for comparison: E-TF-IDF-BERT with LSTM and CNN Hybrid Model: The E-TF-IDF algorithm preprocesses the data, and the output vectors from BERT are fed into LSTM and CNN to extract both local and global features. These features are then fused for classification. LSTM-CNN Model: Word vectors preprocessed by Word2Vec are first input into the LSTM model for global feature extraction. The resulting vectors are combined with the original features and input into the CNN model. CNN uses convolutional kernels of various sizes to extract local features, followed by a pooling layer for downsampling and dimensionality reduction. The reduced features are input into an attention mechanism layer. LSTM and CNN Hybrid Model: Combines the outputs of LSTM and CNN to achieve feature extraction and classification. All four models (including the proposed model) were trained on both the NetEase and THUCNews datasets.

4.3. Experimental Environment

Processor: Intel Core i7-12700H
Memory: 16 GB
GPU: NVIDIA GeForce RTX 3070
Software Environment: Windows platform with Python 3.7.0 for program development

4.4. Evaluation Metrics

The following evaluation metrics were used in this study:

DOI: 10.54254/2755-2721/108/2025.20922

Accuracy: The proportion of correctly predicted samples out of the total samples.

$$Accuracy = \frac{TP + FN}{TP + TN + FP + FN}$$

Precision: The proportion of true positives among all predicted positives.

$$Precision = \frac{TP}{TP + FP}$$

Recall: The proportion of true positives among all actual positives.

$$Recall = \frac{TP}{TP + FN}$$

F1-Score: The harmonic mean of precision and recall.

$$F1 - score = \frac{2 * Precision * Recall}{Precision + Recall}$$

4.5. Comparative Model Results Analysis

The final results of the four models in terms of accuracy, precision, recall, and F1-score are shown in Table 1:

Table 1: Model Evaluation Metrics

	Accuracy	Precision	Recall	F1-score
Ours (THUCNews)	0.989	0.988	0.98	0.984
Ours (NetEase)	0.988	0.987	0.982	0.984
E-TF-IDF-BERT with LSTM and CNN (THUCNews)	0.97	0.972	0.974	0.973
E-TF-IDF-BERT with LSTM and CNN (NetEase)	0.971	0.973	0.97	0.971
LSTM-CNN (THUCNews)	0.91	0.88	0.965	0.872
LSTM-CNN (NetEase)	0.903	0.871	0.875	0.872
LSTM and CNN Hybrid (THUCNews)	0.968	0.968	0.967	0.969
LSTM and CNN Hybrid (NetEase)	0.968	0.968	0.967	0.969

From the table, it can be observed that the proposed model achieves the highest accuracy of 0.989 and 0.988 on the THUCNews and NetEase datasets, respectively, outperforming the other models. The precision of 0.988 and 0.987 indicates that the proposed model performs better in avoiding misclassification of negative samples. Overall, the proposed model achieves superior performance across all four evaluation metrics.

To further highlight the advantages of the proposed model, the training process of different models on the validation sets was analyzed. The changes in val_accuracy with epochs for each model on both datasets are shown in Figure 5:

DOI: 10.54254/2755-2721/108/2025.20922

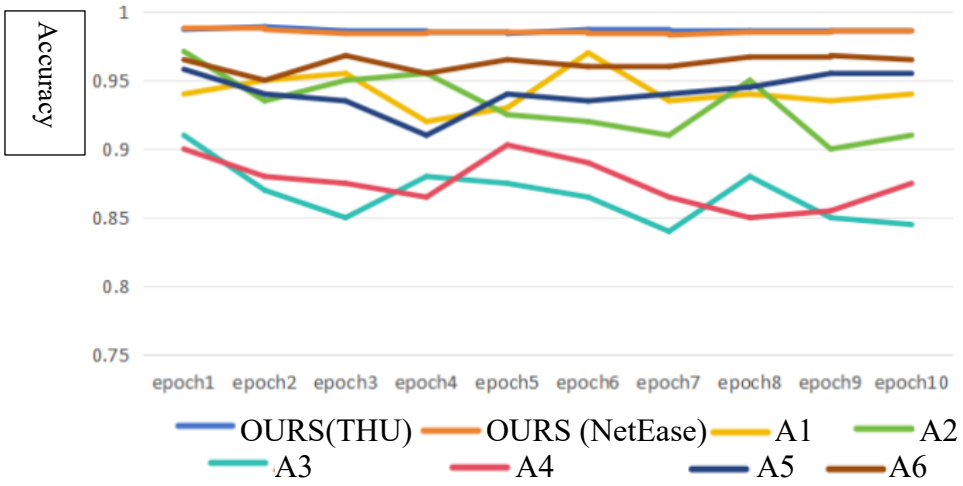

Figure 5: Changes in val_accuracy with Epochs

A1: E-TF-IDF-BERT with LSTM and CNN (THUCNews)
A2: E-TF-IDF-BERT with LSTM and CNN (NetEase)
A3: LSTM-CNN (THUCNews)
A4: LSTM-CNN (NetEase)
A5: LSTM and CNN Hybrid (THUCNews)
A6: LSTM and CNN Hybrid (NetEase)

From Figure 5, it can be seen that the curve of the proposed model remains relatively smooth as the number of epochs increases, indicating fewer fluctuations. In contrast, the curves of the other models exhibit more significant variations, suggesting that the proposed model not only converges faster but also consistently maintains a performance advantage.

5. Conclusion

This paper proposes a news text classification method based on hybrid neural networks, aiming to optimize feature extraction and model structure. By combining the feature extraction capabilities of deep learning with the stability of traditional machine learning, the proposed model enhances the accuracy and efficiency of news text classification. The model achieves favorable results in evaluation metrics such as F1-Score, Recall, and Precision. Comparative experiments with other models further demonstrate the advantages of the proposed approach.

References

[1] *Meng C ,Todo Y ,Tang C , et al.MFLSCI: Multi-granularity fusion and label semantic correlation information for multi-label legal text classification[J].Engineering Applications of Artificial Intelligence,2025,139(PB):109604-109604.*

[2] *Zhou C,Sun C,Liu Z,et al. A C-LSTM Neural Network for Text Classification[J]. Computerence,2015,1(4):39-44.*

[3] *Sun X ,Huang J ,Fang Y , et al.MREDTA: A BERT and transformer-based molecular representation encoder for predicting drug-target binding affinity.[J].FASEB journal : official publication of the Federation of American Societies for Experimental Biology,2024,38(19):e70083.*

[4] *KIM Y. Convolutional Neural Networks for Sentence Classification [J/OL].arXiv:1408.5882[cs.CL].(2014-08-25).https://arxiv.org/abs/1408.5882v2.*

[5] *Zheng S ,Guo C ,Tu D , et al.Spectral super-resolution for high-accuracy rice variety classification using hybrid CNN-Transformer model[J].Journal of Food Composition and Analysis,2025,137(PA):106891-106891.*

[6] *DEVLIN J, CHANG M W, LEE K,et al. BERT: Pretraining of deep bidirectional transformers for language understanding[EB/OL].[2024-05-06]. https://eva.fing.edu. uy/pluginfile.php/524749/mod_folder/content/0/BER*

DOI: 10.54254/2755-2721/108/2025.20922

T%20Pre-training%20of%20Deep%20Bidirectional%20Transformers%20for%20Language%20Understanding.pdf.

[7] *Gao, K. Y., Mou, L., & Zhang, Y. B. (2022). PCC-BiLSTM-GRU-Attention combined model prediction method. Computer Systems and Applications, 31(7), 365–371. http://www.c-sa.org.cn/1003-3254/8580.html*

[8] *Shen Qing,Wen han Yi,and Comite Ubaldo."E-Commerce Live Streaming Danmaku Classification Through LDA-Enhanced BERT-TextCNN Model."International Journal of Information Technologies and Systems Approach (IJITSA) 17.1(2024):1-23.*

[9] *Xuefeng S ,Min H ,Fuji R , et al.A dual-ways feature fusion mechanism enhancing active learning based on TextCNN[J].Intelligent Data Analysis,2024,28(5):1189-1211.*