Research on Fund News Classification Method Based on Multi-level Model Fusion

Ju-Xiang Hu[1], Xue-Qiang Lv[1,2], Xin-Dong You[2] and Jian-She Zhou[1*]

[1] Research Center for Language Intelligence of China,Capital Normal University, Beijing 100048,China

juxianghu@163.com,lxq@bistu.edu.cn,1441425098@qq.com

[2] Beijing Key Laboratory of Internet Culture and Digital Dissemination Research,Beijing Information Science and Technology University,Beijing 100101,China

youxindong@bistu.edu.cn

Abstract. Aiming at many challenges in the field of fund news text classification, including the difficulty of capturing context and dealing with unbalanced feature weights by a single deep learning model, this paper proposes a multi-level model fusion based fund news text classification method. Through in-depth analysis of the structure and characteristics of fund news texts, this method constructs an input matrix based on the integration of multi-level models, and then adopts softmax to complete the classification task of fund news texts. The whole model is referred to as PosiTransBiAttention model (that is, the combination of word location embedding, Transformer, BiLSTM and multi-head attention). The unique feature of PosiTransBiAttention model is that it can fully integrate multi-level features, so as to effectively solve the problem of fund news text classification. The experimental results verify the effectiveness of the model in the fund news text classification task, and the accuracy rate reaches 93.85%. The experimental results further demonstrate the significant performance advantages of the model in this field. Through the combination of multi-feature fusion and multi-level model, this method provides a powerful reference and inspiration for solving the text classification problem in similar fields.

Keywords: Word location characteristics, Attention mechanism,Multi-feature,Text classification

1 Introduction

With the continuous introduction of regulatory policies and the vigorous development of the financial field, a large number of fund news has emerged, and how to accurately classify a large number of fund news is particularly critical. The text data of fund news usually presents the characteristics of many professional terms, unstructured, non-standard terminology, and diversity of language expression, so there are often many challenges in using a single deep learning model for text classification. It is difficult for a single deep learning model to fully consider the context of fund news text and balance the weight of features. In order to solve these problems effectively, the high integration of attention mechanism and multi-features can improve the accuracy of text classification.

With the rapid development of deep learning model and its wide application in the field of natural language processing, more and more researches have transferred the relevant technology to different fields and different tasks, and have made remarkable progress. Liu K et al. [1], aiming at the problem of unknown words in the field of financial public opinion analysis, proposed an unknown word processing strategy, which inferred the semantics of unknown words based on the context information of known words in the text. Compared with traditional text classification methods, this method has outstanding advantages in dealing with unknown words or unrecognizable characters. Zhao W al. [2]proposed AD-CharCGNN algorithm based on CharCNN and GRU to solve the problem of incomplete and insufficient information in text classification in the financial field. The algorithm extracted features from some financial texts and combined them with spatial and temporal domains to

EAI MLICOM 2023, December 17, Beijing, People's Republic of China

DOI 10.4108/eai.17-12-2023.2342373

classify financial texts. This algorithm provides a new perspective and solution for the text classification problem in the financial field. Wan C X et al.[3] proposed a financial text classification model based on BERT-CNN for the problem of causal sentence recognition in the financial field. The model is mainly divided into two key steps. First, the BERT model is used to add CNN structure to the specific task layer, so that the model can encode financial text and obtain rich semantic features. Secondly, the self-attention mechanism is used to classify the text by combining the local text representation and BERT output as input. Tan M J et al. [4]introduced the multi-feature fusion method to detect financial news topics. In this method, the text features of financial news are extracted by subject events, feature word extraction, news semantics, financial named entity recognition, etc., and the nest-cohesive hierarchical clustering algorithm is carried out after feature fusion. This method effectively improves the accuracy of financial news topic detection.

The above research methods mainly focus on feature fusion, but do not conduct in-depth analysis on the text data of fund news, nor do they systematically study the influence of word position and multi-head attention mechanism on the text classification task in multi-feature fusion. Therefore, this paper proposes a fund news classification method based on multi-level model fusion, which aims to deeply study the interaction between multiple features and improve the accuracy of fund news text classification.

2 Related Work

Text classification task in financial field is a focus in deep learning and natural language processing. In recent years, with the rapid development of technology, a large number of domain knowledge involving text feature fusion, attention mechanism, multi-model fusion and other fields have emerged, and have been widely used in various industries. In this dynamic context, research has made significant progress in the following areas:

In the study of text classification based on BERT and its variant models, the emergence of BERT and its variant models provides a new perspective for the extraction and fusion of text features. These models can capture the semantic information of text data efficiently and enrich the dimension of feature representation. Devlin J et al. [5] introduced the theoretical knowledge and related applications of BERT model. The BERT model is pre-trained by the bidirectional transformer structure to obtain excellent results in multiple natural language processing. Liu Y et al. [6] optimized and modified the BERT model and improved the robustness and performance of the model by using larger training data and more training steps on the basis of the BERT model. Zhang Z et al. [7] proposed a Chinese-oriented knowledge enhancement training model, which utilizes entity and relational information to improve the model's semantic understanding ability through multi-task learning. ERNIE performs well in Chinese text processing tasks, especially in named entity recognition and relation extraction. Lan Z et al. [8] proposed the ALBERT model, which reduces the number of parameters of BERT model through parameter sharing, thereby improving training efficiency and model performance. ALBERT is leading the way on both multi-task and single-task natural language processing benchmarks. Sanh V et al. [9]proposed a lightweight BERT model, which compressed the complex BERT model into a small model by knowledge distillation method. Despite the reduced model size, DistilBERT maintained comparable performance across multiple text classification tasks.

The research of text classification based on the fusion of text features explores the combination of different features, such as word embedding, syntactic embedding, emotion embedding and emotion symbol embedding. These strategies attempt to capture text information from multiple perspectives and combine different features organically to obtain a more comprehensive and representative text representation. Zhou P et al. [10]combined the bidirectional LSTM model and two-dimensional maximum pooling for text classification to improve the accuracy and efficiency of classification. Zhang X et al. [11]proposed a text classification method based on character-level convolutional neural networks, which provides a new perspective for solving text classification problems. Zhang W et al. [12] fully mined the semantics of the report by constructing a financial sentiment dictionary, used the K-proximity algorithm to classify the attitudes involved in the report, and proposed a dynamic optimization fusion factor strategy.

In the study of text classification based on attention mechanism fusion, attention mechanism plays a key role in multi-feature fusion. Mechanisms such as self-attention and multi-headed attention are introduced to strengthen the correlation between different features. In this way, researchers are able to capture important features more accurately, which in turn improves classification performance. Vaswani A et al.[13] have introduced Transformer model in detail and proposed a self-attention mechanism, which can improve the expression ability of text features by considering the correlation between words in the text. Yang Z et al.[14] 's study introduced hierarchical attention networks for document classification tasks and used multi-level attention mechanisms to capture correlations at different levels in text. Jia Z Z et al. [15] proposed a multi-feature fusion hierarchical classification method based on attention mechanism to capture the personality and commonality among features, aiming at ignoring the relationship and mutual influence between various dialogue features. Xu

X Y[16] took CNN-BiLSTM, a commonly used emotion classification model, as the basic model, and proposed the SF-CNN-BiLSTM-ATT model, which integrated emotional features and multi-head attention mechanism. The model uses multi-vector feature splicing to enrich semantic features. Zeng S M [17] proposed an attention-based multi-feature fusion emotion analysis method by extracting time series and local features through bidirectional long short-term memory model and convolutional neural network.

In the research of text classification based on multi-model fusion, multi-model fusion is one of the frontier topics in the field of deep learning. The researchers combined different deep learning models, such as CNN, BiLSTM, and LSTM, with each other and weighted them using attention mechanisms. This multi-model synergy can effectively combine the advantages of different models to obtain more accurate classification results. Du C et al. [18]realized Chinese text classification by combining multiple neural network models to obtain more accurate classification results. Yu F et al.[19] proposed Dilated Residual Networks and discussed the method of introducing void convolution into deep learning models. Xu G et al. [20]proposed a social emotion classification method, which adopted a multi-model fusion strategy. The researchers combined different emotion classification models, including traditional machine learning methods and deep learning methods such as convolutional neural networks (CNN) and recurrent neural networks (RNN). Each model captures different features of text data and improves the comprehensive ability of sentiment classification through fusion weighting. Lin M et al. [21] proposed a method based on multi-feature and multi-model fusion by combining multiple feature extraction techniques and multiple models.

Although the above relevant studies have conducted extensive exploration in the field of text classification such as BERT model, text feature fusion, attention mechanism, and multi-model fusion, the research has not conducted a comprehensive and in-depth discussion on the introduction of word location feature fusion, context relationship, and attention mechanism. Therefore, this paper proposes a text classification method of fund news based on the combination of word position, context and attention mechanism. This method first introduces the word location information into different deep learning models, then adopts BiLSTM model to extract the context of fund news text, and finally integrates multi-head attention mechanism to improve the accuracy of fund news text classification. When combining the results of different models, multi-head attention mechanism can better mine the correlation between features, so as to capture the semantic information of text more accurately. At the same time, the multi-head attention mechanism also helps to solve the problem of unbalanced feature weights, so that the key features can affect the classification results with higher weights. In summary, the main contributions of this study are as follows:

(1) This paper proposes a fund news classification method based on multi-level model fusion. In the fund news text classification task, the method integrates the word location features with a variety of different deep learning models, then adopts BiLSTM to extract the context of the fund news text, and introduces multi-head attention mechanism to balance the weight between different features, thus significantly improving the accuracy of the text classification.

(2) By conducting comprehensive comparative experiments on various deep learning models, this paper verifies the effectiveness and superiority of PosiTransBiAttention model. With the help of multi-angle comparative experiments, the performance of these models in text classification tasks is demonstrated, which provides a strong support for practical application.

3 Model Construction

The framework of PosiTransBiAttention model is shown in Figure 1. The structure of the model is mainly divided into input layer, word position embedding layer, coding layer, feature layer, attention layer and output layer. In this paper, these abstract representations of different layers are cleverly fused and spliced to make full use of the information extracted from each level. The fusion strategy of this research aims to capture and integrate various features in the input text, so as to achieve a more comprehensive and accurate task of fund news text classification. Subsequently, by introducing an attention mechanism, the model is able to focus more on key information, further improving the classification ability.

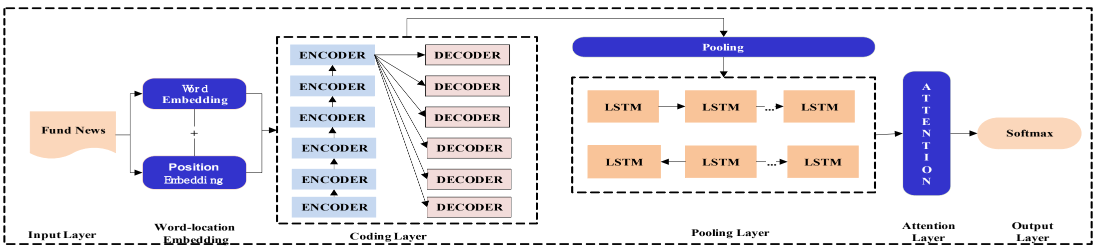


Fig.1. PosiTransBiAttention frame diagram

3.1 Input Layer

The input layer plays a crucial role in the PosiTransBiAttention model framework, which is not only the starting point of the model, but also undertakes several key tasks. These tasks include not only transforming the original fund news text data into a form suitable for deep learning model processing, but also the necessary pre-processing steps that provide a solid foundation for subsequent model processing.

The functionality of the input layer covers several key aspects. First, it is responsible for pre-processing the original fund news text. This includes tasks such as word segmentation, removing stops, and processing punctuation marks to ensure that the data is ready for subsequent processing. Second, the input layer involves the determination of the initial parameters of the deep model, a process that is crucial for the training and learning of the model, ensuring that the model can start from a good initial state when processing text. Finally, the category label of fund news text is quantified by One-Hot coding, and the classification of fund news text is identified in a unique way.

Label vectorization is the process of converting category labels in a classification task into vector representation, and one of the common methods is to use one-hot encoding for vectorization representation. In this process, each class is given a unique integer identifier, which is then represented as a vector of 0 and 1, where only the elements in the position corresponding to the class are 1, and the rest are 0. This representation allows for clear boundaries in the vector space between each class, providing a clear goal for the model.

Suppose there are N categories, of which the I-th category is vectored to the one-HOT-encoded vector represented as y_i. Each vector y_i has N elements representing a point in an n-dimensional vector space.

y_i[j] = 1, if j == i

y_i[j] = 0 if j ≠ i

Both i and j are the indexes of the class, with values ranging from 1 to n. In the vector y_i, only the elements in the i-th position are 1, and the elements in the rest positions are 0, thus achieving the vectorized representation of one-hot encoding.

3.2 Word position Embedding Layer

Word location embedding is a method used to add location information to sequence data. In natural language processing, the meaning of words can have different effects depending on where they appear. For example, the position of a word in a sentence can determine its role in the sentence structure (subject, predicate, etc.). The concept of word location embedding is to assign a unique encoding to each location, giving each location a unique vector representation to enable the model to capture information about different locations. By adding the positional embeddings to the embeddings of the input tags, the model can consider both the semantics of the tags and the positional information to better understand the structure of the sequence. The calculation formula of word position coding is shown in formulas (1) and (2) :

$$PE_{(pos,2_i)} = \sin(\frac{pos}{10000^{2i/d_{model}}}) \tag{1}$$

$$PE_{(pos,2_{i+1})} = \cos(\frac{pos}{10000^{2i/d_{model}}}) \tag{2}$$

Where pos is the position of the word in the sentence, the value range is [0,max sequence length],i is the dimension of the word embedding, the value range is [0,max embeding dimension], and dmodel is the embeding dimension value.

3.3 Coding Layer

The encoder layer further enhances the ability of sequence modeling. Traditional sequential models, such as recurrent neural networks (RNNS) and long and short term memory networks (LSTMS), may encounter problems such as gradient disappearance or gradient explosion when dealing with long distance dependencies. Transformer uses a self-attention mechanism that allows each location to relate to other locations in the sequence, regardless of distance. This allows Transformer to capture long distance dependencies while maintaining good training stability. At the same time, word position embedding helps encode the position information of the input sequence, and the Transformer encoder layer can capture the relationships and features within the sequence to extract valuable representations.

3.4 Feature Layer Construction

3.4.1 BiLSTM feature extraction

The fund news text is presented as a sequence of words, where each word is associated with the surrounding words. BiLSTM in this paper can effectively capture these context relations, accurately grasp the semantic and structural characteristics of the text, and extract useful features of the fund news text. In addition, fundnews texts often have complex sentence structures and long-distance dependencies, while BiLSTM, through forward and backward information transfer, is able to capture patterns in sentences more comprehensively, leading to a better understanding of the meaning of sentences. BiLSTM feature extraction algorithm is shown in Figure 2:

Algorithm 1 BiLSTM Feature Extraction

Require: Input:

- X: Input sequence of positions and transforms,
- where $X = [x_1, x_2, \ldots, x_T]$.
- H: Hidden state dimension of the BiLSTM.

Output:

- F: Extracted features, where $F = [f_1, f_2, \ldots, f_T]$.

Initialize BiLSTM parameters Θ.
Initialize hidden states: $h_0^{\text{forward}} = \mathbf{0}$, $h_0^{\text{backward}} = \mathbf{0}$.
for $t = 1$ to $T - 1$ **do**
  Concatenate input: $x_t = [x_t^{\text{position}}, x_t^{\text{transform}}]$.
  Forward pass through BiLSTM:

$$h_t^{\text{forward}} = \text{LSTM}_{\text{forward}}(x_t, h_{t-1}^{\text{forward}}; \Theta)$$
$$h_t^{\text{backward}} = \text{LSTM}_{\text{backward}}(x_t, h_{t+1}^{\text{backward}}; \Theta)$$

  Concatenate forward and backward hidden states:
  $h_t = [h_t^{\text{forward}}, h_t^{\text{backward}}]$.
  Store the extracted feature: $f_t = h_t$.
end for
Return $F = [f_1, f_2, \ldots, f_T]$.

Fig. 2. BiLSTM feature extraction algorithm

3.4.2 TextCNN feature extraction

In order to compare with the feature effect extracted by BiLSTM, this paper adopts TextCNN to extract the feature of fund news text. TextCNN is a convolutional neural network that can effectively capture local features in the text data of fund news. By introducing TextCNN model, local features with different scales can be extracted from fund news text data, which represent the structure, pattern and correlation of fund news text. By merging the TextCNN layer with the word position embedding layer and the Transformer encoder layer, the model gradually builds a more complex abstract representation. These abstract representations can simultaneously capture sequence-level relationships, lexical-level meanings, and local feature patterns to provide richer information for downstream tasks. TextCNN feature extraction algorithm is shown in Figure 3:

Algorithm 1 TextCNN Feature Extraction

Require: Input:

- X: Input text sequence, where $X = [x_1, x_2, \ldots, x_T]$.
- C: Number of filters.
- K: Filter sizes.
- F: Number of features.

Output:

- $TextCNNFeatures$: Extracted features, where $TextCNNFeatures = [f_1, f_2, \ldots, f_F]$.

Initialize TextCNN parameters Θ.
Initialize convolutional filters: $W_1, W_2, \ldots, W_C$.
Initialize bias terms: $b_1, b_2, \ldots, b_C$.
for $c = 1$ to C **do**
  Convolution and max-pooling:
  for $k = 1$ to K **do**
    Convolve the filter W_c with the input sequence X:
    $c_k = \text{Convolution}(x, W_c)$
    Apply max-pooling to c_k:
    $p_k = \text{MaxPooling}(c_k)$
  end for
  Concatenate the max-pooled features: $p = [p_1, p_2, \ldots, p_K]$
  Apply a non-linear activation function: $h_c = \text{ReLU}(p + b_c)$
end for
Concatenate the output features from all filters: $TextCNN - Features = [h_1, h_2, \ldots, h_C]$
Return $TextCNN - Features$

Fig. 3. TextCNN feature extraction algorithm

3.5 Attention Layer

The attention layer mainly trains the input with multiple attention mechanisms and then integrates them into the model. The multi-head attention mechanism adopts a parallel approach when dealing with different feature relationships in the input. This level of parallelism is driven by different heads of attention [15]. Multiple layers of attention can be seen as a combination of multiple independent attention mechanisms, each of which is called a "head." Each header maps the input query (Q), key (K), and value (V) to a different representation subspace via independent linear transformations. By combining the independent computing power of different heads, the multi-head attention layer provides the model with greater fusion capabilities, allowing it to better understand and represent the various dimensions of the input sequence. In this design, the mechanism of multiple attention layers makes full use of the complementary role of multiple attention heads, so that the model can more effectively capture the internal correlation of the input sequence. This approach helps to improve the model's ability to understand the relationship between different information, thus improving the model's ability to perform complex tasks. The algorithm of the multi-head attention mechanism is shown in Figure 4:

Algorithm 1: Multi-Head Attention

Data: Input sequence $X \in \mathbb{R}^{n \times d}$, Number of attention heads H, Query dimension d_q, Key dimension d_k, Value dimension d_v

for $h \in [1, H]$ **do**

- Query, Key, and Value projections;
- $Q_h = X \cdot W_{Q_h}$;
- $K_h = X \cdot W_{K_h}$;
- $V_h = X \cdot W_{V_h}$;
- Attention score calculation;
- $AttentionScores_h = \frac{Q_h \cdot K_h^T}{\sqrt{d_k}}$;
- Attention weight calculation;
- $A_h = \mathrm{Softmax}(AttentionScores_h)$;
- Multi-head attention output calculation;
- $MultiHeadOutput_h = A_h \cdot V_h$;

Sum the outputs of all heads to get the final output

$FinalOutput = \sum_{h=1}^{H} MultiHeadOutput_h$;

Fig. 4. Multi-head attention mechanism algorithm

3.6 Output Layer

This paper uses softmax function to output the classified text in the fund news text. softmax is a commonly used activation function, often used for multi-class classification problems, and its main role is to convert the raw output of the model into a probability distribution that represents the predicted probability for each class. For the Fund News 5 classification problem, the softmax function converts the 5-dimensional output vector of the model (usually the raw, unnormalized score) into a 5-dimensional probability distribution vector, where each element represents the probability of the corresponding category. By exponenting and normalizing scores, the softmax function emphasizes the relative relationship of each category score to the overall score, making the prediction results more intuitive. The calculation formula is shown in formula (3) :

$$P_i = \frac{e^{zi}}{\sum_{j=1}^{K} e^{zj}} \quad (1 \leq k \geq 5) \tag{3}$$

Where, the input vector z = (z1, z2... , zn), n is the total number of categories.

4 Experiment And Analysis

4.1 Data Preprocessing

4.1.1 Preprocessing

In the pre-processing stage, firstly, this study adopts regular expression method to remove useless characters in fund news text, such as special symbols (@, *), etc. Second, by using a public stop list, the fund news text is removed from the stop word, aiming to exclude those common words that are not important in the text analysis. The purpose of this step is to extract more meaningful information from the text for subsequent processing and analysis.

4.1.2 Text Analysis

In order to facilitate the calculation of word position embedding vector, this paper adopts jieba for word segmentation of fund news text. At the same time, the length of the word sequence after each sample word segmentation is counted, and the frequency distribution of the text length is displayed by the bar chart. This analysis helps determine the maximum dimensions of the model for more accurate modeling in subsequent processing. The sequence length analysis of words in sentences is shown in Figure 5:

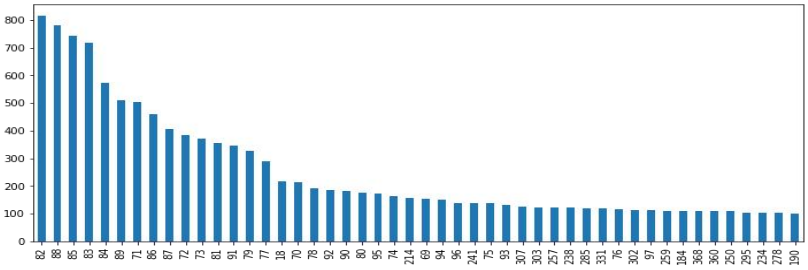


Fig. 5. Length distribution of word sequences

4.1.3 PosiTransBiMulti Model Structure

In this paper, optimizer, maximum number of iterations, different batch cycles, different learning rate cycles and early stop method are used to select the best parameters of the model. The initial iteration number is set to 100 times, and the early stop method is used to terminate the experiment in advance to achieve the convergence of the model. The structural parameters of the PosiTransBiAttention model are shown in Table 1, which records the hierarchical structure of the PosiTransBiAttention model in detail. The total parameters and trainable parameters are 600,133.

Table 1. Structural parameters of PosiTransBiAttention model

Layer (type)	Output Shape	Parameter
inputs (InputLayer)	(None, 100)	0
positional_embedding	(None, 100, 32)	195200
transformer_encoder	(None, 100, 32)	19040
bidirectional	(None, 100, 128)	49664
bidirectional	(None, 100, 128)	98816
embedding	(None, 100, 32)	192000
concatenate	(None, 100, 160)	0
multi_head_attention	(None, 100, 32)	33568
global_max_pooling1d	(None, 160)	0
global_max_pooling1d	(None, 32)	0
dropout	(None, 160)	0
dropout	(None, 32)	0
dense	(None, 64)	10304
dense	(None, 32)	1056
concatenate	(None, 96)	0
dropout	(None, 96)	0
dense	(None, 5)	485

4.2 Experimental Data

This paper collects and selects 50,000 pieces of fund news text data as experimental data, which are divided into training set, test set and verification set according to 8:1:1. Fund news data is divided into 5 categories, namely regulations and policies, macro news, industry news, fund market conditions and comment outlook, category numbers are 0-4.

4.3 Experimental evaluation indexes

In this paper, three evaluation indicators, including Precision, Recall and F value, are selected to evaluate the model effect. The

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{4}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{5}$$

$$\text{F} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{6}$$

calculation formula for evaluation indicators is as follows:

TP indicates the number of samples of the correctly identified category, FP indicates the number of samples of the correctly identified category but not actually of the category, TN indicates the number of samples of the correctly identified category without marking errors, and FN indicates the number of samples of the incorrectly identified category.

4.4 Experimental Results And Analysis

4.4.1 The influence of word position and context on the classification of fund news text

To verify the impact of word position and context embeddings on fund news text classification, this paper uses Transformer text classification model as a benchmark experiment, and designs the following comparison model:

1. Transformer model: As the benchmark model of this paper, Transformer captures different types of relationships through multiple attention heads, so as to better represent multi-level information in the text.
2. BiLSTM model: In order to effectively capture the context and semantic information in the text, BiLSTM is used to extract features from the fund news text.
3. TextCNN model: In order to compare the BiLSTM model, a convolutional neural network architecture (CNN) architecture is adopted to extract local features in the text through the convolutional layer and pooling layer.
4. MLP model: In order to compare BiLSTM model, MLP model is used to learn the nonlinear relationship of fund news text.

5, PosiTrans model (PositionalEmbedding Transformer) : PosiTrans model is a combination of location embedding and Transformer model, designed to capture the location information in the text. By introducing positional embedding, PosiTrans can better understand the order and positional relationships of words in text.

6. PosiTransBi model (PositionalEmbedding Transformer BiLSTM) : PosiTransBi model combines location embedding with BiLSTM to capture the semantic and positional information of text more comprehensively.

7. PosiTransTex model (PositionalEmbedding Transformer TextCNN) : PosiTransTex model is a combination of location embedding and text convolutional neural network (TextCNN) model.

8, PosiTransMLP model (PositionalEmbedding Transformer MLP) : PosiTransMLP model is a combination of location embedding and multi-layer perceptron (MLP) model.

Through a series of designed comparative experiments, this paper experimentally compares and analyzes the performance of various deep learning models on fund news text processing, revealing the advantages and disadvantages of each model. The experiment improves the existing model by introducing position embedding, which improves the model's understanding of word order and contextual relationships, thus enhancing the accuracy of text representation. The improved model combines position embedding with different deep learning architectures, which not only retains the advantages of the original model, but also enhances the sensitivity to position information, thus improving accuracy, contextual understanding, efficiency, and generality. This provides valuable reference and inspiration for processing other types of complex natural language text. The experimental results are detailed in Table 2.

Table 2. The results of the best parameter embedding of word position and context

Index	Experiment	Precision	Recall	F Score	F Improvement
1	Transformer(baseline)	0.9010	0.8904	0.8881	0.00%
2	BiLSTM	0.9102	0.9038	0.9023	1.42%
3	TextCNN	0.9131	0.9074	0.9062	1.81%
4	MLP	0.9116	0.9016	0.8991	1.10%
5	PosiTrans	0.9128	0.9012	0.8992	1.11%
6	**PosiTransBi**	**0.9283**	**0.9240**	**0.9234**	**3.53%**
7	PosiTransTex	0.9281	0.9234	0.9229	3.47%
8	PosiTransMLP	0.9107	0.9040	0.9025	1.44%

As can be seen from Table 2, the F value of fund news text for text classification by Transformer feature extraction is 0.8881. In terms of classification effect, compared with single models such as BiLSTM, TextCNN and MLP, Transformer model is not significantly superior. Compared with the single model, the fund classification effect after embedding the position characteristics of words has shown obvious improvement, in which the F-value of PosiTrans model has increased by 1.11% over the baseline, and the F-value of PosiTransBi model has increased by 3.53% over the baseline. The F value of PosiTransTex model increased by 3.47% over the baseline, and the F value of PosiTransMLP model increased by 1.44% over the baseline. The experimental results show that by introducing the word location information, all models show a certain degree of ability improvement in the fund news text classification task. This is because word location information can help the model better understand the structure and context of the text, so as to improve the ability to capture the semantic meaning of the text. The combination of different model structure and word location information strengthens the characteristics of the model to different degrees, and thus achieves better classification effect in this task.

4.4.2 The influence of multi-head attention mechanism on the classification effect of fund news text

In order to verify the effect of attention mechanism on fund news text classification, this paper integrates multi-head attention mechanism with multi-feature fusion model introducing word position, and designs the following comparative experiments:

1. MultiHeadAttention model: In order to analyze the attention of different parts of the input sequence, multi-head attention mechanism is introduced to capture rich semantic information.

2. PosiTransAttention model (PositionalEmbedding Transformer MultiHeadAttention) : PosiTransAttention model is a Transformer model that combines location embedding and multi-head attention mechanism.

3, PosiTransBiAttention model (PositionalEmbedding Transformer BiLSTM MultiHeadAttention) : The PosiTransBiAttention model combines location embeddedness, BiLSTM and multipronged attention to consider the location, semantics and multipronged attention information of text.

4. PosiTransTexAttention model (PositionalEmbedding Transformer TextCNN) MultiHeadAttention: The PosiTransTexAttention model is a model that combines location embedding, Transformer, text convolutional neural networks (TextCNN), and multi-head attention.

5. PosiTransMLPAttention model (PositionalEmbedding Transformer MLP MultiHeadAttention) : PosiTransMLPAttention model is a model combining location embedding, Transformer, multi-layer perceptron (MLP), and multi-head attention.

6, PosiTransBiTex model (PositionalEmbedding Transformer BiLSTM) TextCNN :PosiTransBiTex model is a comprehensive model that combines location embedding, Transformer, BiLSTM and text convolutional neural network (TextCNN).

7,PosiTransBiTexAttention model (PositionalEmbedding Transformer BiLSTM TextCNN) MultiHeadAttention :PosiTransBiTexAttention model is a model that combines location embedding, Transformer, bidirectional short and long duration memory network (BiLSTM), text convolutional neural network (TextCNN), and multi-head attention.

Through a series of designed comparative experiments, this paper verifies the effectiveness of the combination of word position, context and multi-head attention mechanisms in different deep learning models. This study provides more in-depth analysis and empirical results for the classification of fund news texts, highlighting the key role of these key factors in improving classification performance. The experimental results are shown in Table 3:

Table 3. Experimental results of optimal parameters of multi-head attention mechanism

Index	Experiment	Precision	Recall	F Score	F Improvement
9	MultiHeadAttention	0.9030	0.8936	0.8922	0.41%
10	PosiTransAttention	0.9367	0.9334	0.9228	3.47%
11	**PosiTransBiAttention**	**0.9385**	**0.9366**	**0.9364**	**4.82%**
12	PosiTransTexAttention	0.9378	0.9350	0.9348	4.66%
13	PosiTransMLPAttention	0.9300	0.9264	0.9260	3.79%
14	PosiTransBiTex	0.9046	0.8998	0.8991	1.09%
15	PosiTransBiTexAttention	0.9331	0.9304	0.9301	4.19%

As can be seen from Table 3, the F-value of text classification using a single MultiHeadAttention model to extract fund text features increased by 0.41% over the baseline, indicating that the multi-head attention mechanism can enhance the model's understanding of key information and semantic relationships in text to a certain extent, thus slightly improving the classification effect. In-depth study of other models found that the F-value of PosiTransBiAttention model increased by 3.47% compared with the baseline, and that of Positransbiattention model increased by 4.82% compared with the baseline. F-value of PosiTransTexAttention model is improved by 4.66% compared with baseline, and F-value of PosiTransMLPAttention model is improved by 3.79% compared with baseline. These experimental results collectively show a clear trend that multi-head attention mechanisms play a significant role in enhancing the model's capture of intrinsic information in text, thus enabling the model to make more accurate classification decisions.

However, F-value of PosiTransBiTex model increased by 1.09% compared with baseline, and F-value of PosiTransBiTexAttention model increased by 4.19% compared with baseline. The experimental results show that, The feature extraction of BiLSTM and TextCNN seems to be mutually exclusive. However, after the introduction of multi-head attention mechanism, the experimental results have been significantly improved through the weight balance.

4.4.3 PosiTransBiAttention Model Training Analysis

The training process of PosiTransBiAttention model is shown in Figure 6. From the indicators of the training process, the loss value is gradually reduced and the accuracy is gradually improved by adopting the early stop method in this paper. The results show that the model has good learning and generalization ability when dealing with the task of fund news text classification.

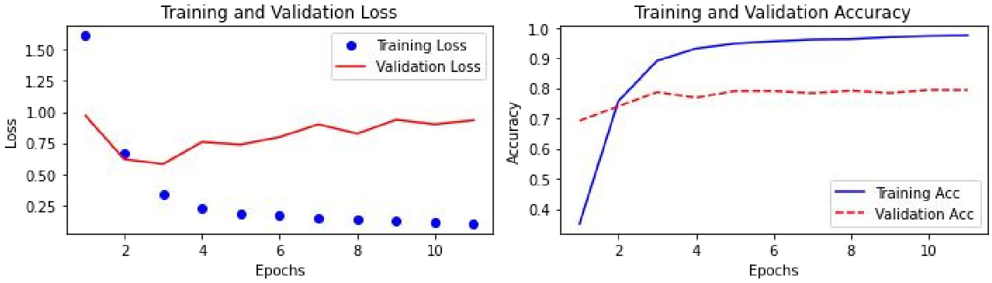


Fig.6. Optimum parameters of PosiTransBiAttention

4.4.4 PosiTransBiAttention Model Confuses Matrix Results

The detailed classification results of PosiTransBiAttention are shown in Figure 7. This method performs well in the task of fund news text classification, with high accuracy, recall rate and F-value. The predictive ability of each category is balanced, and the model can effectively capture the features of different categories. This shows that the model has robust learning and generalization ability in multi-class text classification.

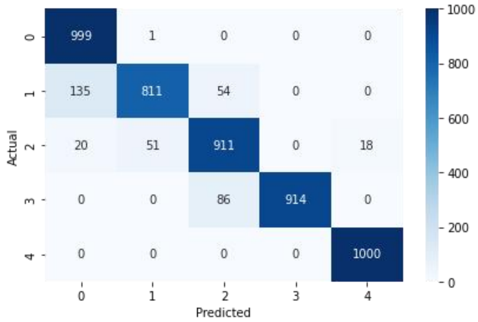
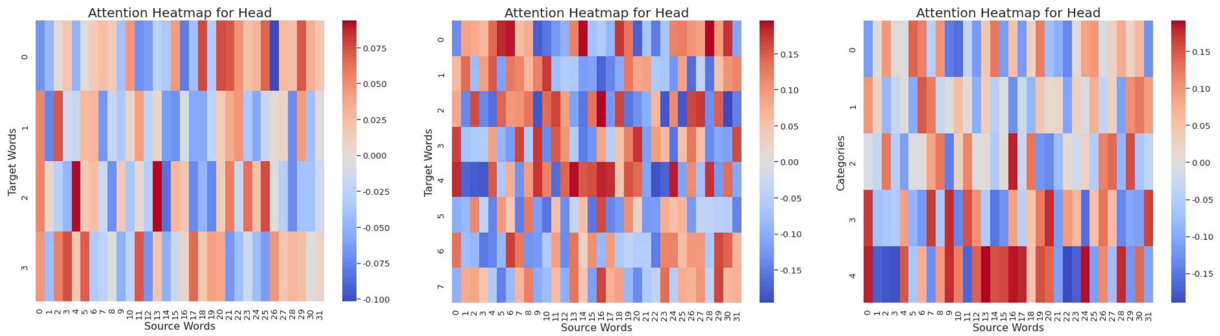


Fig.7. PosiTransBiAttention classification results

4.4.5 Attention weight analysis of PosiTransBiAttention model

The PosiTransBiAttention model consists of two key attention parts, which are the coding layer and the attention layer. The distribution of the PosiTransBiAttention model is shown in Figure 8. Figure 8(a) shows the attention mechanism of the Transformer layer, where the number of attention heads ranges from 0 to 3. The heat map clearly shows the distribution of the model's attention across different words, thus reflecting the focus of attention within the model. Figure 8(b) shows the attention distribution of the multi-head attention mechanism, with the number of attention heads ranging from 0 to 8. The heat map shows the distribution of the model's multi-head attention over different words, with each head focusing on different features or information, and the multi-head attention mechanism helps the model understand the text more fully and capture diverse information. Figure 8(c) shows the weight distribution of bull attention mechanism among fund news text categories and terms. Heat maps indicate competition or sharing of attention weights between different categories, that is, some words may have higher attention weights for more than one category. By analyzing heat maps, we can better understand how attention mechanisms help the model determine the importance of words in different category classifications.



(a)Self-attention weight (b)Multi-head attention weight (c)Category weight

Fig.8. Distribution of attention in PosiTransBiAttention model

5 Closing remarks

This paper aims to solve the problem of text classification of fund news, and significantly improve the performance of text classification by deeply exploring the integration of word position, context and multi-attention mechanism. In order to verify the experimental effect, this paper constructs a series of composite models and compares them with the traditional single model. Different feature fusion methods have different effects on the task of fund news text classification, and there are complex interconnections among features. This research method skillfully fuses and splices different levels of features together, effectively combines lexical position, context and multi-head attention mechanism, thus significantly improving the accuracy of fund news text classification.

In the course of the study, it is found that the quality of word segmentation has an impact on text classification. Therefore, the next step will focus on improving the segmentation effect, and consider the introduction of semantic vector and other methods to further improve the classification performance of the model. By considering many factors comprehensively, it is expected to optimize the text classification model of fund news more comprehensively, and provide a more accurate and effective solution to solve the text classification problem in practical applications.

Acknowledgement

The project was supported by the National Language Commission(No.ZDII 47-17),the Natural Science Foundation of Beijing under Grants (No.4212020), National Natural Science Foundation of China under Grants (No.62171043), National Language Commission project under Grants (No.ZDI145-10),R&D Program of Beijing Municipal Education Commission under grant (No.KM202111232001).

References

[1] Liu K, Ergu D, Cai Y, et al. A new approach to process the unknown words in financial public opinion[J]. Procedia Computer Science, 2019, 162: 523-531.

[2] Zhao W, Zhang G, Yuan G, et al. The study on the text classification for financial news based on partial information[J]. IEEE Access, 2020, 8: 100426-100437.

[3] Wan C X, Li B. Financial causal sentence recognition based on BERT-CNN text classification[J]. The Journal of Supercomputing, 2022: 1-25.

[4] Tan M J,LV X,Tao F F.Study of Finacial News Topic Detection Based on Multi-feature Fusion[J].Computer Engineering,2019,45(03):293-299+308.

[5] Devlin J, Chang M W, Lee K, et al. Bert: Pre-training of deep bidirectional transformers for language understanding[J]. arXiv preprint arXiv:1810.04805, 2018.

[6] Liu Y, Ott M, Goyal N, et al. Roberta: A robustly optimized bert pretraining approach[J]. arXiv preprint arXiv:1907.11692, 2019.

[7] Zhang Z, Han X, Liu Z, et al. ERNIE: Enhanced language representation with informative entities[J]. arXiv preprint arXiv:1905.07129, 2019.

[8] 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.

[9] Sanh V, Debut L, Chaumond J, et al. DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter[J]. arXiv preprint arXiv:1910.01108, 2019.

[10] Zhou P, Qi Z, Zheng S, et al. Text classification improved by integrating bidirectional LSTM with two-dimensional max pooling[J]. arXiv preprint arXiv:1611.06639, 2016.

[11] Zhang X, Zhao J, LeCun Y. Character-level convolutional networks for text classification[J]. Advances in neural information processing systems, 2015, 28.

[12] Zhang W,Zhu H Q,Gao Z G.Feature mining for financial texts and research on dynamic factor fusion strategy[J].Computer Engineering and Applications,2023,59(08):297-305.

[13] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need in Advances in Neural Information Processing Systems, 2017[J]. Search PubMed, 5998-6008.

[14] 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.

[15] Jia Z Z,Gao P F,Ma Y L, et al.Multi-feature fusion attention-based hierarchical classification method for dialogue act[J/OL].Journal of Computer Applications,2023.08.22:1-9.

[16] Xu X Y. Sentiment classification study based on Multi-feature fusion and Attention mechanism[D].AnHui University of Science and Technology,2021.

[17] Zeng S M. Research on sentiment analysis method of Multi-feature fusion based on Attention[D].Harbin Engineering University,2020.

[18] Du C, Huang L. Text classification research with attention-based recurrent neural networks[J]. International Journal of Computers Communications & Control, 2018, 13(1): 50-61.

[19] Yu F, Koltun V, Funkhouser T. Dilated residual networks[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 472-480.

[20] Xu G, Li W, Liu J. A social emotion classification approach using multi-model fusion[J]. Future Generation Computer Systems, 2020, 102: 347-356.

[21] Lin M, Wu D, Meng J, et al. A multi-feature-based multi-model fusion method for state of health estimation of lithium-ion batteries[J]. Journal of Power Sources, 2022, 518: 230774.