Computer Life
ISSN: 1819-4818 | Vol. 12, No. 2, 2024

The Text Classification Method Based on BiLSTM and Multi-Scale CNN

Bo He, Yongfen Yang*, Lei Wang, Jingxuan Zhou

School of Computer Science and Engineering, Chongqing University of Technology, Chongqing, 400054, China

* **Corresponding author**: Yongfen Yang (Email: yangff0518@163.com)

Abstract: Text classification is an important fundamental technique for efficiently managing the huge amount of textual information on the Internet. In the past few years, due to the unprecedented success of deep learning research, the task of text categorization has gradually become the center of gravity in the field of natural language processing, and the algorithms applied to it have also transitioned over time from traditional machine learning methods to neural network models based on deep learning, as well as the emerging attention mechanisms and pre-trained language models with good results. This paper outlines the traditional machine learning methods and deep learning methods involved in text classification, comprehensively analyzes the research progress and achievements of deep learning models for text classification, and compares the advantages and disadvantages of each. Then the evaluation metrics used for text classification and commonly used labeled datasets are introduced. Finally, the challenges faced by the text classification task and the difficulties to be further researched are summarized and outlooked to provide reference and support for future researchers in this field.

Keywords: Text classification; Neural networks; Deep learning; Natural language processing.

1. Introduction

Natural Language Processing (NLP) is a cutting-edge field at the intersection of computer science, artificial intelligence, and linguistics [1]. Text classification, as a basic and necessary task in this field, aims to classify text data into specific categories based on the text content, and has been widely used in spam classification [2], sentiment analysis, topic classification [3], question and answer systems [4], and personalized news recommendation [5-6]. Early text classification is mostly done manually, and this method is time-consuming and labor-intensive, with poor portability and low accuracy. However, with the rapid development of science and technology and the wide application of computers, text information has been fully electronic, and a huge amount of text comes to the surface. In the face of huge text data, the traditional manual classification methods can no longer be processed and refined within a reasonable timeframe to help users make decisions. Based on this, researchers have utilized computers to perform text classification tasks and proposed various text classification algorithms, which not only accelerate the classification speed but also improve the accuracy of classification.

Deep learning is a machine learning method that enables computers to automate the learning process by integrating multiple layers of neural networks. In comparison to traditional machine learning methods, deep neural network models not only reduce the cost of manual feature design required for dealing with different problems and realise automated machine learning, but also can effectively deal with large-scale data and complex nonlinear relationships, which is highly suitable for application in the field of text classification. This paper presents an overview of the current development of deep learning in the field of text classification, detailing the latest research progress and technical approaches to text classification.

2. Text classification method based on traditional machine learning

When doing tasks related to text classification, how to extract features is a key issue. Traditional machine learning tasks require manual extraction of text features. Other commonly used traditional machine learning classification models are Support Vector Machines (SVM)[7], Naive Bayes (NB), K-Nearest Neighbor (KNN), and Random Forest Algorithm. The parameters of NB are small, it is not too sensitive to missing data, the algorithm is simple and efficient, suitable for high dimensional data, and the results are stable. efficient, suitable for high-dimensional data, and stable. However, the independence assumption is often not valid in practical applications, which may affect the classification accuracy. Support vector machine can solve high-dimensional and nonlinear problems. It has high generalization ability, but the training time is long, and it is sensitive to parameter selection and kernel function. KNN decides the class of the new text mainly by calculating the distance between the new text and each text in the training set, and selecting the k neighbors with the closest distance for voting. Therefore, it is more suitable than other methods for datasets that are to be classified by intersection or overlap of class domains.

Furthermore, as data sizes continue to grow, the computational efficiency of these algorithms is declining, which constrains their capacity to process large-scale datasets and ultimately affects the accuracy of the final classification. Consequently, during this period, researchers commenced the investigation of more efficient and automated feature extraction methodologies, with the gradual introduction of deep learning algorithms to address these issues.

3. Text classification method based on Deep learning

Deep learning is a subfield of machine learning that aims to learn features and patterns from large amounts of data through a multilayer neural network structure that mimics the neuronal connections of the human brain. Each layer of the neural network performs a nonlinear transformation of the input data, progressively extracting higher-level abstract features. The advantage of this approach is that it obviates the need for manual feature design. The network is able to learn useful features from data automatically and is capable of handling high-dimensional, unstructured data when dealing with complex data.

3.1. Convolutional neural network based method

CNN (Convolutional Neural Networks, CNN) is one of the typical algorithms for deep learning [8].The structure of CNN mainly includes input layer, convolutional layer, pooling layer, fully connected layer, output layer and so on. Among them, the convolutional layer and pooling layer are the typical structures of CNN that distinguish it from other neural networks.The model structure of CNN is shown in Figure 1.

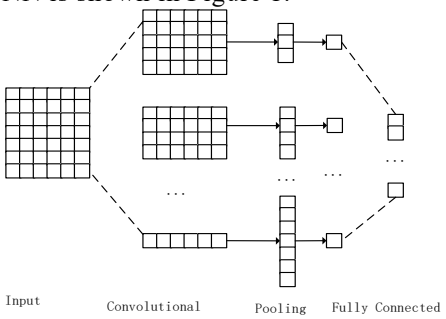

Figure 1: CNN model structure

Kim [9] previously employed Convolutional Neural Networks (CNN) for text categorisation. The convolutional neural network (CNN) architecture was employed to represent text at the character level. This involved combining a multi-task learning framework with a recurrent neural network (RNN) structure, which enabled the model to learn across multiple related tasks. Additionally, a recursive structure was employed to capture contextual information, and a maximum pooling layer was used to identify key components in the text. Although distributed representation methods are capable of automatically learning features from large-scale datasets, they are limited by poor interpretability, long training times, and the necessity of numerous training parameters. In order to enhance the classification efficacy of the TextCNN model, literature [10-12] devised a deep CNN model by augmenting the number of layers and expanding the field of view of the convolutional kernel to procure information. The results demonstrate that the deep CNN model constructed by increasing the number of layers from the vertical direction does not significantly enhance the classification efficacy of short text.

Consequently, numerous scholars have adopted alternative methodologies to enhance the TextCNN model by modifying the configuration of the convolutional and pooling layers of the model in a horizontal direction and extracting a greater number of short text features. Guo et al.[13] constructed an enhanced CNN model with the help of jump convolution and K-Max pooling operation to refine the short text feature extraction. At the same time, they retained the first area of the largest feature values in the pooling layer to obtain short text features from multiple dimensions. Wang et al. [14] enhanced the TextCNN model by constructing an N-gram discontinuous sliding window and a K-Max average pooling operation.

In addition to the research on the structure of the convolutional neural network (CNN), literature [15-17] has also conducted research on the effect of different granularity levels of input on the classification effect. The classification effect of word granularity and word granularity is generally superior to that of input by sentence, due to the inherent lossy compression process involved in input at the sentence level. This enhanced model, to a certain extent, enhances the efficiency of CNN in acquiring text features with a more comprehensive feature representation. Nevertheless, due to the paucity of content in the short text itself, the improved model still encounters the challenge of insufficient information in acquiring text features.

3.2. Long and short-term memory network method

Long short-term memory (LSTM) is a specific type of recurrent neural network (RNN), which employs a reverse chronological order gradually backpropagation method. However, when the text sequence is lengthy, it becomes challenging to address gradient disappearance or gradient explosion issues, which can impede the ability to capture long-distance dependence relationships between text. In light of these considerations, numerous researchers have sought to enhance the efficacy of these networks. Among the most impactful of these enhancements is the introduction of a gating mechanism. Hochreiter et al. [18] proposed the Long Short-Term Memory Network (LSTM), whose core structure, comprising a single LSTM unit, is depicted in Figure 2. In practice, it has demonstrated efficacy in the domains of time series prediction and natural language processing.

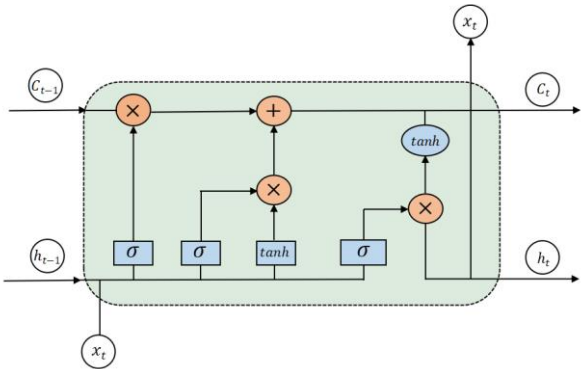

Figure 2: LSTM cell structure

According to Figure 2, the formula for updating the state of the LSTM structure is as follows:

$$v_t = \sigma(W_v \cdot [h_{t-1}, x_t] + b_v) \quad (1)$$

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \quad (2)$$

$$\widetilde{C_t} = tanh(W_C \cdot [h_{t-1}, x_t] + b_C) \quad (3)$$

$$C_t = v_t \times C_{t-1} + o_t \times \widetilde{C_t} \quad (4)$$

$$j_t = \sigma\left(W_j \cdot [h_{t-1}, x_t] + b_j\right) \quad (5)$$

$$h_t = j_t \times tanh(C_t) \quad (6)$$

where $W_{\mathrm{v}} \in \mathbb{R}$、$W_o \in \mathbb{R}$、$W_{\mathrm{C}} \in \mathbb{R}$、$W_{\mathrm{j}} \in \mathbb{R}$,are the corresponding weight sizes of the model, b_{v}、b_o、b_{C}、b_{j} are the biases, c_t is the stored state cell at time t, σ is the Sigmoid activation function, h_t denotes the final output of the LSTM model, h_{t-1} is the output of the previous time step, c_{t-1} is the previous cell state at a time step, and $tanh$ denotes the hyperbolic tangent activation function.

Furthermore, the researchers introduced a comparison model, Tree-LSTM, which outperforms TextRNN on the SST-1 dataset and achieves superior classification results. This is due to the fact that short texts exhibit a specific structure, and LSTM is a network that transfers information in a linear manner according to the temporal order. This makes it challenging for LSTM to learn structural information such as internal dependencies and the syntax of short texts.

In order to capture the semantic information of the sequence more comprehensively, PALIWAL et al. [19] further optimised the LSTM network structure and proposed a bidirectional LSTM model (BiLSTM). This model is capable of scanning from the forward and backward directions of the sequence simultaneously, and is able to take into account the sequence information of both past and future moments in the feature extraction. The BiLSTM model consists of a forward LSTM and a backward LSTM to form the hidden state sequences $(\overrightarrow{h_1}, \overrightarrow{h_2}, \overrightarrow{h_3}, \cdots, \overrightarrow{h_n})$ and $(\overleftarrow{h_1}, \overleftarrow{h_2}, \overleftarrow{h_3}, \cdots, \overleftarrow{h_n})$. By superposition, the forward hidden states from time $t = 1$ to $t = T$ and the reverse hidden states from time $t = T$ to $t = 1$ are weighted in positional order to obtain a new sequence of hidden states, and the output is $H: [h_1, h_2, h_3, \cdots, h_n]$. The calculation formula is as follows:

$$h = \left[\vec{h} \oplus \overleftarrow{h}\right] \quad (7)$$

The essential information and long-term dependencies of textual contextual semantics are completely encapsulated following a bi-directional encoding process, thereby enabling the model to utilise a more comprehensive array of features.

Additionally, scholars have proposed the integration of the BiLSTM model with the CNN model. Zhou et al.[20] have developed the BLSTM-2DPooling and BLSTM-2DCNN models, which consider both the temporal and textual dimensions of the input text, thus enabling the capture of more comprehensive semantic features. J. Jin et al.[21] proposed a short Chinese text classification model based on CNN and BiLSTM, which can focus on extracting the key information of the text to improve the accuracy of the text. Xu et al.[22] proposed a multi-scale BiLSTM-CNN sentiment classification model, which can classify sentiment polarity in a more detailed way. In the hybrid BiLSTM-CNN, the input text information is first decoded by the BiLSTM model and then subjected to a convolutional operation. This captures the semantic information between words and further reduces the feature dimensions through the maximum pooling operation. This reduces the parameters of the model and simultaneously extracts the important features, thus effectively reducing the overfitting of the model to the data.

3.3. 2.3 Graph Neural Network method

With the growing interest in graph neural networks, GNN-based models have demonstrated remarkable performance in encoding the syntactic structure of sentences in text classification tasks. To enhance the efficacy of the network, Velickovic et al. [23] introduced the graph attention network (GAT), which employs the attention mechanism as an aggregation function to aggregate the information of the central node and the neighbouring nodes, thereby enhancing the interpretability of the GCN, whose weights are calculated as:

$$e_{ij} = LeakyReKU\left(a^T\left[Wh_i \parallel Wh_j\right]\right) \quad (8)$$

$$\alpha_{ij} = \frac{exp(e^{ij})}{\sum_{k \in N_i} exp(e^{ij})} \quad (9)$$

In [24-25], a bi-graph convolutional network model for aspect-based sentiment analysis was proposed which takes into account both complementarity of syntactic structures and semantic relevance. In a related vein, Literature [26] put forth the concept of encoding words using directed acyclic graphs and devised a directed acyclic neural network to operationalize this concept. The model offers a more intuitive approach to simulating the flow of information between the context of a remote dialogue and the nearby context. In a novel approach, Literature [27] proposes a graph-based model for emotion-triggering paths. This model employs commonsense knowledge to enhance the semantic dependencies between candidate clauses and emotion clauses. Yang T. et al. [28] subsequently proposed a heterogeneous graph attention network, which modelled documents, topics, and entities as nodes of a textual graph and constructed edges between topic-document, entity-document, and entity-entity to capture relational information. A dual attention mechanism is also designed to capture the importance of different neighbouring nodes as well as the importance of different types of nodes. This reduces the noise information, enhancing the interpretability of the model.

3.4. Pre-trained models

Pre-training models play an important role in the field of deep learning and NLP. In 2018, a deep pre-training model, BERT, was proposed for NLP tasks using the Transformer architecture [29-30]. The BERT model architecture is shown in Figure 3. Due to its excellent performance, the BERT model is widely used in downstream tasks of NLP. This type of model is trained with a large range of unlabelled data to produce rich contextual semantic information, which is able to extract the connection between words very well. Consequently, most researchers are moving towards pre-training models, and there are many models based on BERT and its variants in various lists of natural language tasks. Cui Yiming et al. [31] proposed a pre-training language model specifically for the Chinese language, MacBERT, which modifies the word masking based on BERT. This replaces the masked words with similar words, thus reducing the gap between the pre-training and fine-tuning phases. In order to address the limitation of BERT input length, the HadaBERT model was proposed [32], which comprises a local encoder and a global encoder. The local encoder encodes the document in segments using

multiple BERT models, while the global encoder synthesises the results of the segmented encoding into the final representation of the document based on Attention.

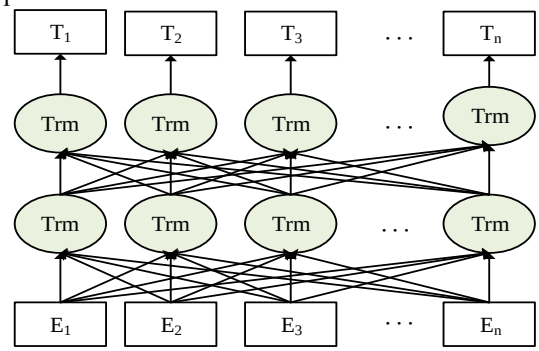

Figure 3: BERT model structure

Despite the emergence of numerous enhanced models, the most widely utilised one is BERT, taking into account the resource consumption and classification effectiveness. Language models are capable of effectively learning global semantic representations by being pre-trained under large-scale preconditioning, and they significantly facilitate natural language processing tasks, including text classification [33]. The question of how to apply pre-trained models that have been trained on large datasets to downstream tasks is a highly worthwhile area of research.

Transfer learning is the research idea of pre-training a model in large batches before applying it to another contextual model. Li Bion et al [34] performed sarcastic text detection by combining it with knowledge transfer from other resources, making full use of the transferred resource knowledge to improve the model and enhance the model performance. Sun Xiaoyan et al [35] designed a new migration learning strategy for cross-domain migration of false comments, which can effectively capture the same features in the text, followed by the use of semi-supervised collaborative mechanism to identify false information by combining the migrated data. The combination of migration learning and semi-supervised collaboration mechanism improves the accuracy of false comment detection. The related research on migration learning mainly improves the idea on the pre-training model, training datasets to fine-tune the model, and various improvement methods are taken and verified in combination with the actual task.

3.5. Overview of the methodology

In recent years, numerous deep learning models for text classification have been proposed, as illustrated in Table 1. These models have been applied to a range of tasks, including sentiment analysis (SA), topic labelling (TL), aspect-based sentiment analysis (ASBS), short text categorisation (STC), conversation emotion recognition (ERC) and emotion cause extraction (ECE). In order to enhance the efficacy of model classification, researchers have endeavoured to implement a plethora of methodologies, including the incorporation of neural network models or attention mechanisms, or the improvement of conventional neural network models. Furthermore, some researchers have investigated text classification techniques based on pre-trained models. These employ an unsupervised approach to automatically mine semantic knowledge, which is then used to construct pre-trained targets. This enables the learning of structural information in the speech-captured text, with classification performance that is unparalleled by other methods.

Table 1: Different Models for Text Classification

Model	Time	Method	Application
CNN	2018	LSTM-CNN	SA
	2022	CRFA	STC
	2022	TextConvoNet	SA
LSTM	2021	BiLSTM-CRF	ASBS
	2022	Hierarchical-LSTM	SA
	2022	Attention-based LSTM	SA, STC
GNN	2017	GAT	Twitter, Laptop
	2021	KAG	SINA CITY NEWS
	2021	DG-LSTM	IPC,CPC,CrPC
BERT	2019	RoBERTa	AGnews,IMDB
	2021	BERTweet	SA,ERC
	2022	Bert-Base	LIAR

The four text classification methods mentioned above represent the principal models in the field of deep learning. Each of them is effective in handling different types of data and tasks, thus exhibiting unique strengths. Initially employed for image processing, convolutional neural networks (CNN) have also been used for text classification[11]. The text is initially converted into word embeddings, after which local features are extracted by a 1D convolutional layer and a pooling layer. These layers are capable of capturing local patterns in the text, such as n-grams. Convolutional operations are easily parallelisable and are suitable for large-scale data processing. Furthermore, the convolutional kernel is shared across locations in the text, reducing the number of parameters. The disadvantages are also more prominent. It is difficult to capture long-distance dependencies, and when the feature selection depends on the size of the convolution kernel, it is necessary to adjust the size of the convolution kernel in order to capture features of different lengths. LSTM makes up for some of the shortcomings of CNN by The introduction of memory units and gating mechanisms (input gates, oblivion gates, and output gates) enables the capture of long-term dependencies in sequences[19]. Furthermore, it allows for the efficient processing of long texts, the capture of long-distance dependencies, and the suitability for the processing of sequential data. However, the training time is lengthy, the demand for computational resources is considerable, and the difficulty of parallelisation is high, which presents a challenge for parallelisation.

GNN represents the text as a graph structure, and learns

the node representation through the Message Passing mechanism (Message Passing) and the aggregation of neighbouring node features. The construction of text graphs is a common method, with two main approaches: the Word Co-occurrence Graph and the Dependency Tree Graph[25]. These methods are suitable for text data represented by various graph structures, as they are able to capture the complex relationships and structures between words. However, the construction and computation overheads are considerable, particularly for large-scale data sets, and the performance is contingent upon the quality of the constructed graph structure. BERT is based on the Transformer architecture, which enables contextual, bidirectional understanding through bidirectional encoders. The model is initially trained on a large corpus of textual data, after which it is fine-tuned for specific downstream tasks[34]. Its bi-directionality enables it to understand the meaning of words in different contexts, and the pre-trained model can be applied to a variety of downstream tasks, reducing the need for labelled data. It performs well in many NLP tasks, especially text classification. However, the high computational resource requirements of pre-training and fine-tuning, particularly the need for high-performance GPUs with numerous parameters, complex deployment and optimisation, and the inherent complexity of the models, result in slower inference, making it difficult to apply to real-time tasks. The selection of an appropriate model, tailored to the specific requirements of a given task, represents a key strategy for leveraging the potential of deep learning.

4. Evaluation metrics

The most commonly employed evaluation metrics in text classification tasks include accuracy, precision, recall, F1-score and confusion matrix. The aforementioned evaluation metrics facilitate the assessment of the model's performance in classification tasks.

The accuracy of a classifier is defined as the ratio of correctly classified samples to the total number of samples. The metric gauges the model's overall performance across all categories.

$$Accuracy = \frac{TP+FN}{TP+TN+FP+FN} \quad (10)$$

The proportion of samples predicted by the model to be in the positive category that are actually in the positive category is referred to as the accuracy of the model. The metric gauges the precision of the model in identifying positive categories.

$$Precision = \frac{TP}{TP+FP} \quad (11)$$

Recall is the proportion of samples that are actually in the positive category that are correctly predicted to be in the positive category by the model. It measures the ability of the model to identify positive category samples.

$$Recall = \frac{TP}{TP+FN} \quad (12)$$

The F1 value represents the reconciled mean of precision and recall, which takes into account both the precision and recall of the model.

$$F1 = \frac{precision \times recall \times 2}{precision+recall} \quad (13)$$

The confusion matrix is a two-dimensional matrix that is employed to illustrate the classification outcomes of the model. Each row represents the true category, while each column represents the category predicted by the model. The fundamental elements of the confusion matrix comprise true positives (TP), false positives (FP), true negatives (TN) and false negatives (FN). As shown in Table 2.

Table 2: Confusion matrix

True Value	Predicted Value	
	Positive	Negative
Positive	TP(True Positive)	FN(False Negative)
Negative	FP(False Positive)	TN(True Negative)

5. Datasets

The training and evaluation of text classification models necessitates the availability of a substantial number of labelled datasets. The labelled datasets selected for the validation of deep learning-based text classification models are predominantly specific domains, such as movie reviews, news, social comments, and so forth. In the majority of cases, they are employed to classify the sentiment polarity of the reviews. It can be observed that Chinese labelled datasets have the potential for further improvement, both in terms of the number of instances and the breadth of the domains represented. This section presents an overview of open-source datasets in text classification research, with a focus on commonly used text labelling datasets. The number of documents in the training set, the number of categories in the test set, the average sentence length, and the source are presented in tabular form. As illustrated in Table 3.

Table 3: Common Text Label Datasets

Datasets	Number of documents in training set	Number of documents in test set	categories	Average length of sentences	source
THUCNews	41000	11200	14	18.1	-
SogouNews	50000	10000	10	29.3	
Twitter	8204	3005	3	19.0	
IMDB	25000	12500	10	239.0	
Yelp	1566574	380000	2	153.0	
AG News	120000	7600	4	7.0	
SST-1	8544	2210	5	18.0	
SST-2	67000	18000	2	19.0	
DBpedia	560000	70000	14	55.0	
R8	5482	2189	8	65.7	
Google Snippets	10060	2280	8	18.0	
TREC	5952	500	6	10.0	
MR	7108	3554	2	20.4	

6. Prospecting

The research on the application of deep learning in text

classification has made considerable progress, yet there remain significant challenges in the areas of pre-training model enhancement, expanding the feature research model integration, and so forth, which require further investigation and research.

However, contemporary scholars rarely apply the pre-training model and feature fusion idea simultaneously to optimise the model. Instead, they conduct research on a single direction, such as introducing the BERT pre-training model or not applying the pre-training model to optimise the feature extraction process. As deep learning continues to evolve, techniques such as knowledge distillation and transformers are continually proposed. The integration of these techniques into models offers the potential to leverage the strengths of each while enhancing the model's generalisation and robustness. Consequently, text classification methods based on the integration of multiple models have attracted considerable interest from numerous scholars.

Nowadays, the parameter tuning of models has the potential to enhance the performance of classification. Nevertheless, the challenge remains in meeting the substantial training data requirements for deep learning. Deep learning models have extremely high requirements in terms of the quantity of training data and the time required for computation, which are far from being comparable to those of other algorithms. The accuracy of the training results of deep neural networks is largely contingent upon the quantity of training data. Concurrently, the training time of deep neural networks increases in proportion to the complexity of the network model. One avenue for future research will be to identify methods for maintaining the performance of the model while reducing the size of the model continuously.

The performance of deep learning models is contingent upon the quality and balance of the sample data. In the event that the sample data is of poor quality and imbalanced, it will result in a decline in classification accuracy. Furthermore, the currently available short text labelling datasets are relatively small and concentrated in a few specific domains. Consequently, it is of paramount importance to reinforce related research and construct high-quality datasets based on multiple domains with the objective of achieving more accurate classification performance for text.

7. Conclusion

This paper presents a comprehensive overview of text classification methods and applications based on deep learning. It begins with a brief introduction to traditional machine learning methods and then proceeds to analyse the current state of research in this field. Finally, it presents a detailed analysis of relevant labelled datasets. Finally, the future development of deep learning methods in text classification will be discussed.

Acknowledgments

This research is supported by the 2024 High-Quality Innovation Project for Postgraduates of Chongqing University of Technology(NO.gzlcx20243174).

References

[1] Tiejun Zhao, Muven Xu, Antony Chen. A review of natural language processing research[J]. Journal of Xinjiang Normal University (Philosophy and Social Science Edition), 2023,1-23.

[2] WU SH,CHEN S P.Spam message recognition based on TFIDF and Self-Attention-Based Bi-LSTM[J].Computer Application Systems,2020,29(9):171-177.

[3] GAN C Q, FENG Q D, ZHANG Z F. Scalable multi-channel dilated CNN-BiLSTM model with attention mechanism for Chinese textual sentiment analysis[J]. Future Generation Computer Systems, 2021, 118(1): 297-309.

[4] ROY A M. An efficient multi-scale CNN model with intrinsic feature integration for motor imagery EEG subject classification in brain-machine interfaces[J]. Biomedical Signal Processing and Control, 2022, 74(4): 1-14.

[5] Ming Li,Ying Li,Wangqin Lou,Lisheng Chen. A hybrid recommendation system for Q&A documents[J].Expert Systems With Applications,2022.

[6] Chuhan Wu, Fangzhao Wu, Yongfeng Huang, and Xing Xie. 2023. Personalized News Recommendation: Methods and Challenges. ACM Transactions on Information Systems 41, 1 (2023), 1–50.

[7] Cortes C, Vapnik V. Support-Vector Networks J . Machine Learning,1995,20(3):273-297.

[8] SONG Zhongshan,NIU Yue,ZHENG Lu, et al. Multiscale CNN convolutional and global relation for Chinese text classification model[J]. Computer Engineering and Applications,2023 ,59(20):103-110.

[9] KIM Y.Convolutional neural networks for sentence classifification[J].arXiv:1408.5882,2014.

[10] CONNEAU A, SCHWENK H,BARRAULT L,et al.Verydeep convolutional networks for text classification [J].arXiv:1606.01781,2016.

[11] LE H T,CERISARA C,DENIS A.Do convolutional networks need to be deep for text classification[J]. arXiv: 1707.04108,2017.

[12] JOHNSON R,ZHANG T.Deep pyramid convolutional neu-ral networks for text categorization[C]//Proceedings of the55th Annual Meeting of the Association for Computa-tional Linguistics(Volume 1:Long Papers),2017:562-570.

[13] GUO J,YUE B,XU G,et al.An enhanced convolutionalneural network model for answer selection[C]//Proceed-ings of the 26th International Conference on World WideWeb Companion,2017:789-790.

[14] WANG H,HE J,ZHANG X,et al.A short text classification method based on n-gram and CNN[l].Chinese Jour-nal of Electronics,2020,29(2):248-254.

[15] Adams B., McKenzie G.Crowdsourcing the character of a place: character-levelconvolutional networks formultilingual geographic text classification[J]. Transactionsin Gis, 2018, 22(2): 394-408.

[16] Chen Zhuang, Qian Tieyun. Transfer capsule network for aspect level sentimentclassification[C]//Proceedings of the 57th Annual Meeting of the Association forComputational Linguistics. Florence: ACL press, 2019: 547-556.

[17] Guo Bao, Zhang Chunxia, Liu Junmin, et al. Improving text classification withweighted word embeddings via a multi-channel TextCNN model[,. Neurocomputing.2019,363:366-374.

[18] HOCHREITER S,SCHMIDHUBER J.Long short-term memory[J].Neural computation,1997,9(8):1735-1780.

[19] SCHUSTER M, PALIWAL K K. Bidirectional recurrent neural networks. IEEE Transactions on Signal Processing, 1997, 45(11):2673-2681.

[20] ZHOU P,QI Z,ZHENG S,et al.Text classification improvedby integrating bidirectional LSTM with two-dimensionalmax pooling[J].arXiv:1611.06639,2016.

[21] HUANG JI,LIN JO,HE YJ,et al.Chinese shorttext classification algorithm based on local semantics andcontext [1].Computer Engineering and Applications, 2021,57(6):94-100.

[22] XU XK,ZHOU Z YA multi-scale BILSTMCNNbased emotion classification model for Wechat tweetsand its application [J].Information Science, 2021 , 39(5):130-137.

[23] VELICKOVIC P,CUCURULL G,CASANOVA A,et al.Graph attention networks[Jl.arXiv:1710.10903,2017

[24] XU B B,CEN K Y,HUANG J,et al.A survey on graphconvolutional neural network[I].Chinese Journal of Computers,2020,43(5):755-780.

[25] Li R,Chen H,Feng F,et al.Dualgraph convolutiona networks foraspect-basedanalysis[C]sentimentonline:Proceedings of the Fifty-ninth Annual Meeting ofthe Association for Computational Linguistics and theEleventh International Joint Conference on NaturalLanguage Processing,2021.

[26] Shen w,Wu S,Yang Y,et al.Directed acyclic graph ne-tworkfor conversational emotion recognition[J/OL].(2021-5-27)[2022-9-10]

[27] Yan H,Gui L,Pergola G,et al.Position bias mitigation: Aknowledge-aware graph model for emotioncauseextraction [J/OL].(2021-6-7)[2022-9-10]

[28] YANG T,HU L,SHI C,et al.HGAT:heterogeneous graphattention networks for semi-supervised short text classification[1]. ACM Transactions on Information Systems, 2021,39(3):32.

[29] KENTON J D M-W C, TOUTANOVA L, K. BERT: pre - training of deep bidirectional transformers for language understanding IC].Proceedings of NAACL-HLT,2019: 4171-4186.

[30] DAI Z, YANG Z, YANG Y, et al. Transformer- XL: attentive language models beyond a fixed -length context [C]Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics,2019: 2978-2988.

[31] Cui Yiming, Che Wanxiang, Liu Ting, et al. Pre-training with whole word masking for chinese bert[J]. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 2022,29: 3504-3514.

[32] Kong Jun, Wang Jin, Zhang Xuejie. Hierarchical bert with an adaptive fine-tuningstrategy for document classification[J]. Knowledge-Based Systems, 2022,238:107872.

[33] Jindian Su, Shanshan Yu, Xiaobin Hong. A self-supervised pretraining method for Chinese spelling error correction[J]. Journal of South China University of Technology (Natural Science Edition), 2023, 51 (09): 90-98.

[34] Li Biang, Ma Hongchao,Zhou Qinglei. Sarcasm detection based on transfer learning[J]. Computer Application Research, 2021.38(12):3646-3650.

[35] Sun Xiaoyan,Qiao Yali. False comment recognition gate based on migration and semi-supervised symbiotic fusion. Journal of Nanjing University (Natural Science),2022.58(05):846-855.

[36] Soni S,Chouhan SS,Rathore SS.TextConvo Net: Aconvolutional neural network based architecture for textclassification [J/OL]. (2022-3-10)[2022-9-10]

[37] ZHANG X,ZHAO J,LECUN YCharacter-level convolutional networks for text classification[C]//Advances inNeural Information Processing Systems,2015,28:649-657

[38] WANG J,WANG Z,ZHANG D,et al.Combining knowl-edge with deep convolutional neural networks for shorttext classification[C]//26th International Joint Conferenceon Artificial Intelligence,2017:2915-2921.

[39] WU C Y,DIAO Q,QIU M,et al.Jointly modeling aspects,ratings and sentiments for movie recommendationC]// Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2014:193-202.

[40] SOCHER R,PERELYGIN A,WU J,et al.Recursive deepmodels for semantic compositionality over a sentimentreebank [C]//Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing,2013:1631-1642

[41] YAO L,MAO C,LUO YGraph convolutional networksfor text classification[C]//Proceedings of the 33rd AAAIConference on Artificial Intelligence,2019:7370-7377

[42] PHAN X H,NGUYEN L M,HORIGUCHI S.Learning toclassify short and sparse text & web with hidden topics firom large-scale data collections[C]//Proceedings of the17th International Conference on World Wide Web,2008 :91-100.

[43] LI X,ROTH D.Learning question classifiers[C]//19th International Conference on Computational Linguistics ,2002

[44] PANG B,LEE L.Seeing stars:exploiting class relationships for sentiment categorization with respect to rating scales[J].arXiv:cs/0506075 ,2005.