Proceedings of Machine Learning Research vol 278:1–10, 2025 2025 2nd International Conference on Machine Learning and Intelligent Computing

Research on Part of Speech Enhanced Text Classification Based on Rotation Position Encoding and Hierarchical Features fusion Text Classification

Mingjian Li* 1102338778@QQ.COM and **Huaiyang Li**
Xichang Minzu Preschool Normal College: Information Center, XCMY, LiangShan, China

Editors: Nianyin Zeng, Ram Bilas Pachori and Dongshu Wang

Abstract

In response to the current problems of missing contextual information, incomplete feature representation, and difficulty in semantic parsing in text classification. This article proposes a text classification framework that combines feature fusion and vocabulary enhancement. Firstly, use WoBERT to encode the text, collect dynamic word vectors, and effectively integrate vocabulary into characters to enhance boundary interaction; Secondly, rotation position encoding is introduced into the character vector to obtain relative distance information between characters and improve feature embedding; Subsequently, to enhance the feature capture capability, the D-mixup structure was introduced to cross fuse the relative distance and CLS information, and continuously extract the global representation of the text in depth; Finally, the Multi Sample Dropout method is used to calculate the loss of multi-level mixed global representations, improving the learning ability of the model. In the experiments on the THUCnews and SMP2020 datasets, the F1 values of the proposed model were 94.72% and 78.38%, respectively, indicating better performance than the current research methods. This indicates that the model proposed in this article can effectively improve generalization and robustness, enhance text classification performance, and is easy to implement, providing reference ideas for future research.

Keywords: text classification; rotation position encode; feature fusion; WoBERT; multi level mix.

1. Introduction

Text Classification is an important task in the field of natural language processing (NLP), which aims to assign predefined categories to texts. It plays a key supporting role in downstream applications such as sentiment analysis, intelligent question answering, spam processing, and topic labeling.

Currently, with the development of neural networks and large model technology, deep learning has gained new momentum in text classification. However, there are still some shortcomings in deep learning-based classification algorithms. For example, although CNN can capture different granularities of text well through convolution kernels of different window sizes, it ignores the order of the text; LSTM obtains the global semantic features of the text through recursive calculation, but lacks the ability to capture local semantic features, and cannot obtain the key information in the text; in addition, previous studies have ignored the impact of vocabulary on performance, which can easily lead to noise caused by the phenomenon of one word having multiple meanings, making word segmentation ambiguous.

In response to the above problems, a method of text classification is proposed that combines rotational position encoding and hierarchical feature fusion, starting from the directions of external knowledge embedding and feature mixing. The research explores the underlying features such as positional relationships, lexical representations, and probabilistic generalization, to further solve the problem of multiple meanings for one word, strengthen the model's ability to understand text, and improve the ability to mine the potential characteristics of characters.

2. Related Work

Deep learning uses neural networks to automatically train data features. Its application in text classification focuses on obtaining more critical and comprehensive feature representations. Commonly used neural networks include convolutional neural networks (CNN), recurrent neural networks (RNN), attention mechanisms, and pre-trained models.

CNN was initially used for image classification, and because it can perform operations such as convolution and pooling on multiple sequences, it has also been applied to natural language processing tasks. Kim (2014)proposed the TextCNN algorithm, which was the first to apply CNN to text classification. By customizing the kernel size of the convolution, local text feature information at different positions can be extracted. Conneau et al. (2017) built on this foundation and studied the problems of deep networks in text classification tasks, and proposed the character-level deep convolutional neural network VDCNN. Johnson and Zhang (2017) further explored the connections between text information and proposed Deep Pyramid Convolutional Neural Networks (DPCNN). Subsequent researchers have made variations and improvements to CNN, and have successively proposed models such as TextGCN, GCN, and GCNII.

The BERT (Devlin et al., 2019) model learns character feature representations by running self-supervised learning methods on a large amount of corpus data. It has achieved first place in multiple tasks in the field of NLP and has given new impetus to neural networks.Many researchers have seen emerging research directions and have introduced improvements to BERT one after the other. Models such as RoBERTa (Liu et al., 2019), ALBERT (Lan et al., 2019) have all optimized modeling performance to some extent (Vaswani et al., 2017).

Papers (YANG et al., 2023; HUANG and LI, 2022; Ning et al., 2024) use methods such as BERT, attention mechanisms, and adversarial training to extract features, which can model textual information well and achieve good results. However, there are still problems such as ambiguous word order relationships, monotonous linear features, and weak generalization ability, which can lead to overfitting.

3. WR-Mix Model

The WR-Mix (WoBERT-RoPE-MD-Mixup) model mainly consists of a word embedding layer, a feature extraction layer, a feature fusion layer, and an output layer, as shown in Figure 1.

3.1. Word Embedding Layer

WoBERT (Su, 2020) is a word-level pre-training model proposed by Su Jianlin. The core lies in the improved Tokenizer algorithm, which enables the tokenizer to embed Chinese words well. Finally, MLM training is performed on the basis of RoBERTa-wwm-ext to further improve the text representation effect. The input structure of the (Ni et al., 2024) is shown in Figure 2:

3.2. Feature Extraction Layer

Rotary Position Embedding (RoPE) (Su et al., 2021) is a relative position encoding model proposed by Su Jianlin. It encodes absolute position information into each dimension by rotation, so that the model can capture the relative position information between characters in the sequence and has good extrapolation. It is widely used in large models such as LLaMA and ChatGLM.

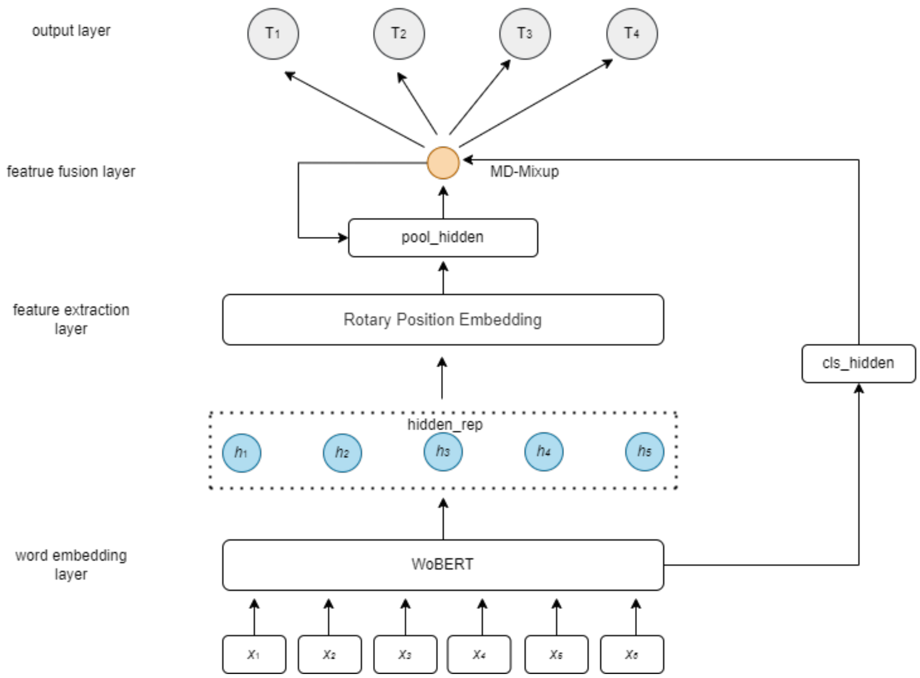

Figure 1: Model diagram.

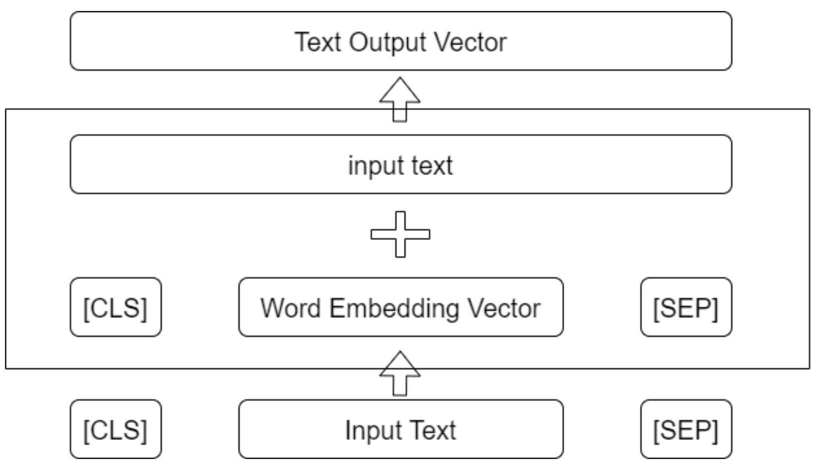

Figure 2: WoBERT model diagram.

$$< f_q(q_m, m), f_k(k_n, n) >= g(q_m, k_n, m-n) \tag{1}$$

$$\begin{aligned}
&f_q(q_m, m) = q_m e^{im\theta} = (W_q x_m) e^{im\theta} \\
&f_k(k_n, n) = k_n e^{in\theta} = (W_k x_n) e^{in\theta} \\
&g(q_m, k_n, m-n) = Re[q_m {k_n}^* e^{i(m-n)\theta}] = Re[(W_q x_m)(W_k x_n)^* e^{i(m-n)\theta}]
\end{aligned} \tag{2}$$

$$\theta = [\theta_0, ..., \theta_{d/2-1}] \tag{3}$$

BERT model outputs mainly two types of information: sequence-level and word-level. WoBERT still uses this method. As shown in the following formula:

$$\text{hidden\_rep,cls\_head} = \text{WoBERT}\,(\text{token\_ids,\_attention\_mask}) \tag{4}$$

hidden_rep is output at the word level and is the embedded representation of all characters, with a size of [batch, sequence, hidden_size].*cls*_head is the sequence-level output, which represents the semantic information of the text segment and has a size of [batch,hidden_size].

In the past, text classification algorithms directly used *cls_hidden* as the text feature for loss calculation, ignoring the original representation of characters and words, which resulted in the loss of semantic information. In this paper, we study the further use of character embedding to supplement semantics. As shown in the following equation:

$$\text{position\_hidden} = \text{RoPE}\ (\ \text{hidden\_rep}\ ) \tag{5}$$

Equation 5 indicates that word vectors are modeled using rotation position encoding to capture relative positional information between characters and words, which better handles long text sequences.

3.3. Feature Fusion Layer

For the extracted feature knowledge, previous algorithms often use multi-level methods such as stitching, addition, and attention mechanisms for knowledge fusion, which can easily lead to overfitting of the model due to the accumulation of channels. In this regard, this study draws on Multi Sample Dropout and D-Mixup (Li et al., 2024) to propose an improved hierarchical feature fusion method, MD-Mixup, as shown in Table 1:

Table 1: MD-Mixup algorithm

```
Begin
pool_hidden = Mean(position_hidden)
out = 0
loss = 0
For i In Range(circle_num):
    cls_hidden = D_Mixup(cls_hidden,pool_hidden)
    out += FCL(cls_hidden)
    loss += Cross_entropy(out,label)
Return out,loss
```

3.4. Output Layer

The fused mixed features are used as the linear output of the category, and the measure representing the probability of each category is the logits. The classification loss of the model is calculated using the cross-entropy of multiple classifications, and the specific formula 6 is as follows:

$$\text{Loss} = -\frac{1}{N}\sum_{i=1}^{N}\log\left(\frac{\exp\left(\log it_{i,y_i}\right)}{\sum_j \exp\left(\log it_{i,j}\right)}\right) \tag{6}$$

4. Experiment

4.1. Datasets

The evaluation was performed using the THUCNews and SMP2020 datasets,and the specific quantity distribution is shown in Table 2.

Table 2: Dataset

Name	Train	Dev	Test
THUCNews	180000	100000	100000
SMP2020	8606	2000	3000

4.2. Experimental Setup

Hardware settings: PyTorch-1.10.0 framework development, GPU is NVIDIA GeForce-RTX-3090 (24G).

Model parameters: WoBERT hidden layer output 780; RoPE hidden layer output 780; Dropout is 0.5; optimizer is Adam; learning rate is 1e-5. Since the THUCNews data is sufficient, the SMP2020 data is sparse and the text lengths are different, different settings were used for some parameters: epoch 3, batch 128, and pad 32 for THUCNews; epoch 5, batch 64, and pad 150 for SMP2020.

4.3. Evaluation method

$$\text{Accuracy } = \frac{TP+TN}{TP+TN+FP+FN} \tag{7}$$

$$\text{Precision } = \frac{TP}{TP+FP} \tag{8}$$

$$\text{Recall } = \frac{TP}{TP+FN} \tag{9}$$

$$F1 = \frac{2\times \text{ Precision } \times \text{ Recall}}{\text{Precision } + \text{ Recall}} \tag{10}$$

where TP denotes true examples, TN denotes true negative examples, FP denotes false positive examples, and FN denotes false negative examples.

4.4. Experimental Result

Tables 3-4 above shows that:

(1) Compared with Mengzi, WFDR_MGMCHNN, TLIFC-RoBERTa, GLFDF, IRPN_RFFL, the F1 score is improved by 0.43, 2.17, 1.01, 2.22, 0.72 percentage points respectively; Increase the Acc score by 0.49, increase by 2.16, decrease by 0.25, and increase by 2.22 percentage points, respectively.

(2) Compared with ERNIE+DPCNN+BiGRU, RBA, MFS-BiLSTM, BCBA, ChineseBERT-MCFF improved the F1 score by 3.72, 3.43, 3.11, 2.88, and 2.31 percentage points, respectively; and the Acc score by 0.49, 3.45, 3.14, 2.89, and 2.33 percentage points, respectively.

(3) Compared with methods such as GLSTCM-HNN, TLIFC-RoBERTa, BERT-FPnet-1, ERNIE+DPCNN+BiGRU, and ChineseBERT-MCFF, which introduce knowledge labels, feature projections, or combinations of DPCNN, BiLSTM, and multi-head attention networks, the method in this paper only uses a single RoPE neural network for feature extraction and fusion, which is conducive to reducing model complexity and more accurately capturing the semantic representation of the text.

Table 3: THUCNews comparative test results Unit: %

Model	Acc	F1
Mengzi (Chen et al., 2024)	94.29	94.29
WFDR_MGMCHNN (Ma and Huang, 2024)	92.56	92.55
TLIFC-RoBERTa (Yu et al., 2024)	**94.97**	93.71
GLFDF (Zheng and Zhang, 2024)	92.50	92.50
IRPN_RFFL (Lu et al., 2024)	-	94.00
WR-Mix (our)	94.72	**94.72**

Table 4: Comparison of the results of the SMP2020 test in %

Model	Acc	F1
ERNIE+DPCNN+BiGRU (Yang et al., 2023)	75.16	75.18
RBA (Zhang et al., 2022)	75.48	75.47
MFS-BiLSTM (LI et al., 2021)	75.79	75.79
BCBA (Bao et al., 2021)	76.04	76.02
ChineseBERT-MCFF (Gao et al., 2023)	76.60	76.59
WR-Mix (our)	**78.93**	**78.90**

4.5. Ablation Experiment

To verify the impact of each part of this model on the whole, ablation experiments were conducted on two datasets: 1) WoBERT: only WoBERT is used for feature extraction; 2) WR-Mix (W/o RoPE): the feature extraction layer is removed, and the initial hidden_rep is used as the input of the fusion layer; 3) WR-Mix (W/o MD-Mixup): the feature fusion layer is removed, and the features are concatenated using the additive method.The specific results are shown in Figures 3 and 4.

The above chart shows that

(1) in the THUCNews dataset, +RoPE improves the Acc and F1 scores of the baseline model by 0.15 and 0.05 percentage points, respectively; +MD-Mixup improves the Acc and F1 scores of the baseline model by 0.36 and 0.26 percentage points, respectively; and the model in this paper improves the Acc and F1 scores of the baseline model by 0.46 and 0.26 percentage points, respectively.

(2) In the SMP2020 dataset, the +RoPE model's ACC and F1 scores decreased by 0.43 and 0.3 percentage points compared to the baseline model; the +MD-Mixup model's ACC and F1 scores increased by 1.27 and 1.39 percentage points compared to the baseline model; and the model in this

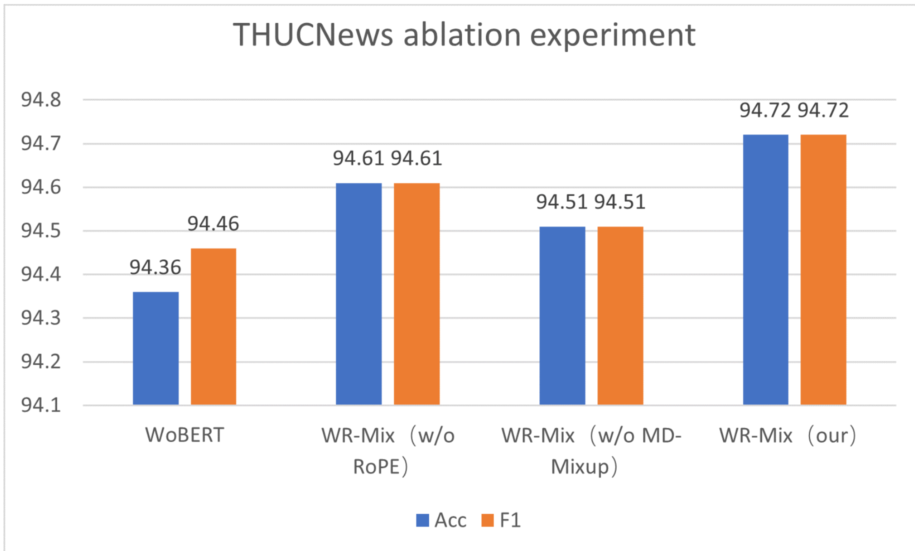

Figure 3: THUCNews ablation experiment diagram.

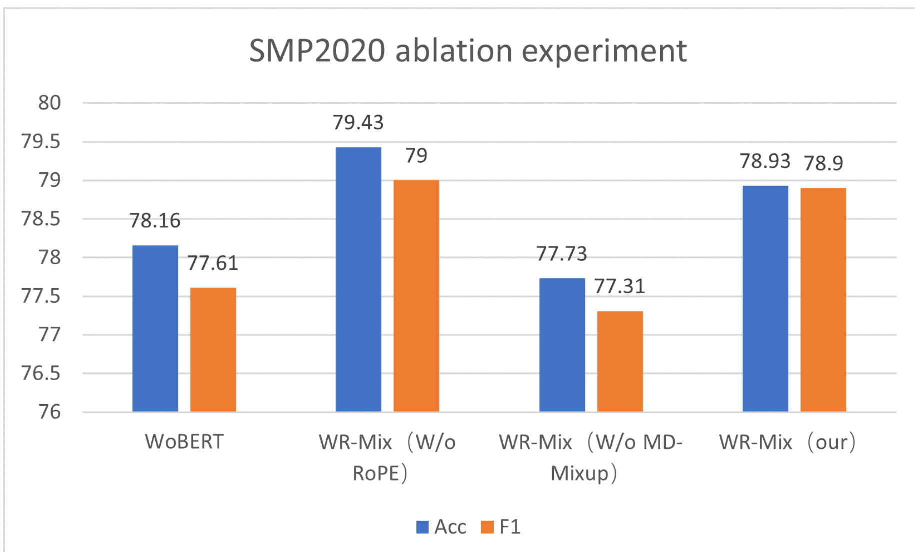

Figure 4: SMP2020 ablation experiment diagram.

paper's ACC and F1 scores increased by 0.77 and 1.29 percentage points compared to the baseline model.

From the above analysis, we can see that: MD-Mixup alone has a better effect than RoPE, while adding the RoPE indicator to the SMP2020 dataset actually reduces it. This is because the SMP2020 data format is not fixed and sparse compared to THUCnews, and it is understood that a single overlay network has limited model performance, which can easily lead to overfitting problems. By using the MD-Mixup regularization technique, features can be effectively discretized, vector interactions can be strengthened, and hidden information can be more accurately captured.

5. Conclusions

This paper proposes a word-level enhancement text classification method based on rotation position encoding and Hierarchical Features fusion. Based on WoBERT, rotation position encoding is used to model the relative information-rich underlying information of sequence output features. The proposed MD-Mixup fusion method strengthens vector interaction and data connection, effectively improving problems such as context loss and information loss. The effectiveness and feasibility of the model are verified on the THUCNews and SMP2020 datasets.

Acknowledgments

Research Project by Liangshan Minzu Education Big Data Research Center of Xichang Minzu Preschool Normal College (25LMJD01)

References

Tong Bao, Ni Ren, Rui Luo, Baojia Wang, Gengyu Shen, and Ting Guo. A bert-based hybrid short text classification model incorporating cnn and attention-based bigru. *Journal of Organizational and End User Computing (JOEUC)*, 33(6):1–21, 2021. ISSN 1546-2234. doi: 10.4018/JOEUC.294580.

Xuesong Chen, Heng Li, and Haochang Wang. Short text classification combining attention mechanism and mengzi model. *Computer and Modernization*, (09):101–106+120, 2024. ISSN 1006-2475.

Alexis Conneau, Holger Schwenk, Loïc Barrault, and Yann Lecun. Very deep convolutional networks for text classification. In Mirella Lapata, Phil Blunsom, and Alexander Koller, editors, *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers*, pages 1107–1116, Valencia, Spain, April 2017. Association for Computational Linguistics.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. BERT: Pre-training of deep bidirectional transformers for language understanding. In Jill Burstein, Christy Doran, and Thamar Solorio, editors, *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186, Minneapolis, Minnesota, June 2019. Association for Computational Linguistics. doi: 10.18653/v1/N19-1423.

Hui Gao, Yaling Xun, and Linqing Wang. Text sentiment analysis based on multi-channel fusion feature network. *Computer Technology and Development*, 33(11):175–181, 2023.

Zhongxiang HUANG and Ming LI. Albert combined with bi-directional network for text categorization. *Computer and Modernization*, (010):000, 2022.

Rie Johnson and Tong Zhang. Deep pyramid convolutional neural networks for text categorization. In Regina Barzilay and Min-Yen Kan, editors, *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 562–570, Vancouver, Canada, July 2017. Association for Computational Linguistics. doi: 10.18653/v1/P17-1052.

Yoon Kim. Convolutional neural networks for sentence classification. In Alessandro Moschitti, Bo Pang, and Walter Daelemans, editors, *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1746–1751, Doha, Qatar, October 2014. Association for Computational Linguistics. doi: 10.3115/v1/D14-1181.

Zhenzhong Lan, Mingda Chen, Sebastian Goodman, Kevin Gimpel, Piyush Sharma, and Radu Soricut. ALBERT: A lite BERT for self-supervised learning of language representations. *CoRR*, abs/1909.11942, 2019.

Mingjian Li, Yuquan Zhou, and Huaiyang Li. Entity extraction research with relative position encoding and masked conditional random fields : Named entity recognition. In *2024 3rd International Conference on Electronics and Information Technology (EIT)*, pages 858–863, 2024. doi: 10.1109/EIT63098.2024.10762428.

Wei-Jiang LI, Fang QI, and Zheng-Tao YU. Sentiment classification method based on multi-channel features and self-attention. *Journal of Software*, 32(9):2783, 09 2021. doi: 10.13328/j.cnki.jos.005992.

Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. Roberta: A robustly optimized BERT pretraining approach. *CoRR*, abs/1907.11692, 2019.

Qianhui Lu, Yu Zhang, Mengling Wang, Tingwei Wu, and Yuzhong Shan. A text classification model for nuclear power equipment quality based on improved cyclic pooling network. *Computer Applications*, 44(07):2034–2040, 2024. ISSN 1001-9081.

Changlin Ma and Hao Huang. Text classification based on multi granularity channels and dual reinforcement. *Journal of Huazhong University of Science and Technology (Natural Science Edition)*, 52(11):133–139, 2024. ISSN 1671-4512. doi: 10.13245/j.hust.240402.

Yuan Ni, Shihao Liao, and Jian Zhang. Chinese named entity recognition based on wobert and adversarial learning. *Computer Engineering*, 50(11):119–129, 2024. ISSN 1000-3428. doi: 10.19678/j.issn.1000-3428.0068258.

Zhaoyang Ning, Qing SHEN, Xiulanand HAO, and Knag ZHAO. A short text classification framework based on improved adversarial learning and fused features. *Computer and Modernization*, (04):66–76, 2024. ISSN 1006-2475.

Jianlin Su. Speed up without dropping points: word granularity-based chinesewobert, Sep 2020. URL `https://spaces.ac.cn/archives/7758`.

Jianlin Su, Yu Lu, Shengfeng Pan, Bo Wen, and Yunfeng Liu. Roformer: Enhanced transformer with rotary position embedding. *CoRR*, abs/2104.09864, 2021.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention is all you need. *CoRR*, abs/1706.03762, 2017.

Senqi Yang, Xuliang Duan, Zhan Xiao, Songsong Lang, and Zhiyong Li. Agricultural news text classification based on ernie+dpcnn+bigru. *Computer Applications*, 43(05):1461–1466, 2023. ISSN 1001-9081.

Xingrui YANG, Shouwei ZHAO, Ruxue ZHANG, Yehui Tao, and Xingjun Yang. Bilstm-attention text classification model with improved bert word vectors. *Sensors and Microsystems*, 42(10): 160–164, 2023. ISSN 2096-2436. doi: 10.13873/J.1000-9787(2023)10-0160-05.

Hang Yu, Yanling Zhou, Mengxin Zhai, and Han Liu. Text categorization based on pre-trained model and label fusion. *Computer Applications*, 44(3):709–714, 3 2024.

Jun Zhang, Li Zhang, Fanfan Shen, Hai Tan, and Yanxiang He. Roberta fusion of bilstm and attention mechanism for implicit sentiment analysis. *Journal of Computer Engineering & Applications*, 58(23):142–150, 2022. ISSN 1002-8331.

Wenjun Zheng and Shunxiang Zhang. A text classification model with dynamic fusion of global and local features. *Journal of Xi'an Engineering University*, 38(4):97–105, 2024.