Research on Application of Emergency Text Classification Based on Combination Model

Long Lyu[1], Yinghua Song[1] and Dan Liu[1]

[1]School of Safety Science and Emergency Management, Wuhan University of Technology, Wuhan Hubei 430070, China.
Email:18381305006@163.com

Abstract. Based on the powerful learning ability of the deep learning model, the text uses convolutional memory network (CLSTM) to process the text information of the emergency. Firstly, the convolutional neural network is used to learn the local spatial feature information of the text, and then the text is extracted by the long and short memory neural network. The time feature information is used to scale the feature information by using the Softmax layer to finally obtain the event text category. Through experimental comparison, the precision, recall and comprehensive values of CLSTM model reached 0.879, 0.877 and 0.848, respectively, which were significantly higher than MLP and CNN models.

1. Introduction

In the past three decades, the rapid development of the Internet has made the network an important carrier of information dissemination. Every day, a large amount of network information is generated around the world, and it is transmitted on the Internet in the form of sound, text and images. Faced with such a huge amount of information, it is very important for users to get the information they care about in a short time. At the same time, the network platform will generate a large amount of news information, including the information content of emergencies such as natural disasters, social security incidents, public health incidents and accident disasters[1]. Compared with ordinary events, emergencies have obvious characteristics: sudden Sex, destructive, complex, limited information, uncertainty and variability, persistence and publicity. Because of its sudden occurrence, it will have a greater impact in society in a short period of time. At the same time, the information held by the public is very limited. It may develop in an uncontrollable direction in a short period of time and cause negative emotions to the society. Therefore, the relevant departments need to monitor the situation in real time. The emergence of incidents, as well as the media's attention to emergencies, to avoid serious impacts on society and the public.

   Text classification can be roughly divided into three periods from the technical development process: traditional text classification, shallow learning text classification and deep learning text classification. The traditional text classification technology is based on the mathematical statistics knowledge. The machine learning method is used to calculate the probability that the text belongs to various categories. The most probable one is the final classification. The typical method is the naive Bayesian method[2], decision tree. Method[3] and KNN method[4]. The shallow learning text classification technology is also calculated by using mathematical statistics. The typical methods are support vector machine (SVM)[5], BP neural network[6], integrated learning method[7]. As a hot field, the deep learning method mainly constructs a neural network model by simulating the information transfer between neurons in organic organisms, extracts text information, and finally realizes classification function. The commonly used model has deep neural network (DNN)[8], Convolutional

Neural Network (CNN)[9], Recurrent Neural Network (RNN)[10], and Deep Confidence Network (DBN)[11].

In this paper, the Convolutional Memory Network (CLSTM) depth combination model is used. The model expresses the incident text information in word vector mode. First, the word vector is input to the convolutional neural network (CNN), and the convolution kernel is used to extract the local spatial features of the text. Then, the feature information is input to the long-term memory neural network (LSTM) to learn the text time feature information, the second is connected to the Softmax layer scaling feature information, and finally the full-connection layer is connected to the burst event text category.

2. Introduction to The Classification Process of Emergency Texts

The classification of emergency texts consists of two main steps: sample collection and processing and combined model training. The process is shown in Figure 1:
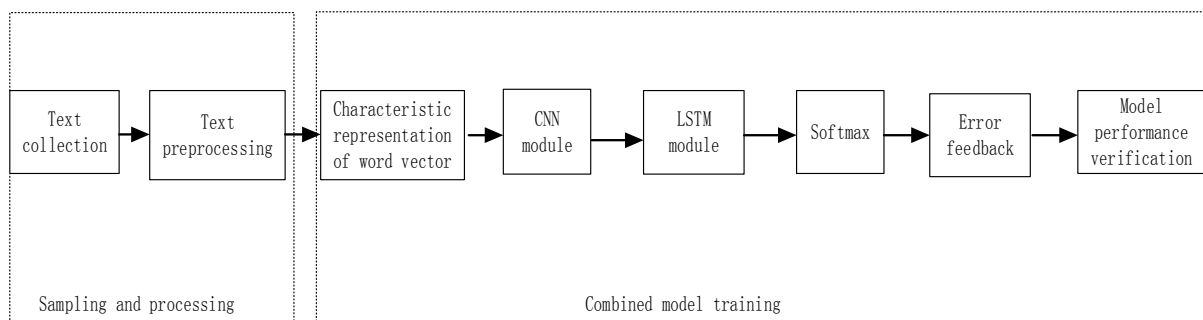

Figure 1. Emergency file text classification flow chart

2.1.Emergency Text Collection

Incidents fall into four categories: public health events, social security events, accident disaster events, and natural disaster events. Through reptile technology, this article searches for 23 categories of secondary category events in Baidu News in the form of keywords: explosions, virus infections, shipwrecks, earthquakes, vicious killings, child trafficking, aircraft crashes, tsunami disasters, traffic accidents, Financial fraud, terrorist attacks, mine accidents, storms, mass incidents, forest fires, sandstone disasters, food poisoning, urban catastrophic fires, hazardous chemicals accidents, plutonium incidents, juvenile crimes, vaccine incidents and rainwater disasters, each One class crawled 150 articles for a total of 3,450 articles. Through manual proofreading, the irrelevant text is eliminated, and 667 valid texts are finally saved, and saved in the txt file in the form of title, time, and body.

2.2.Text Preprocessing

At this point, the data text belongs to the original sample, and the text needs to be pre-processed. The main operations include: filtering non-Chinese information, eliminating the information that is insignificant to the missing text; Chinese word segmentation, so that the text represents the text feature in the form of words; Eliminate the useless words such as connectives, relational words, and modal words in the words; dataset segmentation, randomly divide the text according to the 'training set: test set=4:1' ratio; text length processing, keep the text information fixed length, redundant Intercepting, insufficient zero padding.

2.3.Text Feature Representation

There are three commonly used feature representation methods: one-hot method, TF-IDF method and word vector method.

The One-hot method only considers whether words appear, ignores the weight of words in the text, and when the amount of text data is large, there will be a phenomenon of dimensional explosion and feature stagnation, which is a big performance for computer memory and model. The challenge.

Although the TF-IDF method considers the importance of words in the text category, it ignores the semantic relationship between words. When the amount of text data is large, there will also be a phenomenon of dimensional explosion and feature stagnation.

The word vector method maps the features of each word into a vector of several dimensions. All words can be represented by vectors of fixed length. At the same time, the word vector method also takes into account the semantic relationship before and after the words, meaning or function similar. The vector space distance of words is also small, in other words, the similarity is high. The word vector mainly implements feature extraction functions by two kinds of neural networks: CBOW (continuous bag-of-word) and Skip-gram (continuous skip-gram gram). The CBOW model predicts intermediate words based on several words before and after the sentence; the Skip-gram model predicts several words in the future based on one word. The objective function of the two models is:

$$\text{loss} = \sum_{w \in c} log p(w|Context(w)) \tag{1}$$

In this paper, the gensim module based on the Skip-gram model is used to train the word vector. If the word vector dimension is too small, the word vector feature cannot be fully represented. If the dimension is too large, the resource is wasted. Therefore, it is necessary to select the appropriate dimension.

2.4. CNN Module

The CNN network is a feedforward neural network whose artificial neurons can respond to a surrounding area of a part of the coverage, which includes a convolutional layer and a pooling layer. The core advantage of this network lies in its shared convolution kernel. The convolution kernel can extract information of several words at a time. The convolution kernel can be shared in one operation, reducing model parameters. The principle is:

The corresponding word vector of the word x_i in the convolutional layer after k-dimensional embadding is $£^k$, $x_i \in £^h$, a sentence $X_{1:n}$ can be formalized as:

$$X_{1:n} = x_1 \oplus x_2 \oplus \cdots x_n \tag{2}$$

A window with h words is represented as:

$$X_{i:i+h-1} \in £^{hk} \tag{3}$$

A filter matrix of size h*k is converted into a one-dimensional vector of length h*k, expressed as:

$$W \in £^{hk} \tag{4}$$

$$c_j = f(W \cdot X_{i:i+h-1} + b) \tag{5}$$

The eigenvalues extracted by a filter are:

$$\hat{c} = \max(c) \tag{6}$$

Filters can get the feature vector:

$$z = [\hat{c_1}, \hat{c_2}, \hat{c_3} \dots \widehat{c_m}] \tag{7}$$

2.5. LSTM Module

The LSTM unit structure has an input gate, an output gate, and a forget gate. W_f, W_i, and W_c, and W_o are neural network weight parameters of the forgetting gate, the input gate, and the output gate, respectively; b_f, b_i and b_c, b_o represent the neural network offsets of the forgetting gate, the input gate and the output gate, respectively; $\tilde{C_t}$ and h_t represent the state values of the neural unit at time t. Then there are the following calculations:

A, forget gate:

$$f_t = \sigma(W_f \cdot [h_{t-1}, P(t:)] + b_f) \tag{8}$$

B, input gate:

$$i_t = \sigma(W_i \cdot [h_{t-1}, P(t:)] + b_i) \tag{9}$$

$$\widetilde{C}_t = \tanh(W_c \cdot [h_{t-1}, P(t:)] + b_c) \tag{10}$$

C, output gate:

$$o_t = \sigma(W_o \cdot [h_{t-1}, P(t:)] + b_o) \tag{11}$$

$$h_t = o_t * \tanh(f_t * C_{t-1} + i_t * \widetilde{C}_t) \tag{12}$$

2.6.Model Evaluation

In the classification problem, the common prediction results are shown in Table 1:

Table 1. Classification result possibility

		actual results	
		positive	Negative
forecast result	positive	True positive(TP)	False positive(FP)
	Negative	False Negative(FN)	True Negative(TN)

The model is evaluated using precision, recall, and composite value (F1) as follows:

$$\text{precision} = \frac{TP}{TP+FP}, recall = \frac{TP}{TP+FN}, \frac{2}{\text{F1}} = \frac{1}{accuracy} + \frac{1}{recall} \tag{13}$$

3. Analysis of The Results

3.1.Experimental Tools

Dataset: 667 news reports of emergencies (including 23 categories)

  System: Windows 10 64-bit operating system

  Processor: AMD Ryzen7 3700U

  Programming tool: anaconda3

  Toolkit: Jieba module, gensim module, pandas module, sklearn module, numpy module, keras module

3.2.Model Assistant

The parameters of the emergency text classification model based on deep learning have great influence on the performance of the model. In order to optimize the performance of the model, based on the same sample set, the accuracy is obtained by repeating the simulation three times and changing the parameters of a single model each time. The average value of the recall rate and the comprehensive value is the best parameter when the comprehensive value is the largest. In order to analyze the performance of the CLSTM combined model, the MLP and CNN models are trained and compared with the same method. The optimal parameters of the model are shown in Table 2-4.

Table 2. MLP model parameters and optimal values

Parameter	Best Value	Parameter	Best Value	Parameter	Best Value
Number of hidden layer units	640	Hidden layer activation function	Relu	Dropout	0.6

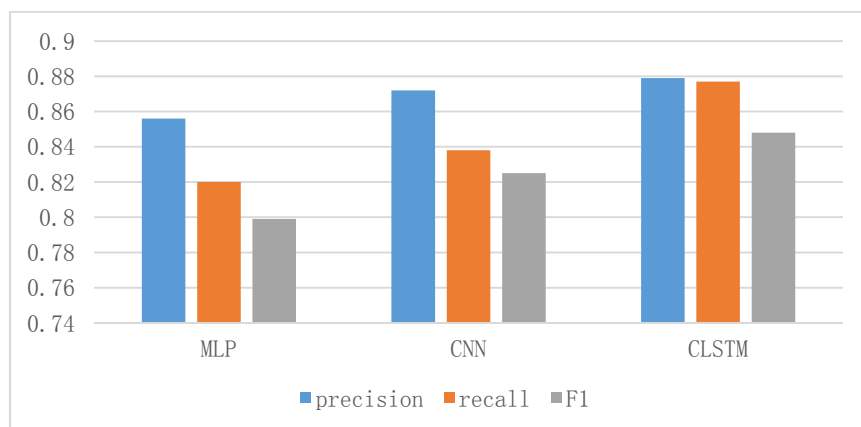**Table 3.** CNN model parameters and optimal values

Parameter	Best Value	Parameter	Best Value	Parameter	Best Value
Sentence length	400	Number of word dimensions	260	Number of kernel	768
Convolution kernel activation function	Tanh	Dropout	0.7	Number of hidden layer units	128
Hidden layer activation function	Elu				

Table 4. CLSTM model parameters and optimal values

Parameter	Best Value	Parameter	Best Value	Parameter	Best Value
Sentence length	400	Number of word dimensions	260	Number of the first kernel	768
First layer convolution kernel activation function	Tanh	Number of the next kernel	1024	Next layer convolution kernel activation function	Softplus
LSTM unit number	256	LSTM Dropout	0.8	Full connection layer dropout	0.7
Number of fully connected layer units	512	Full connection layer activation function	Rlu		

3.3.Model Comparison Analysis

Compare the models according to the optimal state values of all the tunable parameters in each model obtained in the previous step. All model uniform parameters have batch training size (batch_size) and sample training times (epochs), set to: batch_size =16, epochs=20. The results of each model are shown in the following table and figure:

Figure 2. Model performance comparison

Through experimental comparison and analysis, we can know:

(1) Comparison of model complexity: MLP≺CNN≺CLSTM; accuracy ratio comparison: MLP≺CNN≺CLSTM; recall ratio comparison: MLP≺CNN≺CLSTM; comprehensive value comparison: MLP≺CNN≺CLSTM. It can be seen that the more complex the model structure, the more layers there are, the more the number of neural units, the stronger the learning ability.

(2) The CLSTM model accuracy rate is 0.879, which is significantly higher than MLP (0.856) and

CNN (0.872); its recall rate is 0.877, which is 5.7 percentage points higher than MLP (0.82) and CNN (0.838), respectively, and 3.9 percentage points; The comprehensive value takes into account the balance between accuracy and recall. The overall performance of the model can be considered. The integrated value of the CLSTM model (0.848) is significantly higher than MLP (0.799) and CNN (0.825).The data shows that the model has higher accuracy than the other two models.

4.Conclusion

In this paper, the CRNN model is used to learn text information. Firstly, the CNNN local information extraction capability is used to obtain the text space information, and then the LSTM model is used to obtain the sequence information to extract the text time feature information. Through experiments, the CLSTM model has a better classification effect in the emergency news classification.

5.References

[1]   Yinghua Song. Introduction to emergency management of emergencies [M]. Beijing: China Economic Publishing House, 2009: 15-17, 21-25, 32-33.
[2]   Zhibin Xiong, Jianfeng Zhu, Chengguo Yin et al. Automatic classification of tourism emergencies based on Weka[J].Software Guide, 2016, 15(04): 154-156.
[3]   Hothorn T, Lausen B, Benner A, et al. Bagging survival trees, Statistics in Medicine, 2004, 23: 77-91.
[4]   Xu Wei, Lingyu Liu, Mingming Wang. Research on Incident Detection and Trend Research Based on Cross-media Analysis[J]. Systems Engineering - Theory & Practice, 2015, 35(10): 2550-2556.
[5]   Zhang D, Xu H, Su Z, et al. Chinese comments sentiment classification based on word2vec and SVM perf [J]. Expert Systems with Applications, 2015, 42(4): 1857-1863.
[6]   Xinsheng Liu, Li Wei. Design and Implementation of a Text Classification System for Tourism Emergencies Based on BP Neural Network[J]. Computer and Modernization, 2011(07): 192-194+198.
[7]   Hothorn T, Lausen B. Double 2bagging: Combining calssifiers by boot strap aggregation, Pattern Recognition, 2003, 36: 1303-1309.
[8]   Rao A, Spasojevic N. Actionable and political text classification using word embedding and lstm [J]. Compute Science, arXiv: 1607.02501, 2016.
[9]   Kim Y. Convolutional neural networks for sentence classification [J]. Eprint Arxiv, 2014, 1181.
[10]   Socher R, Lin C.Y, Ng A.Y, et al. Parsing natural scenes and natural language with recursive neural networks [C]. International Conference on International Conference on Machine Learning. Omnipress, 2011: 129-136.
[11]   Shubao LHU, Mingyue Wang, Pei Xiang et al. An information text classification algorithm for deep learning[J]. Journal of Harbin University of Science and Technology, 2017, 22(02): 105-111.