Research Article

Research on News Text Classification Based on Deep Learning Convolutional Neural Network

Yunlong Zhu

Commission for Discipline Inspection, Zhengzhou Technical College, Zhengzhou 450121, China

Correspondence should be addressed to Yunlong Zhu; zhuyunlong123@zcmu.edu.cn

Received 2 September 2021; Revised 31 October 2021; Accepted 9 November 2021; Published 8 December 2021

Academic Editor: Yinghui Ye

Copyright © 2021 Yunlong Zhu. 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.

Aiming at the problems of low classification accuracy and low efficiency of existing news text classification methods, a new method of news text classification based on deep learning convolutional neural network is proposed. Determine the weight of the news text data through the VSM (Viable System Model) vector space model, calculate the information gain of mutual information, and determine the characteristics of the news text data; on this basis, use the hash algorithm to encode the news text data to calculate any news. The spacing between the text data realizes the feature preprocessing of the news text data; this article analyzes the basic structure of the deep learning convolutional neural network, uses the convolutional layer in the convolutional neural network to determine the change value of the convolution kernel, trains the news text data, builds a news text classifier of deep learning convolutional neural network, and completes news text classification. The experimental results show that the deep learning convolutional neural network can improve the accuracy and speed of news text classification, which is feasible.

1. Introduction

With the rapid development of Internet technology, we are in the era of information explosion. While enjoying the convenience brought by rich online information, we are also facing the severe challenge of how to quickly and effectively extract data information from massive information. Therefore, people pay more and more attention to the research and analysis in the direction of information retrieval technology and data mining technology, which makes the research related to natural language processing develop rapidly. Among them, the main task of text classification technology, which plays a key role in massive text data processing, is to solve the problem of chaotic text information to a certain extent. The basic principle is to extract relevant text features from the original text content and finally judge the category label, which has attracted more and more attention from scholars. And then, there are many applications and research [1]. However, with the rise of big data, massive online Internet text information has a series of new features, such as more complex formats, more cumbersome types, faster update speed, and more difficult labeling.

Especially after mobile phone users are popularized on a large scale, microblog, headline news, and other social life are enriched, and various short texts are also growing rapidly. All these changes will bring new challenges to text classification [2]. With the advent of the era of artificial intelligence, machine learning is one of the important disciplines in this field. In these decades of development, the research of machine learning-related algorithms has been quite mature, and a series of breakthroughs have been made in practical application. 2006 was the first year of deep learning. After its concept was proposed again, it quickly became the core research field of scholars all over the world. Up to now, as the core research object in the field of machine learning, deep learning has attracted extensive attention from contemporary Internet big data and artificial intelligence [3]. Since Google, Microsoft, IBM, Baidu, and other large Internet technology companies began to focus on the research and development of deep learning technology, it has made great breakthroughs in the fields of image, speech recognition, and natural language processing. Deep learning establishes a multilayer neural network structure by simulating the hierarchical structure of human brain, extracts the
distributed features of input data layer by layer from the bottom to the top, and finally establishes a good mapping function to describe the abstract relationship from the bottom signal to the high semantics [4]. Obviously, as an algorithm rising in the whole big data environment, deep learning will become one of the hot research directions in the future. Therefore, using artificial intelligence algorithm to classify news text has become a hot issue in this field.

Literature [5] proposed a short text classification method based on keyword similarity. Firstly, the word 2vec word vector model is obtained through a large number of corpus training. Then, the keywords of each type of text are obtained through textrank, and the de duplication operation is carried out in the keyword set as the feature set. For any feature, the similarity between each word in the short text and the feature is calculated through the word vector model. The maximum similarity is selected as the weight of the feature. Finally, k-nearest neighbor (KNN) and support vector machine (SVM) are selected as the classifier training algorithm. Based on the Chinese news title dataset, the classification effect is improved by about 6% on average compared with the traditional short text classification method, which verifies the effectiveness of the method. However, this method does not preprocess the text before text classification, resulting in more similar information, which affects the classification results. Reference [6] proposes an ACO-WNB classification algorithm based on improved information gain. Firstly, according to the word frequency distribution of feature words in the dataset, an adjustment factor is added to enhance/suppress the contribution/interference of feature words, select features with strong discrimination to form feature subsets, and improve the accuracy of Ig processing unbalanced datasets. Then, the ant colony optimization algorithm (ACO) is combined with the weighted naive Bayesian model, and ACO is used to iterate and globally optimize the weights to generate ACO-WNB classifier to improve the classification efficiency of text data. The improved algorithms are compared and analyzed with typical news datasets. The experiments show that Ig can effectively remove redundant high-frequency features and has better feature selection ability for unbalanced datasets. ACO-WNB classifier has higher accuracy and better classification efficiency for actual text data. The classification process of this method is complex and has some limitations. Literature [7] proposed CRNN text classification algorithm based on attention mechanism. Taking the pretrained word vector as the input, the convolutional neural network (CNN) is used to extract the features of the text vector; Bi-gating loop unit (BI Gru) is used to capture the word order of text, extract the context dependency of the text, and identify the importance of different features combined with the attention mechanism. The highway network is used for feature optimization. The model is tested on three English corpora: 20 newsgroups, sst-1, and sst-2. The experimental results show that the model effectively improves the accuracy of classification tasks. However, this method has the problem of high noise in classification.

In view of the shortcomings of the above methods, this paper proposes a research on news text classification based on deep learning convolutional neural network. The weight of news text data is determined by VSM vector space model, and the information gain of mutual information is calculated to determine the characteristics of news text data. On this basis, the hash algorithm is used to encode the news text data, calculate the spacing between any news text data, and realize the feature preprocessing of news text data. This paper analyzes the basic structure of deep learning convolutional neural network, determines the change value of convolution kernel with the help of convolution layer in convolution neural network, trains news text data, constructs news text classifier of deep learning convolution neural network, and completes news text classification. The technical route of this paper is as follows:

**Step 1**: determine the weight of news text data through VSM vector space model, calculate the information gain of mutual information, and determine the characteristics of news text data.

**Step 2**: the hash algorithm is used to encode the news text data, calculate the spacing between any news text data, and realize the feature preprocessing of news text data.

**Step 3**: analyze the basic structure of deep learning convolution neural network, determine the change value of convolution kernel with the help of convolution layer in convolution neural network, train news text data, construct news text classifier of deep learning convolution neural network, and complete news text classification.

**Step 4**: experimental analysis

**Step 5**: conclusion

### 2. Classification of News Text Based on Deep Learning Convolutional Neural Networks

#### 2.1. News Text Data Feature Extraction

Due to the complex format of news text data directly acquired by the Internet, computers cannot directly understand the text content, so the representation of news text characteristics is to transform the unstructured or structured news text of the Internet into structured text intelligible by the computer. In this paper, we first characterize its feature [8] with the help of the VSM vector space model. The model is based on the number of feature words appear statistics, mainly including three steps: first, calculate the word frequency, then calculate the inverse document frequency, and finally calculate the TF-IDF. The model is based on the feature representation of the word vector, if the reference source is not found in error. In presenting the weight size of each feature item in each text set, each text can be measured by the reference source [9]. If a feature word corresponds to a word vector, all feature words in all text sets correspond to the corresponding dimension of the space. The weights are calculated here using the TF-IDF method based on word frequency, the number of single appearances, giving a large weight [10] to persuasive features in each document. The TF-IDF weights are calculated as follows:

\[
w_{ij} = tf_{ij} \times idf_{ij} = tf_{ij} \times \log \left( \frac{N}{nj} \right).
\]  

(1)
Among them, $w_{ij}$ represents feature terms, $tf_{ij}$ represents the number of news text appearances, $idf_{ij}$ represents the frequency of news text occurrence, $\log (N/n_{ij})$ represents the derivative of text occurrence, and $N$ represents the total number of news text.

After the feature weights of the calculated news text data, the entropy changes of the news text data. Information gain is an entropy-based method. First, calculate the change of information entropy when the feature item appears, that is, the information gain, and then select the information gain according to the size value to measure the importance of the information of the final classification; feature importance is proportional to the amount of information, that is, the more the feature carries information represents the greater the word feature importance [11]. The following formula is the calculation method of the information gain:

$$g(t) = - \sum_{i=1}^{m} P(C_i) \log P(C_i) + P(t) \sum_{i=1}^{m} (C_i),$$

where $P(C_i)$ represents the proportion or probability of $C$ category documents in all sets of documents, $P(t)$ represents the probability of having feature items in a document, and $m$ is the number of document categories.

Mutual information is a variable that reflects the correlation between two variables. It works as follows: mutual information size represents the degree of correlation between characteristics and category, and the larger the association, the closer the former. Mutual information metric method is as follows:

$$M(t, C) = \log \frac{P(t \pm c)}{P(t) \times p(c)}.$$  \hspace{1cm} (3)

Among them, $t$ represents the feature term and $C$ represents the category.

According to the calculation of the mutual information of the above news text, the extraction of the characteristic information of the news text is completed, i.e.,

$$\varphi_j = \arg \min \sum_{i=1}^{n} g_i + v|Y|^2.$$  \hspace{1cm} (4)

In the formula, $y$ represents the proportional coefficient of news text feature data characteristics, and the key value of news text features [16].

2.2. News Text Data Feature Preprocessing. To realize the effective classification of news text, this paper introduces the deep hash algorithm to effectively preprocess the above news text features. The deep hash algorithm can effectively handle the characteristics of news text. When the characteristics of news text differ, the algorithm of [12] by removing Hamming distance has the advantage of simple operation and high outlet efficiency [17].

The basic idea of hash function is that in the original news text feature data, the feature points are relatively close, and the impact of these two points occurs more frequently, when setting the set of news text features to be processed as follows:

$$H = \{h_i\}_{i=1}^{n}.$$  \hspace{1cm} (5)

Set the distance metric function for two points in the news text feature to the following:

$$g_i = h_i \sum \gamma_p.$$  \hspace{1cm} (6)

According to the distance between the two points obtained above, since the hash algorithm has certain unidirectional, it is irreversible, i.e.,

$$\sigma \rightarrow R(\sigma).$$  \hspace{1cm} (7)

In the formula, $R(\sigma)$ represents the output after the hash and $\sigma$ represents the original news text.

The deep hash algorithm is considered as a density function [13] when dealing with noise in features in news text, and each density function corresponds to features of multiple news text, i.e.,

$$\text{size}(R(\sigma)) < \text{size}(\sigma).$$  \hspace{1cm} (8)

According to the features of the corresponding news text, [14] is converted to obtain the following:

$$E = \{R_i(\sigma)\} > 0.$$  \hspace{1cm} (9)

Finally, the preprocessing of the news text feature data based on the deep hashing algorithm is implemented, i.e.,

$$V_I = \frac{1}{2} S_t Y(h).$$  \hspace{1cm} (10)

In the formula, $V_I$ represents the preprocessing results of news text feature data and $Y(h)$ represents the weight value of preprocessing [18].

2.3. Classification of News Text Based on Deep Learning Convolutional Neural Networks. Deep learning convolutional neural network is an AI learning algorithm based on the neural network topology and optimizes the convolutional layer in neural networks into a convolutional kernel and simultaneously completes the propagation [15] in different directions in signal processing. The technical core of the method is in the transformation of the convolutional wave, through which detailed information of the data can be obtained in any case. Its expansion structure is shown in Figure 1.

In the study of this paper, the change of convolutional kernels in the convolutional layer completes the classification of news text. The convolutional layer structure is shown in Figure 2.

In Figure 2, the $a_1, a_2 \ldots a_n$ represents the feature data that the structure needs to be entered, and $n$ is the number of feature data; $g_1, g_1 \ldots, g_n$ represents the number of convolutional kernels, $b$ and $c$ represent different weight values.
of the structure, and $d_1, d_2 \cdots d_n$ represents the amount of the final output of this topology [19].

In a convolutional neural network, the $a_1, a_2 \cdots a_n$ news text feature data is input into it, and the processed data features are as follows:

$$Q = \left( \sum_{i=1}^{N} \frac{b a_i - c_j}{d_i} \right).$$ \hspace{1cm} (11)

In the formula, $c_j$ represents the translation factor of the convolutional kernel number, $a_i$ represents the scaling factor, $b a_i$ represents the weight value after the feature data input, and $\varepsilon$ represents the excitation function [19].

Calculate the output values obtained from the above feature data to obtain the final training sample data, i.e.,

$$T_k = \sum_{i=1}^{n} \mu_{ik} Q.$$ \hspace{1cm} (12)

In formula (12), the $\mu_{ik}$ represents the weight value of the convolutional kernel.

In this network, because it is prone to certain errors in the forward propagation, the forward propagation in the neural network is corrected to achieve a more accurate classification of news text. Therefore, the error correction is accomplished by using the gradient descent method.

$$W = \sum_{i=1}^{n} (x_i - y_i).$$ \hspace{1cm} (13)

In formula (13), $x_i$ is the ideal output of the $K$ feature data and $y_i$ is the actual output value of the $K$ feature data.

In the error values obtained above, their weight values and scaling factors are corrected according to the gradient descent method, i.e.,

$$b = b_j + \Delta b_{j+1},$$

$$\Delta \mu_{j+1} = -\tau \frac{\partial W}{\partial \mu_{j}}.$$ \hspace{1cm} (14)
In the formula, $\tau$ represents the learning efficiency values [20].

From the above analysis, a convolutional neural network classifier constructs the data trained on news text, i.e.,

$$F_i = \frac{\sum_{n=1}^{\infty} u_i d_{ij}}{1}. \quad (15)$$

The classification is completed according to the news text classifier obtained by formula (15), and the classification process of the method is shown in Figure 3.

### 3. Experimental Analysis

#### 3.1. Design of Experimental Scheme

**Experimental environment of this paper:** The operating system is Ubuntu14.04. The processor is the Intel Core i5. The dataset used by the development tool for Pycharm community edition 3.4 experiments was collected from recent hot news content from Sina Weibo and manually annotated. The four datasets were DBMC-1, DBMC-2, DBMC-3, and DBMC-4. Among them, $C$ represents the number of categories of comment content. Here is the positive and negativity of the comments’ content. The SN represents the number of sentences per news text. The average length of each sentence is 7. The $N$ represents the size of the dataset. The $|V|$ represents the size of the forming dictionary. Test represents the proportion of the test set to the dataset. Details are shown in Table 1.

The convolutional neural network structure used in this experiment mainly consists of 1 word embedding layer, 1 convolutional layer, and 1 pooling layer, the dimension of the word vector is set to 128, the size of the convolutional kernel window is set to $3 \times 128, 4 \times 128, 5 \times 128$, etc., and the number of convolutional nuclei is 128.

#### 3.2. Design of Experimental Indicators

Verify the effectiveness and accuracy of the classification method. Among them, the classification accuracy is a percentage, the higher the value represents the higher the efficiency of classification, about low classification time consuming, the better the classification efficiency.

![Figure 4: Comparison of news text classification accuracy for different methods.](image)

![Table 2: Time-consuming analysis of sample news text classification (s).](table)

| Number of iterations | Methods of this paper | Document [5] methods | Document [6] methods |
|----------------------|-----------------------|----------------------|----------------------|
| 20                   | 2.4                   | 4.9                  | 3.7                  |
| 40                   | 2.4                   | 4.0                  | 3.5                  |
| 60                   | 2.5                   | 4.9                  | 3.2                  |
| 80                   | 2.4                   | 4.8                  | 3.6                  |
| 100                  | 2.4                   | 4.7                  | 3.2                  |

#### 3.3. Analysis of Experimental Results

The proposed method, the literature [5] method, and the results are shown in Figure 4.

From the data in Figure 4, the proposed method, literature [5] method, and literature [6] method classified more than 90%, while the other two methods showed the rise, which is due to the spacing between any news text data; code any news text data, train the news text classifier of deep learning convolutional neural network, and complete the classification of news text.

To further verify the effectiveness of this method, the time-consuming classification of the proposed method, the literature [5] method, and the literature [6] method are compared, and the results are shown in Table 2.

Analyzing the data of experimental results in Table 2, we can see that with the number of iterations, the time-consuming method of sample news text classification is different with the proposed method, literature [5] method, and literature [6] method. When the number of iterations is 20, the proposed method has a time-consuming $N$. Text classification of sample news is about 2.4 s. The literature [5] method takes about 4.9 s, of sample news text classification. The literature [6] method takes about 3.7 s, for the classification of sample news text. When the number of iterations is 60, the proposed method takes about 2.5 s, The literature [5] method takes about 4.9 s, of sample news text classification. The literature [6] method takes about 3.6 s. As shown by the comparison of the experimental data, the classification of this paper is shorter and has faster speed. This is due to analyzing the basic structure of deep learning.
convolutional neural network, determining the change value of its convolutional layers in the convolutional neural network, training news text data, building a news text classifier of deep learning convolutional neural networks, and completing the classification of news text.

4. Conclusion

In order to improve the classification accuracy of news text, a new classification method based on deep learning convolutional neural network is proposed. Determine the weight of the news text data through the VSM vector space model, calculate the information gain of mutual information, determine the characteristic news text data of the news text data, calculate the distance between any news text data, and analyze the basic structure of the deep learning convolutional neural network. Determine the change value of the convolution kernel, train the news text data, and build a news text classifier. The experimental results show that the deep learning convolutional neural network can improve the accuracy of news text classification and increase the classification speed.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The author declares that he has no conflict of interest.

References

[1] R. Asgarnezhad, M. Soltanaghaei, and S. A. Monadjemi, “An application of MOGW optimization for feature selection in text classification,” The Journal of Supercomputing, vol. 16, no. 15, pp. 154–160, 2020.

[2] A. Esuli, A. Moreo, and F. Sebastiani, “Funnelling: a new ensemble method for heterogeneous transfer learning and its application to cross-lingual text classification,” ACM Transactions on Information Systems, vol. 37, no. 3, pp. 1–30, 2019.

[3] M. Oleynik, A. Kugic, Z. Kasac, and M. Kreuzthaler, “Evaluating shallow and deep learning strategies for the 2018 n2c2 shared task on clinical text classification,” Journal of the American Medical Informatics Association, vol. 26, no. 11, pp. 1247–1254, 2019.

[4] M. A. Ibrahim, M. U. Ghani Khan, F. Mehmood, M. N. Asim, and W. Mahmood, “GHS-NET a generic hybridized shallow neural network for multi-label biomedical text classification,” Journal of Biomedical Informatics, vol. 116, no. 1, article 103699, 2021.

[5] Z. Zhenhao, G. Yi, H. Meiqi, and W. Jixiang, “Research on short text classification based on keyword similarity,” Application Research of Computers, vol. 37, no. 1, pp. 26–29, 2020.

[6] N.-j. Qiu, P. Gao, P. Wang, and Y. Tao, “Research on ACO-WNB classification algorithm based on improved information gain,” Computer Simulation, vol. 36, no. 1, pp. 295–299, 2019.

[7] R. Chen, C.-g. Ren, Z.-y. Wang, Z.-j. Qu, and H.-p. Wang, “Attention based CRNN for text classification,” Computer Engineering and Design, vol. 40, no. 11, pp. 3151–3157, 2019.

[8] Z. Chen and J. Ren, “Multi-label text classification with latent word-wise label information,” Applied Intelligence, vol. 14, no. 9, pp. 1–14, 2020.

[9] A. Gcs and B. Mi, "Stacked DeBERT: all attention in incomplete data for text classification,” Neural Networks, vol. 136, no. 1, pp. 87–96, 2021.

[10] J. Krishna, H. Purohit, and H. Rangwala, “Diversity-based generalization for neural unsupervised text classification under domain shift,” ECML-PKDD, vol. 15, no. 24, pp. 1145–1152, 2020.

[11] B. Cowley, M. Filetti, and K. Lukander, “The psychophysiology primer: a guide to methods and a broad review with a focus on human–computer interaction,” Foundations and Trends in Human–Computer Interaction, vol. 9, no. 3, pp. 151–308, 2016.

[12] I. Baldini, D. Wei, K. N. Ramamurthy, M. Yurochkin, and M. Singh, “Your fairness may vary: group fairness of pre-trained language models in toxic text classification,” International Conference on Information Technology in Medicine & Education, vol. 45, no. 1, pp. 15–21, 2021.

[13] J. Jang, Y. Kim, K. Choi, and S. Suh, “Sequential Targeting: an incremental learning approach for data imbalance in text classification,” International Conference on Information Technology in Medicine and Education, vol. 45, no. 1, pp. 147–154, 2020.

[14] I. Chalkidis, M. Fergadiotis, S. Kotitsas, P. Malakasiotis, N. Aletas, and I. Androutsopoulos, “An empirical study on large-scale multi-label text classification including few and zero-shot labels,” International Conference on Information Technology in Medicine and Education, vol. 45, no. 2, pp. 123–128, 2020.

[15] T. Igamberdiev and I. Habernal, “Privacy-preserving graph convolutional networks for text classification,” International Conference on Information Technology in Medicine and Education, vol. 7, no. 1, pp. 868–871, 2021.

[16] Y. Xu, L. Gui, and T. Xie, “Intelligent recognition method of turning tool wear state based on fusion information and BP neural network,” Shock and Vibration, vol. 2021, no. 8, Article ID 7610884, p. 10, 2021.

[17] C. Y. Peng, U. Rahany, S. W. Kuo, and Y. Z. Chen, “Sound detection monitoring tool in CNC milling sounds by K-means clustering algorithm,” Sensors, vol. 21, no. 13, 2021.

[18] D. C. Mackintosh-Franklin, “An evaluation of formative feedback and its impact on undergraduate student nurse academic achievement,” Nurse Education in Practice, vol. 50, no. 4, article 102930, 2021.

[19] Y. Chen, X. Wang, and X. Du, “Diagnostic evaluation model of English learning based on machine learning,” Journal of Intelligent & Fuzzy Systems, vol. 40, no. 2, pp. 2169–2179, 2021.

[20] X. Liu, “Feature recognition of English based on deep belief neural network and big data analysis,” Computational Intelligence and Neuroscience, vol. 2021, no. 6, Article ID 5609885, p. 10, 2021.