Chinese Emergency Event Recognition Based on BiGRU-AM Model

Haoran Yin, Jinxuan Cao*, Guodong Wang, Houlu Zhang
People's Public Security University of China, Beijing, China

*Corresponding author: caojinxuan@163.com

Abstract. A Chinese emergency event recognition based on bidirectional gated recurrent unit BiGRU-AM model with attention mechanism is proposed to resolve the limitation of traditional event recognition methods and the poor interpretability of general recurrent neural networks in respect of information features with different degrees of importance. Firstly, text corpus was trained to generate word vectors, and contextual information features were extracted through BiGRU, and then attention mechanism was introduced into BiGRU network to make feature extraction more selective. Finally, the learned features were activated by softmax function to output recognition results. Simulation results show that this method improves the accuracy and recall rate of emergency recognition, and the F value is superior to other methods.

1. Introduction

An event, as a form of information, is defined as the objective fact that a specific person, thing interacts at a specific time and a specific place [1]. The Internet is full of all kinds of disorderly and unexpected news. These news are interspersed with other types of news, hindering users' clear understanding of emergencies and the classification and storage of related researchers [2], so how implementing emergency identification in the network is one of the problems that need to be solved at present. Event recognition is an important basis for event extraction. Event recognition refers to extracting uniformly formatted, structured, and user-interested event information from unstructured text information, and classifying the events accordingly [3].

At present, there are three methods for event recognition: pattern matching, machine learning, and deep learning. Based on the pattern matching, the feature templates of the corresponding events are designed manually in advance. Jiang Deliang [4] matches the text of the incident according to a predefined pattern, standardizes the extracted results, and finally integrates the obtained information into a structure. In the form of transformation, good results are obtained, but relying on expert domain knowledge, the model generalization ability is not strong. Based on machine learning, which mainly focuses on the discovery of features and how to build a classifier, so that event recognition is transformed into a classification problem, Yao Zhanlei et al. [5] proposed an improved TF-PDF algorithm based on the relatively stable combination of hot words. Identify emergency events; He Zhongshi et al. [6] added semantic role SR features on the basis of extracting traditional features, and then used CRFs to identify emergency events. The method based on pattern matching requires expert domain knowledge, and the artificial design features have a high accuracy rate in the professional field, but it is less portable for event recognition in the open field, and statistical machine learning methods have the principles and
foundation of statistics have avoided the dependence on expert knowledge to a certain extent, and have high portability, but the accuracy rate needs to be improved. In order to solve the above problems, deep learning has received more and more attention in the field of emergency event recognition. Li Hong et al. [7] proposed the DCNNs-LSTM model to realize the emergency event recognition of the adhesive language Uyghur; [8] proposed a recurrent convolutional neural network model with attention mechanism based on language model, which mainly solves the problem of detecting polysemy and multi-event sentences; Huang Xifeng [9] proposed a dynamic masking attention mechanism model. Capturing richer contextual information has achieved leading experimental results in multi-event extraction tasks; Zhang Yajun et al. [10] adopted hybrid supervised deep belief networks to improve the recognition of events and realize the identification of other relevant elements of events.

Based on previous studies, this paper proposes a Bidirectional Gated Recurrent Unit-Attention Mechanism (BiGRU-AM) model that incorporates an attention mechanism. The BiGRU part of the model captures the context feature information of the text corpus, and uses the attention mechanism to weight the extracted feature information to make the feature selection more selective, thereby reducing the noise impact of irrelevant features and enhancing the model's importance to different levels of features. The interpretability of information makes the model have better feature learning ability and generalization ability, thereby improving the recognition effect of network emergencies.

2. Model

2.1. Overall Framework

Because a single recurrent neural network feature learning framework has no distinguishing mechanism for the feature sequence information output at each moment, the significance of the features at each moment is poorly interpretable, and universal weighted bias retransmission of the feature sequence information at each moment is easy to cause The lack of effective feature information has poor robustness. Therefore, an attention mechanism is introduced to automatically weight the feature sequence information output by the neural network, so as to capture feature information of different importance levels, so that the model can focus on extracting the features of several words, and reduce the relatively less important word features against emergencies. Impact of event identification. For example, the trigger word "earthquake, car accident, vomiting, explosion" has a greater impact on the recognition result of an emergency, and the corresponding attention weight coefficient will be higher.

The emergency recognition based on BiGRU-AM model proposed in this paper includes four parts: input layer, BiGRU, attention layer, and output layer. As shown in Figure 1.
2.2. **Input Layer**

This article uses jieba word segmentation to process the text corpus $T$ to remove stop words from the processed text. In deep learning, the text corpus input and output are previously fixed sizes, but the text corpus data is different from the two-dimensional data such as images. The sentence length in the text is not fixed, and fixed-length processing is required. Therefore, the input sequence length is set as $m$ words, the participle text of more than $m$ words intercepts the first $m$ words, and the participle text of less than $m$ word length is filled and complemented.

The pre-processed text is trained using the Word2Vec model, and the text word vector is trained to replace the traditional one-hot method, which can avoid the problem of sparse vector space and dimensional disaster, thereby transforming the vector space into a low-dimensional dense form. The training method uses the Skip-Gram model to obtain the $n$-dimensional word vector $v_t \in \mathbb{R}^n (t=1,2,\ldots,m)$, then the text word vector matrix $V=[v_1, v_2, \ldots, v_m]$, where $V \in \mathbb{R}^{m \times n}$.

2.3. **BiGRU**

In the task of emergency event recognition, the operation of the bidirectional recurrent neural network is to input the word vector matrix into the recurrent neural unit, pass forward and backward, and finally output. The main purpose is to capture the sequence information characteristics of the entire text word vector and obtain context information. This paper uses GRU as the recurrent neural network unit. The GRU simplifies the LSTM with one gate missing and a simpler structure, as shown in Figure 2.

![GRU Unit](image)

**Figure 2. GRU Unit**

The process of BiGRU capturing sequence feature information is as follows:

1) Input the text word vector $v_t$ into the GRU unit, and connect it with the sequence information $h_{t-1}$ at the previous time of the hidden layer, and obtain the update signal $z_t$ and the reset signal $r_t$ through the update gate and reset gate operations:

Update Gate: $z_t = \sigma(W_z \cdot [h_{t-1}, v_t])$  \hspace{1cm} (1)

Reset Gate: $r_t = \sigma(W_r \cdot [h_{t-1}, v_t])$  \hspace{1cm} (2)

2) Determine the importance of $h_{t-1}$ to the candidate memory unit by $r_t \cdot h_{t-1}$, connect it with $v_t$, and obtain the candidate memory unit $\tilde{h}_t$ by $W$ weight and tanh activation function.
Candidate Memory Unit: \( \tilde{h}_t = \tanh(W \cdot [r_t \cdot h_{t-1}, v_t]) \)  

(3)

3) The calculation of the update signal \( z_t \) with \( h_{t-1} \) and \( \tilde{h}_t \) determines how much weight \( h_{t-1} \) passes to the next state, and obtains the current state memory unit \( h_t \):

\[
\text{Current Memory Unit: } h_t = (1 - z_t) \cdot h_{t-1} + z_t \cdot \tilde{h}_t
\]

(4)

4) The above steps are performed separately for the forward GRU and backward GRU neural units, so BiGRU is obtained. By using the forward GRU to calculate the feature information \( h^\leftarrow_t \) at time \( t \) and the backward GRU to calculate the feature information \( h^\rightarrow_t \) at time \( t \), the output feature information \( o_t = [h^\leftarrow_t, h^\rightarrow_t] \) at that time can be obtained.

5) Connect \( o_t \) at each moment to form a eigenvector matrix \( O = [o_1, o_2, \ldots, o_m] \), where \( o_t \in \mathbb{R}^{2n}(t=1,2,\ldots,m) \), \( O \in \mathbb{R}^{m \times 2n} \), feature matrix \( O \) is as input for the next layer.

Where \( w_z, w_r, w \) represents weight; \( \sigma, \tanh \) represents activation function; "*" operator represents matrix element multiplication, and "[[]]" represents vector connection.

2.4. AM Layer

In the task of emergency identification, the operation of the attention layer is to automatically weight the feature information of the previous layer. The main purpose is to make the contribution of important feature information to the results be reflected in the weighting process, so that the model can focus on extracting the features of several words. The specific steps are as follows:

1) After multiplying the input feature vector \( o_t \) by the weight matrix \( w_a \) and adding the offset \( b_a \), \( u_t \) is obtained by the activation function \( \tanh \):

\[
u_t = \tanh(o_t \cdot w_a + b_a)
\]

(5)

2) Let \( e_u = u_t, E = (e_1, e_2, \ldots, e_m), A = (a_1, a_2, \ldots, a_m) \), then \( A = \text{softmax}(E) \), we can get the weight \( a_t \) automatically assigned by each feature vector:

\[
a_t = \frac{\exp(u_t \cdot u_t^T)}{\sum \exp(u_i \cdot u_i^T)}
\]

(6)

3) Multiplying the feature vector \( o_t \) at each time with the weight \( a_t \) of its corresponding time, and obtaining the weighted result \( S \):

\[S = \sum a_t o_t\]

(7)

Where \( w_a, u \) represent weights, and are continuously optimized by model training after initialization; \( o_t \) represents the eigenvector of the output matrix \( O \) of the previous layer; \( b_a \) represents the offset; \( w_a \in \mathbb{R}^{2n \times 2n}, b_a \in \mathbb{R}^{2n}, u \in \mathbb{R}^{2n}, S \in \mathbb{R}^{2n} \).

2.5. Output Layer

The specific operation steps of the output layer are as follows:

1) The output result \( S \) of the previous layer is obtained through the fully connected neural network and the activation function softmax operation to obtain the event category probability distribution \( P \):

\[P = \text{softmax}(S)\]
\[
P = \frac{\exp\left(Sw_{fc} + b_{fc}\right)}{\sum \exp\left(Sw_{fc} + b_{fc}\right)}
\]

(8)

Where \(w_{fc}\) represents the weight and \(b_{fc}\) represents the offset; \(w_{fc} \in \mathbb{R}^{2n \times c}, b_{fc} \in \mathbb{R}^{c}\), \(P = \{p_1, p_2, \ldots, p_n, \ldots, p_c\}\)(\(i \in [1, c]\), c is number of emergency event types.

2) The training process uses a cross-entropy cost function as the loss function, so that the learning rate is \(LR\), the loss value is \(loss\), and the optimizer Adam is used for random gradient descent:

\[
loss = \text{average}\left(-\sum \text{label} \times \text{tf.log p}_i\right)
\]

(9)

\[
\text{AdamOptimizer}\left(LR\right) . \text{minimize}\left(loss\right)
\]

(10)

3) The verification and testing process uses the \(\text{arg.max}\) function. Use this function to get the index value of the maximum probability \(p_i\) in the event category probability distribution \(P\), and match this index value with the actual label of the event type:

\[
\text{label} = \text{arg._max}\left(p_i\right)
\]

(11)

3. Experiment and Analysis

3.1. Emergency Event Dataset

The corpus data used in this paper is the Chinese Emergency Corpus (CEC) set up by the Semantic Intelligence Laboratory of Shanghai University, as well as data obtained through web crawlers. The corpus data are derived from news reports of emergencies on the Internet, and data from the media, which contains five types of emergencies: earthquakes, fires, traffic accidents, terrorist attacks and food poisoning. There are 332 CECs in total; 28,267 data obtained by the crawler, and 25,530 in total after deduplication. In this paper, experiments are performed on the two data sets. The training set, the verification set, and the test set are randomly selected and divided according to 7: 2: 1.

3.2. Train Phase

1) This article uses an open source corpus. The original text corpus is in Extensible Markup Language xml format. Regular expressions are used to remove extra xml elements from the text corpus to extract the main body of the text. The corpus is segmented by jieba word segmentation, and redundant stop words are removed from the text after the word segmentation, such as: "", time "year month day", punctuation marks, etc.

2) The Word2Vec model is used to train the pre-processed text, and the training method uses Skip-Gram to obtain a 200-dimensional vector representing the word vector space of each word segmentation. Count the number of segmentation. The segmentation text of more than 300 words intercepts the first 300 words. The segmentation text of less than 300 words is filled with zero vectors to 300 words. The input word vector matrix is a two-dimensional matrix \(V\) of 300 \(\times\) 200;

3) Set the number of hidden neurons in the BiGRU layer to 200 and use the matrix \(V\) as the input to the BiGRU layer to obtain a 300 \(\times\) 400 feature vector matrix \(O\);

4) Set the attention_size of the attention layer to 400, use the matrix \(O\) as the input of the attention layer, and obtain the weighted summed result vector \(S\);

5) The type of emergency is earthquake, fire, traffic accident, terrorist attack, food poisoning, so set the number of emergency types \(c = 5\), pass the vector \(S\) through the last fully connected neural network, and activate it with the softmax function to get the text Emergency category probability distribution \(P\);
6) The cross-entropy cost function is used as the loss function to obtain the loss value loss. The learning rate LR is set to 1e-3, and the Adam optimizer is called for random gradient descent to minimize the loss value, so that the entire model is trained to convergence. Every 10 iterations in the training process, a validation set check is performed to adjust the relevant parameters of the model.

3.3. Test Phase
1) The data pre-processing and word embedding vectors are the same as the training phase, and the processed test set text data is input to the training convergence BiGRU-AM model;
2) Through BiGRU and the attention layer, after being activated by the softmax function, the predicted probability value $P$ of the emergency type is obtained;
3) After the operation of the arg_max function, the index value of the maximum probability $p_i$ in the event type label vector $P$ is matched with the actual event type label, that is, $\text{label} = \text{arg_max} (p_i)$, and the matching result is output.

3.4. Analysis
The analysis method of the results of the emergency identification uses a common evaluation standard, which is the same as that in [15-19], including precision, recall, and F-Value. In this paper, multiple sets of experiments are performed, including a single neural network CNN (Convolutional Neural Networks), GRU, and a joint model BiGRU-AM, and the comparison with other literatures on the effectiveness of emergency model recognition. The experimental comparison results are shown in Table 1.

| Method   | P/%   | R/%   | F/%   |
|----------|-------|-------|-------|
| SVM[15]  | 79.30 | 59.90 | 63.70 |
| CNN      | 72.73 | 64.00 | 68.09 |
| GRU      | 69.70 | 66.67 | 68.15 |
| BiGRU[16]| 71.10 | 69.00 | 70.00 |
| BiGRU-AM | 75.76 | 69.44 | 72.46 |

Since the CEC corpus is a structured XML language, the preprocessed text features are still more obvious and easier to extract and train, which relatively weakens the advantages of the convolutional layer in extracting local features, which is not general and the small amount of CEC data may lead to insufficient training of the model. Therefore, experiments were performed on various unstructured information data crawled from the Internet using various models. The experimental comparison results are shown in Table 2.

| Method   | P/%   | R/%   | F/%   |
|----------|-------|-------|-------|
| SVM      | 77.26 | 53.43 | 63.17 |
| CNN      | 71.59 | 62.69 | 66.85 |
| GRU      | 70.36 | 64.96 | 67.55 |
| BiGRU    | 73.53 | 67.56 | 70.32 |
| BiGRU-AM | 76.32 | 71.76 | 73.96 |

In Table 1 and Table 2, paper [15] uses the traditional machine learning method SVM (Support Vector Machine) to identify emergencies with an accuracy rate of 79.30%. Although a good recognition effect is obtained, from the F value, The other 4 groups of deep learning network models are significantly better than SVM; CNN uses convolutional neural networks to identify emergencies, and GRU uses one-way recurrent neural network to identify emergencies, with an accuracy rate of 69.70%, and the F value is slightly Higher than CNN, it shows that the context information captured by GRU is more conducive
to the identification of emergencies; the F value of [16] using the BiGRU model has reached 70.00%, which is nearly 2% higher than GRU, indicating that the two-way recurrent neural network passed context information features, the effect of identifying emergencies is better than one-way recurrent neural network; BiGRU-AM introduces the attention mechanism on the basis of two-way recurrent neural network, which effectively improves the accuracy and recall rate, indicating that the attention mechanism can Different weight coefficients are assigned according to the degree of influence of features on the results, so that the model can selectively learn important features and improve the recognition of emergencies.

By comparing the experimental results, from the comprehensive evaluation index F value, the F value of the BiGRU-AM model proposed in this paper has reached 72.46%, which is better than other models, indicating that the BiGRU-AM model has achieved better performance in the task of emergency identification. Good results.

4. Conclusion
In the task of emergency recognition, this paper proposes a BiGRU-AM model. The text corpus is first trained as a word vector, the context information features are extracted through BiGRU, and the attention mechanism is introduced into the BiGRU network. The extraction is more selective, thereby reducing the noise impact of irrelevant features, and improving the problem of poor interpretability of recurrent neural networks in sequence information features with different degrees of importance, so that the model has better feature learning ability and generalization ability. Experimental results show the feasibility and effectiveness of the method for emergency event identification.

Considering that the semantic information of text can be further expanded, the model lacks in the extraction of local feature information. The next step is to incorporate more semantic features to enrich the dimension of the word embedding vector in the input layer, and try to add other network structures to improve the model's ability to capture local features and improve the recognition of emergencies.

Acknowledgments
This work was financially supported by the National Key R&D Program of China (2016YFB0801100, 2018YFC0823200, 2018YFC0809800) and the Fundamental Research Funds for the Central Universities of PPSUC (2019JKF108). Correspondence should be addressed to Jinxuan Cao: caojinxuan@163.com.

References
[1] QIN Yanxia, ZHANG Min and ZHEN Dequan, “A survey on neural network-based methods for event extraction”, Intelligent Computer and Applications, 2018, 8(3):1-10.
[2] Xia Yan, He Lin Pan, Yunlai, et al., “Research on Recognition of Sudden Events on Web Based on Combination of Rules and Statistical Method”, New Technology of Library and Information Service, 2010(10):65-69.
[3] Zhong Z, Jin L, Feng Z. Multi-font printed Chinese character recognition using multi-pooling convolutional neural network[C]// International Conference on Document Analysis & Recognition. 2015.
[4] JIANG De-liang, “Research on extraction of emergency event information based on rules matching”, Computer Engineering and Design, 2010,31(14):3294-3297.
[5] Xu xin, “Research on the Detection of Sudden Events in News Stories of Online Information”, New Technology of Library and Information Service,2011(04):52-57.
[6] HE Zhong-shi, LIU li, XING Xin-lai, et al., “Chinese events recognition based on semantic role”, Computer Engineering & Science, 2013,35(04):181-185.
[7] LI Hong,YU Long, TIAN Shengwei, et al., “Uyghur Emergency Event Extracton Based on DCNNs-LSTM Model”, Journal of Chinese Information Processing, 2018, 32(6):52-61.
[8] SHI Zheer, CHEN Jinxiu, “Event detection via recurrent and convolutional networks based on language model”, Journal of Xiamen University(Natural Science), 2019,58(03):442-448.
[9] Huang Xifeng, “Event extraction via dynamic masked attention”, Application Research of Computers. https://doi.org/10.19734/j.issn.1001-3695.2018.12.0927.
[10] ZHANG Ya-jun, LIU Zong-tian, ZHOU Wen, “Event Recognition Based on Deep Belief Network,” Acta Electronica Sinica, 45(06):1415-1423, 2017.
[11] Hongyu Lin, Yaojie Lu, Xianpei Han, et al. Nugget Proposal Networks for Chinese Event Detection[J]. Computer Science, 2018. (2018-05-01)[ 2018-05-01]. https://arxiv.org/abs/1805.00249.
[12] Tu Min, Liu Xiang, Liu Shuchun, “Natural Language Processing Core Technology and Algorithm with Python”, China Machine Press, 2018.
[13] MIKOLOV T, SUTSKEVER I, CHEN K, et al. Distributed Representations of Words and Phrases and their Compositionality[C]//The Neural Information Processing Systems (NIPS) Foundation. Proceedings of NIPS, December 5-10, 2013, Harrahs and Harveys, Lake Tahoe. USA:NIPS, 2013:3111-3119.
[14] Mikolov T, Chen K, Corrado G, et al. Efficient Estimation of Word Representations in Vector Space[J/OL]. Computer Science, 2014. (2014-02-10)[2014-02-10]. http://arxiv.org/pdf/1301.3781.pdf.
[15] XUAN Xiaoxing, LIAO Tao, GAO Beibei, “Automatic Extraction of Chinese Event Trigger Word,” Computer & Digital Engineering, (3):457-461, 2015.
[16] MA Chenxi, CHEN Xingshu, WANG Wenxian et al., “Chinese Event Detection Based on Recurrent Neural Network.” Netinfo Security, (5):75-81, 2018.
[17] QIU Qizhi, ZHOU Sansan, LIU Changfa, et al., “Emergency Information Extraction Based on Style and Terminology”, Journal of Chinese Information Processing, 2018,32(09):56-65+74.
[18] WANG Hong-bin, GAO Hong-kui, “Research on Event Recognition Based on Word2vec and Dependency Analysis”, Computer Engineering & Software, 2017,38(06):62-65.
[19] WEN Chang, LIU Yu, GU Jinguang, “Evolution relationship extraction of emergency based on attention-based bidirectional long short-term memory network model”, Journal of Computer Applications, 2019,39(06):1646-1651.