

Research Article

Applications of Deep Learning in News Text Classification

Menghan Zhang 

School of Communication, Soochow University, Suzhou 215123, China

Correspondence should be addressed to Menghan Zhang; mhzhang@suda.edu.cn

Received 21 June 2021; Revised 21 July 2021; Accepted 24 July 2021; Published 5 August 2021

Academic Editor: Mian Ahmad Jan

Copyright © 2021 Menghan Zhang. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The advancement in technology is taking place with an accelerating pace across the globe. With the increasing expansion and technological advancement, a vast volume of text data are generated everyday, in the form of social media platform, websites, company data, healthcare data, and news. Indeed, it is a difficult task to extract intriguing patterns from the text data, such as opinions, summaries, and facts, having varying length. Because of the problems of the length of text data and the difficulty of feature value extraction in news, this paper proposes a news text classification method based on the combination of deep learning (DL) algorithms. In order to classify the text data, the earlier approaches use a single word vector to express text information and only the information of the relationship between words were considered, but the relationship between words and categories was ignored which indeed is an important factor for the classification of news text. This paper follows the idea of a customized algorithm which is the combination of DL algorithms such as CNN, LSTM, and MLP and proposes a customized DCLSTM-MLP model for the classification of news text data. The proposed model is expressed in parallel with word vector and word dispersion. The relationship among words is represented by the word vector as an input of the CNN module, and the relationship between words and categories is represented by a discrete vector as an input of the MLP module in order to realize comprehensive learning of spatial feature information, time-series feature information, and relationship between words and categories of news text. To check the stability and performance of the proposed method, multiple experiments were performed. The experimental results showed that the proposed method solves the problems of text length, difficulty of feature extraction in the news text, and classification of news text in an effective way and attained better accuracy, recall rate, and comprehensive value as compared to the other models.

1. Introduction

News has evolved into an institution that disseminates the most up-to-date information to the public. News is disseminated through a variety of channels, including online media, newspapers, television, radio, and other forms of media. In general, news supplied through the media is divided into numerous areas, including health, economics, politics, and sports. Furthermore, at the moment, news is still manually classified into these categories, which means that when submitting a news item, the submitter would first know the entire text of the news to be published, which is then placed in the appropriate category. This is indeed a very tough task for the news unloaders who have a high amount of news items to deal with. As a result, an intelligent system that can automatically identify content according to existing

news category is required to assist news uploading personnel when posting such data. A machine learning strategy that can elucidate these issues is needed to solve these issues and is able to identify the news content types.

Text categorization is a fundamental task in the field of natural language processing (NLP) and is frequently utilized in the retrieval of information, unreliable analysis and identification, analysis of emotions, identification of spam emails, etc. [1–4]. Text mining is a term used to describe the method of extracting patterns or knowledge from unstructured texts. This can be thought of as a data mining or knowledge discovery extension from a structured (character and word) dataset. Text has evolved into one of the most fundamental forms of data storage; text mining is thought to have a greater market viability than data mining. From the recent research studies, it is observed that about 80–85% of

information of the companies and organizations are presented and stored in the textual format. Text mining, on the contrary, includes a significant number of more complex jobs (in comparison to data mining) because it deals with text data that is inherently unstructured and unclear. Information extraction, text analysis, information retrieval, and categorization are all part of the multidisciplinary discipline of text mining. Text mining technique is used as one of the method for the classification of news text [5]. Furthermore, one of the important functionality of text mining is that it identifies the most important patterns in a large amount of text dataset. It is also used for clustering, feature extraction, and information retrieval.

Researchers are interested in NLP applications such as text categorization, machine translation, speech recognition, document summarization, and question answering because of the vast amount of data available online. As a result, research in the NLP sector has progressed, and numerous state-of-the-art models in ML and AI have been developed [6]. Text classification has various applications such as e-commerce platform, blogs, content curator, directories, and news agency documents and text classification [7]. It is a supervised ML technique that uses text documents along with its labels for training the classification models. It can be done at several levels, including sentence, aspect, and document levels. In addition, extracting useful information/features from the sentences can be used to perform the sentiment analysis [8].

In the field of text classification by using deep learning (DL) approaches, researchers at home and abroad have made a lot of exploration. Yinghua et al. [9] proposed a model for the English text classification that extracts the local features by the CNN after the text input matrix is constructed by the word vector model, and the result index of text classification is improved. Tianlong et al. [10] proposed a dynamic CNN model in which they improved the pooling layer of the CNN and retained the semantic order after the pooling layer. In other studies, in order to highlight and extract the most important features from the textual data, the researchers have added an attention mechanism to the CNN model [11, 12]. The experimental results show that the most important features are preserved selectively which improved the text classification result significantly. Jinyuan et al. [13] have used a CNN model to analyze the emotion of twitter text, which solves the problem that the existing emotional classification methods are difficult to excavate the deep semantic features in the text. Ya and Shuang [14] constructed a word vector matrix through TF, IDF, and word2vec and highlighted the important word information in the text.

Yumin et al. [15] used a CNN model to solve the problem of web text classification that reduced the complexity of the model. In other studies conducted by different researchers [16–18], word2vec has been used to construct a word vector matrix for new product reviews and then the emotions are classified using textual data.

However, the abovementioned research of combination depth learning uses a single word vector to express text information and only the information of the relationship

between words is considered, but the relationship between words and categories is ignored which indeed is an important factor for the classification of news text. The following are some of the basic contributions of this paper:

- (i) This study follows the research idea of the combination of DL techniques based news text classification and selects CNN, LSTM, and MLP models to propose a custom MLP model of news text classification based on double input combined depth learning.
- (ii) The proposed model is expressed in parallel with word vector and word dispersion. The relationship among words is represented by word vector as input of CNN module, and the relationship among words and categories is represented by a discrete vector as an input of the MLP module in order to realize comprehensive learning of spatial feature information, time-series feature information, and relationship between words and categories of news text.
- (iii) Multiple experiments were conducted to check the stability and performance of the proposed method.
- (iv) The experimental results show that the performance of the proposed method is way better than the other approaches in terms of prediction accuracy and other performance measures.

The rest of the paper is organized as follows. Section 2 represents the related work, while Section 3 illustrates the proposed model for news text classification. Section 4 demonstrates the experimental results and analysis, and finally, we conclude our paper in Section 5.

2. Related Work

In recent years, deep learning (DL) has become a research trend and hot spot for the research communities across the globe. DL has numerous applications in various fields such as healthcare, education, industries, agriculture, and text processing. Some scholars have used DL technology to extract text feature information. For example, the problem of data dimension explosion and high sparsity is not easy to appear when extracting text features by traditional ML algorithms. Longfeng [19] first proposed the word2vec model, which provides technical support for word vector conversion. At the same time, some scholars proposed to use neural network as a classification model, for example, Alswaidan and Menai [20] have used the word2vec model to generate word vector and then used the CNN model to learn feature information. Based on the word vector, long short-term memory (LSTM) is used to analyze public opinion emotion of emergency network in the process of studying the pre-Qin classics. First, they constructed the classification system, then expressed the text characteristics with TF-IDF, and input it into the Bi-LSTM (bi-directional LSTM) model. The experimental results show that the effect of the DL method was better than that of the ML methods. Compared to the traditional ML methods, the accuracy of the DL methods has been improved significantly.

There are several types of neural networks, and the nature and structure of each network model learning information is different and depends on the nature of the problem. For example, the CNN model is more inclined to extract local spatial feature information because of its unique local connection structure. Recurrent neural network (RNN) model is more inclined to extract the feature information of time series due to its sequence structure (the output of a certain time is part of the input of the next time). Because of the defects of the ability to extract information from a single DL model, some scholars began to study the feature information extraction of the combined DL models. For example, Mu et al. [21] introduced the attention mechanism to study the news text classification based on the combination of the DL models. Cao et al. [22] input the text features into the RCNN (region CNN) neural network model composed of RNN and CNN and applied them to text classification, and the classification performance is improved obviously. Zhao et al. [23] proposed a DL model composed of CNN and LSTM based on attention (CLA) by repeated series of convolutional layers and circulation layer. Firstly, the word coding was implemented in series; secondly, they implemented it to realize the sentence coding, and finally, the implicit emotion analysis task was realized at the last layer using the softmax function, based on attention mechanism. In the study of text classification, Ke and Chen [24], firstly, quantify the text words and then extract the text feature information by neural network and give weight to each feature by TF-IDF algorithm, and finally use the Bayesian classifier to achieve the classification task. Yu [25], discussed the principle, advantages, and disadvantages of CNN and RNN and combined a model method for searching the automatic abstract technology.

In recent years, DL-based text categorization research has made significant progress [26]. DL models, in general, excel in learning the high-level vector representations of words, phrases, and paragraphs from raw text data and have a significant impact on classification outcomes. With the advancement of DL technology, an increasing number of researchers are applying it to the field of NLP. Kim et al. [27] have used CNN to classify text and attained promising classification outcomes. However, because of the differences in expression and word structure between Chinese and English, determining the type of a Chinese text may require multiple words or characters. Keeping the variation of characters and words of both of the languages in consideration, it is mandatory to use segmentation techniques for the extraction of Chinese text. For Chinese text classification, most of the models use the segmentation techniques by segmenting the word first and then forward the segmented features the classification models to classify it [28, 29]. For the categorization of Chinese text, most of the models have used word embedding techniques [30–32]. Most contemporary Chinese text classification models, on the contrary, treat text characteristics as the fundamental unit of text representation, ignoring the useful representation of character features.

However, the abovementioned research, using DL techniques, focuses only on a single word vector to express text information. Only the information of the relationship

between words is considered, but the relationship between words and categories is ignored. In order to overcome the issues mentioned above, this paper follows the research ideas of combination of DL techniques for news text classification and selects CNN, LSTM, and MLP (multilayer perception) models to propose a custom MLP model for news text classification based on double input combined DL technique. It is expressed in parallel with word vector and word dispersion, and the relationship among words is represented by word vector as an input of the CNN module. Furthermore, the relationship between words and categories is represented by a discrete vector as an input of the MLP module, to realize comprehensive learning of local spatial feature information, time-series feature information, and relationship between words and categories of news text.

3. Proposed Model for News Text Classification

This section of the paper illustrates the proposed model used to classify the news textual data. The proposed model works in different steps. When the news events appear on the Internet, the process of news text classification based on the combination of DL techniques is given as follows:

Step 1: input the text into the CNN model to predict whether the text belongs to the news event or not. If the answer/prediction is not, then it does not need to be monitored, otherwise, it enters Step 2.

Step 2 (text preprocessing and feature expression): word segmentation and filtering stop words, using the word2vec model in order to get the text word vector, and calculate the text dispersion vector. The vector input based on discreteness reflects the association information between words and categories. The greater the value is, the greater the contribution to classification is. The input based on a word vector reflects the semantic relationship between words. Using two kinds of feature expression can make the model learn the deep-seated features of the text as much as possible.

Step 3: input the word vector to MCNN (multiple CNN) modules and get the text spatial feature information through two convolution and pooling operations and then input the spatial feature to the LSTM module to learn the temporal feature information. The discrete vector is then given as an input to the MLP module, and the hidden layer neuron learns the relationship between words and categories. The output values of the MLP module and the LSTM module are spliced and fused and given as input to the softmax activation function for feature scaling and output news prediction categories.

3.1. Feature Representation among Words and Categories Based on Discreteness. When we compare the ordinary text with the news text, the news text contains relatively few keywords but has a strong relevance with the theme of the news. If “earthquake” appears in the news text, the text is likely to be earthquake event news. Therefore, this paper uses

key feature words with strong correlation to express news text information and to distinguish ordinary text from event news. In DL techniques, the word vector method uses the spatial distance to represent the similarity degree of word meaning, and the one hot method uses 0 and 1 to represent whether the feature words appear or not, ignoring the proportion weight of words in the text and the association information between words and categories. The following are some of the steps that demonstrate the prescribed scenario:

Step 1: count the number of texts with specific words in a certain category to form the category frequency vector of the words. Let V_{Im} denote the number of texts with the word I in the m^{th} category, and the frequency vector V_I of the word I in the sample set with n categories is

$$V_I = (v_{I1}, v_{I2}, \dots, v_{In}). \quad (1)$$

Step 2: considering the deviation of the class frequency vector caused by the different numbers of texts in each class, the occurrence probability of specific words in each class is calculated. Let Q_i be the total number of I -type texts, and P_I be the category probability vector, which is given as follows:

$$P_I = \left(\frac{V_{I1}}{q_1}, \frac{V_{I2}}{q_2}, \dots, \frac{V_{In}}{q_n} \right). \quad (2)$$

Step 3: calculate the variance of P_I as the dispersion of word I . The greater the variance, the greater the dispersion, indicating that the feature information contained in the classification process is more obvious. Let \bar{p}_I be the probability mean of the word I and σ_I^2 be the probability variance of the word I , and the formula of dispersion is given as follows:

$$\begin{aligned} \bar{p}_I &= \frac{\sum_{i=1}^n v_{Ii}/q_i}{n}, \\ \sigma_I^2 &= \sum_{i=1}^n \left(\frac{v_{Ii}}{q_i} - \bar{p}_I \right)^2. \end{aligned} \quad (3)$$

Step 4: the text discreteness vector is composed of the discreteness of each word, while the length is the total number of words. When a word appears, the corresponding word sequence is the variance value of the word. Otherwise, it is 0. Let Z be the total number of words in the sample set and d be the text dispersion vector; then, it can be expressed as follows:

$$D = (\sigma_1^2, \sigma_2^2, \dots, \sigma_z^2), \quad (4)$$

3.2. Proposed DCLSTM-MLP Model Structure. This section illustrates the proposed DCLSTM-MLP model and its structure. The structure of the custom-map model is shown in Figure 1.

The proposed model is mainly composed of three parts: MLP, MCNN, and LSTM. Among them, MLP takes the feature vector between words and categories based on the

degree of dispersion as the input, and the hidden layer neurons learn the relationship between words and categories and output the feature information vector between words and categories. MCNN takes the word vector containing interword feature information as the input, which has five types of the convolution kernel. After the convolution pooling operation of different types of convolution kernels, it repeatedly extracts the local features of the word vector and outputs the text space feature information vector of multiple types of the convolution kernel. The feature vector extracted by one type of convolution kernel is a time segment, and the feature information extracted by multiple types of convolution kernels is composed of multiple time segments, which are given as input to the LSTM to learn the time-series feature information of multiple time segments and finally output the comprehensive spatial feature vector. The MLP and LSTM output feature vectors are spliced and fused into a new feature information vector and then given as an input to the full connection layer for re-learning. The extracted feature information after re-learning is input to the softmax layer for feature scaling, and finally, the text category is obtained.

4. Experimental Results and Analysis

This section of the paper represents the experimental results carried out via different experiments and the analysis of those results. All the experiments were performed on a laptop computer system having the specification of Intel Core-i7, 9th generation and a processor of 2.7 GHz, 12 GB of RAM, and the operating system installed on the system was Microsoft Windows 10. Anaconda, Jupyter Notebook is used as an IDE for carrying out the simulations. *Python* has been used as a language for the implementation and for generating the simulation results. The list of python packages used in this study was Pandas, Numpy, Matplotlib, Keras, Tensorflow, seaborn, and Sklearn.

4.1. Event News Text Collection. This section represents the collection of event news textual data. We divide the news events into four categories: (I) public health events, (II) social security events, (III) accidents and disasters, and (IV) natural disasters. In this study, we search 10 kinds of natural disaster events including storm disaster, rainstorm disaster, Blizzard disaster, earthquake disaster, tsunami disaster, flood disaster, debris flow disaster, forest fire disaster, sandstorm disaster, and landslide disaster and 5477 news event texts, in the form of keywords on Baidu news, using the crawler technology. We also search the open-source regular news event text set (such as economic, art, political, and other news texts), 2815 news texts, using the same technology, forming a total of 8292 news texts, which were merged.

4.2. Analysis of the Classification Model of Common Events and News Events. The total sample set contains 5477 event news texts and 2815 ordinary news texts by forming a total of 8292 sample set. From the samples' set, 6699 texts are

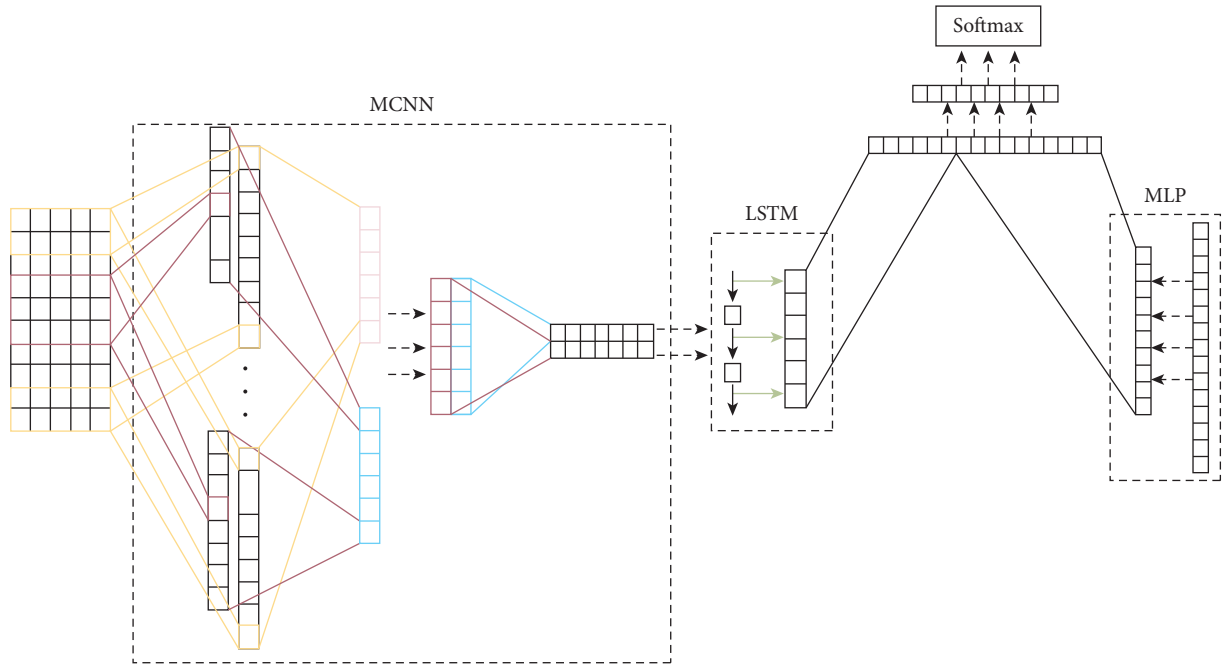


FIGURE 1: DCLSTM-MLP model structure.

randomly selected as the training set and 1593 texts as the test set, and the training set is input into the proposed CNN model. The main parameters of the CNN model include the first mock exam length, the word vector dimension, the number of convolution layers, and the number of neurons in the full connection layer. Each time the single model parameters are changed, and the accuracy of and recall the model and average value of the composite value are obtained by repeating the simulations 3 times. The parameters with the maximum average value are the best, and the best parameters and optimal results are shown in Table 1. Using the parameter setting mentioned in Table 1, the classification model can effectively identify ordinary text and news events and lay the foundation for the next step of news event text classification.

4.3. Analysis of News Text Based on Multiclassification Models

4.3.1. Superiority Test of Dispersion Vector. To test the feature information extraction ability of the proposed test dispersion vector expression method, the news event text is taken as the sample set and compared with the vector space model (VSM), chi-square test, and TF-IDF method. The results attained via all the mentioned techniques are shown in Figure 2. It can be seen from Figure 2 that the proposed method has better information expression ability than the VSM, chi-square test, and TF-IDF method, and the extracted feature information are more abundant and valuable for the classification model.

Based on the DL approach, the parameters of the first mock exam text classification model determine the performance of the model. To compare the DCLSTM-MLP combination model with MLP, Text-CNN, Text-LSTM,

CLSTM, and CNN-MLP models proposed in this paper, we get the accuracy of each single model parameter by changing the single model parameters 3 times on the same training set of the news event text. For the average value of recall and comprehensive value, the parameter with the maximum comprehensive value is the best, as shown in Tables 1–6 .

Based on setting the best state of each model, taking the same test set of news event text as the object, the rate of accuracy, recall, and comprehensive value of each model is obtained. Table 7 shows the results of accuracy, recall, and comprehensive value of all the utilized models.

Figure 3 shows a comparative study of the performance (accuracy, recall, and comprehensive value) of all the utilized models in this study.

From Table 7 and Figure 3, we can conclude the following. In general, the more complex the structure of the text classification model is, the more is the number of neurons, the more the network level is, the stronger the learning ability is, and the stronger the comprehensive performance of the model is. From the perspective of accuracy, it can be observe that $MLP < \text{Text-LSTM} < \text{Text-CNN} < \text{CLSTM} < \text{CNN-MLP} < \text{DCLSTM-MLP}$. From the perspective of recall, $\text{DCLSTM-MLP} > \text{CLSTM} > \text{Text-LSTM} > \text{CNN-MLP} > \text{Text-CNN} > \text{MLP}$. So, in all cases, the proposed DCLSTM-MLP is better than the other models.

Compared with other neural networks, the MLP model has the simplest structure and limited ability to learn features. The structure of the CLSTM model is more complex than that of CNN-MLP, but the former uses the single input method with the word vector as the input value, while the latter uses the double input method with the word vector and the dispersion vector as the input value. The latter has more sufficient input information and more comprehensive model learning content, so the accuracy of the latter is slightly higher. The

TABLE 1: MLP model parameters and optimal values.

Parameter	Dispersion vector dimension	Number of neurons in the first layer	Number of neurons in the second layer
Optimum value	400	1024	256

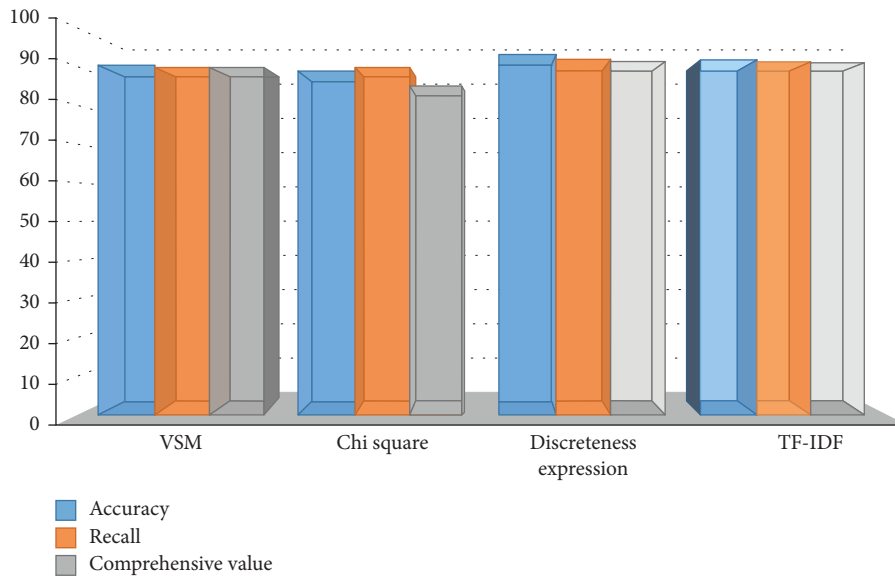


FIGURE 2: Performance comparison of various feature expression methods.

TABLE 2: CNN model parameters and optimal values.

Parameter	Sentence length	Word vector dimension	Number of convolution kernels	The number of neurons in the whole connective layer
Optimum value	80	240	512	256

TABLE 3: Parameters and optimal values of the text-LSTM model.

Parameter	Sentence length	Word vector dimension	Number of convolution kernels	The number of neurons in the whole connective layer
Optimum value	80	240	120	512

TABLE 4: CNN-MLP model parameters and best values.

Parameter	Sentence length	Word vector dimension	Number of convolution kernels	Number of neurons in the first layer	Dispersion vector dimension	Number of neurons in the second layer	The number of neurons in the whole connective layer
Optimum value	80	240	512	1024	400	256	256

TABLE 5: Parameters and optimal values of the CLSTM model.

Parameter	Sentence length	Word vector dimension	Number of convolution kernels	The number of neurons in the hidden layer	Number of neurons in circulatory layer
Optimum value	80	240	512	256	320

TABLE 6: DCLSTM-MLP model parameters and their best values.

Parameter	Sentence length	Word vector dimension	The number of neurons in the second hidden layer	The number of neurons in the whole connective layer	Number of LSTM neurons	Number of neurons in the first hidden layer	The number of convolution kernels in the first layer	The number of convolution kernels in the second layer
Optimum value	400	260	256	512	320	1024	512	256

TABLE 7: Accuracy, recall, and comprehensive value results of all the utilized models.

Method	Accuracy	Recall	Comprehensive value
MLP	88.76	88.25	88.05
Text-CNN	92.46	91.35	91.2
Text-LSTM	92.35	92.2	91.2
CNN-MLP	93.68	91.53	91.52
CLSTM	93	93.05	93.1
DCLSTM-MLP	94.82	94.97	94.83

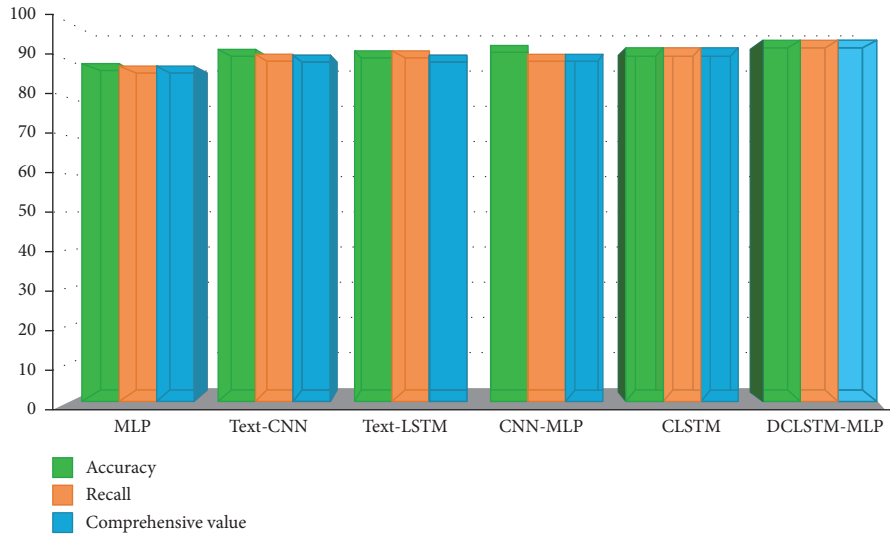


FIGURE 3: Comparison of the models' performance.

DCLSTM-MLP model has the most complex structure. It can learn not only the sequence information and spatial information of text but also the relationship information between feature words and categories. The accuracy of the DCLSTM-MLP model is significantly higher than the other models. The comprehensive value is a parameter to comprehensively measure the accuracy and recall rate, which can objectively reflect the trend of accuracy and recall rate, so the more complex the model is, the higher the comprehensive value is.

The accuracy rate of the DCLSTM-MLP model reached 94.82%, which is significantly higher than that of the other models (MLP: 88.76%, Text-CNN: 92.46%, Text-LSTM: 92.35%, CNN-MLP: 93.68%, and CLSTM: 93.0%). Also, the recall rate of the DCLSTM-MLP is 94.97%, and comprehensive value is 94.83% and is significantly better than that of the other models. Overall, the comprehensive value of the DCLSTM-MLP model is 94.83% which is higher than the other models in the percentage (MLP:

6.06%, text-CNN: 2.36%, text-LSTM: 2.47%, CNN-MLP: 1.14%, and CLSTM: 1.79%), which indicates that the combined model can improve the classification performance significantly.

5. Conclusion

The difference between the key features of news events and ordinary texts, as well as the obvious distinguishing features of all kinds of event news text keywords, is an important and challenging task. This paper solves the abovementioned problems and has the following three innovations. Firstly, a two-level classification model is designed. The first-level model identifies news events, and the second-level model implements news event classification. Secondly, different from the existing research, using word vector to express text feature information, this paper proposes a discrete vector to express text feature information, considering the

contribution of each word in classification, and obtains the contribution of each word in classification by calculating the probability variance of each word. Finally, in the proposed model, the word vector and the dispersion vector are used to express the text features. The dispersion vector is used to represent the relationship between words and categories, while the word vector is used to represent the semantic information between words. Through the experimental comparison and analysis, in the two-level classification model proposed in this paper, the recognition rate of the first-level model is 99.5%, and the accuracy rate of the second-level model is 94.82%, which shows that the model has significant news event recognition and classification ability. The limitation of this paper is that the public stop word list is used in the text preprocessing, and the special stop word list corresponding to news is not constructed, resulting in some feature information being filtered out, and the special stop word list for news events can be established later. At the same time, the number of sample sets is small, and the number of various types of news is uneven, which leads to the limited effectiveness of the model. In the later stage, more sample sets need to be used for experiments. The limitation of this work is the future work of this paper.

Data Availability

The data used to support the findings of this study are included within the article.

Conflicts of Interest

The authors declare that they have no conflicts of interest regarding the publication of this paper.

References

- [1] W. B. Croft, D. Metzler, and T. Strohman, *Search Engines: Information Retrieval in Practice*, Vol. 283, Addison-Wesley Reading, Boston, MA, USA, 2010.
- [2] W. Zhang, X. Tang, and T. Yoshida, "TESC: an approach to TExt classification using semi-supervised clustering," *Knowledge-Based Systems*, vol. 75, pp. 152–160, 2015.
- [3] W. Zhang, Y. Du, T. Yoshida, and Q. Wang, "DRI-RCNN: an approach to deceptive review identification using recurrent convolutional neural network," *Information Processing & Management*, vol. 54, no. 4, pp. 576–592, 2018.
- [4] H. Chen, M. Sun, C. Tu et al., "Neural sentiment classification with user and product attention," in *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pp. 1650–1659, Austin, TX, USA, November 2016.
- [5] A. Abbe, C. Grouin, P. Zweigenbaum, and B. Falissard, "Text mining applications in psychiatry: a systematic literature review," *International Journal of Methods in Psychiatric Research*, vol. 25, no. 2, pp. 86–100, 2016.
- [6] M. Abdul-Mageed and M. T. Diab, "AWATIF: a multi-genre corpus for modern standard Arabic subjectivity and sentiment analysis," in *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC'12)*, pp. 3907–3914, Istanbul, Turkey, May 2012.
- [7] S. Wang, S. Mazumder, B. Liu, M. Zhou, and Y. Chang, "Target-sensitive memory networks for aspect sentiment classification," in *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 957–967, Melbourne, Australia, July 2018.
- [8] L. Zhang, S. Wang, and B. Liu, "Deep learning for sentiment analysis: a survey," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 8, p. e1253, 2018.
- [9] S. Yinghua, L. V. Long, and L. Dan, "Research on emergency news recognition and classification based on combinatorial deep learning model," *Acta Sinica Sinica*, vol. 40, no. 2, pp. 145–151, 2021.
- [10] G. Tianlong, F. Xuan, L. Long, B. Xuguang, and L. Yunhui, "Ethical behavior discrimination method based on social news data set," *Computer Research and Development*, vol. 58, no. 2, pp. 253–263, 2021.
- [11] D. Dandan, *Research and Application of Feature Dimension Reduction in Text Classification*, Nanjing University of Posts and telecommunications, Nanjing, China, 2020.
- [12] L. gaojun and W. Xiaobin, "Text classification of marketing news based on CNN + LST attention," *Computer Technology and Development*, vol. 30, no. 11, pp. 59–63, 2020.
- [13] T. Jinyuan, D. Yufeng, Q. Ruihua, and L. Hongfei, "Automatic summarization generation of Chinese news text based on the bet-pgn model," *Computer Applications*, vol. 41, no. 1, pp. 127–132, 2021.
- [14] G. Ya and F. Shuang, "Sentiment analysis algorithm of news text combined with attention mechanism," *New Industrialization*, vol. 10, no. 7, pp. 15–18, 2020.
- [15] Z. Yumin, *Design and Implementation of Personalized News Recommendation System Based on Knowledge Map*, Beijing Jiaotong University, Beijing, China, 2020.
- [16] P. Yao, W. Mo, and W. Jian, "Text clustering and application prospect of agricultural trade friction news based on improved text representation," *Agricultural Prospect*, vol. 16, no. 6, pp. 80–88, 2020.
- [17] N. Shan, Y. Xin, X. Guangyi, Z. Feng, and Z. Lei, "Chinese news text abstract generation with keywords," *Computer Engineering and Science*, vol. 42, no. 12, pp. 2265–2272, 2020.
- [18] S. Yuxiao, *Research and Implementation of Automatic Text Summarization Method for Forestry News*, Beijing Forestry University, Beijing, China, 2020.
- [19] W. Longfeng, *Research on News Text Classification Based on Deep Learning*, Anhui University of Technology, Hefei, China, 2020.
- [20] N. Alswaidan and M. E. B. Menai, "A survey of state-of-the-art approaches for emotion recognition in text," *Knowledge and Information Systems*, vol. 62, 2020.
- [21] X. Mu, F. Zhu, Y. Liu, E.-P. Lim, and Z.-H. Zhou, "Social stream classification with emerging new labels," *Advances in Knowledge Discovery and Data Mining*, Springer, Berlin, Germany, pp. 16–28, 2018.
- [22] Z. Cao, W. Li, S. Li, and F. Wei, "Improving multi-document summarization via text classification," in *Proceedings of the AAAI Conference on Artificial Intelligence*, San Francisco, CA, USA, February 2017.
- [23] R. Zhao, X. Xiong, X. Zu, S. Ju, Z. Li, and B. Li, "A hierarchical attention recommender system based on cross-domain social networks," *Complexity*, vol. 2020, Article ID 9071624, 13 pages, 2020.
- [24] Y. Ke and L. Chen, "Research on text classification and data analysis of complaints by shanghai gas company," in *Proceedings of the 2021 IEEE 5th Advanced Information*

- Technology, Electronic and Automation Control Conference (IAEAC)*, pp. 1626–1630, Chongqing, China, March 2021.
- [25] M. Yu, *Design and Implementation of Text Summarization System for Economic News*, Beijing University of Posts and Telecommunications, Beijing, China, 2020.
- [26] Y. Le Cun, Y. Bengio, and G. Hinton, “Deep learning,” *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.
- [27] H. Kim, K. Anderson, S. Lee, and J. Hildreth, “Generating construction schedules through automatic data extraction using open BIM (building information modeling) technology,” *Automation in Construction*, vol. 35, pp. 285–295, 2013.
- [28] Y. Kim, “Convolutional neural networks for sentence classification,” in *Proceedings of the Empirical Methods in Natural Language Processing*, pp. 1746–1751, Doha, Qatar, October 2014.
- [29] L. Siwei, L. Kang, H. Shizhu et al., “How to generate a good word embedding?” *IEEE Intelligent Systems*, vol. 31, p. 1, 2016.
- [30] W. Huang and J. Wang, “Character-level convolutional network for text classification applied to Chinese corpus,” 2016, <https://arxiv.org/abs/1611.04358>.
- [31] Y. Li, X. Wang, and P. Xu, “Chinese text classification model based on deep learning,” *Future Internet*, vol. 10, no. 11, p. 113, 2018.
- [32] S. Liu, P.-T. Bremer, J. J. Thiagarajan et al., “Visual exploration of semantic relationships in neural word embeddings,” *IEEE Transactions on Visualization and Computer Graphics*, vol. 24, pp. 553–562, 2017.