An Improved Deep Belief Network for Chinese Emergency Recognition

Haoran Yin¹, Jinxuan Cao*, Luzhe Cao² and Guodong Wang³
People's Public Security University of China, Beijing, China

*Corresponding author e-mail: caojinxuan@163.com, ²74335879@qq.com,
³2018211269@ppsuc.com, ⁴2018212038@ppsuc.com

Abstract. Aiming at the defects that the RBM module in DBN can only re-represent information but cannot extract information features, and can only handle one-dimensional data, the DBN network is improved, and a Conv-DBN model is proposed to recognize emergencies. First, the text corpus is preprocessed, and the word vector matrix generated by Word2Vec is used as input, and then the word vector features are extracted through the visible layer integrated into the convolution operation. Word vector features are used as the input of the next layer. Finally, every layers are fine-tuned through back-propagation at the top layer. The softmax function is used to activate, and the recognition result is output. Simulation results show that the method proposed in this paper has improved accuracy and recall, and F value is better than 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]. Event extraction mainly includes two steps: recognizing the events and analyzing the recognized events, and then extracting the event elements [2]. Event recognition aims to obtain structured event information that is of interest to users from unstructured text information and classify the corresponding events accordingly.

At present, there are three methods of event recognition: pattern matching, shallow machine learning and deep learning. Pattern matching is mainly to design a template corresponding to type events in advance for feature matching. Valery Solovyev et al. [3] proposed a method of using Google Books NGram corpus in Russia to create a vocabulary of dependent models, an event trigger vocabulary, and a frame element vocabulary to match Russian events. But the method of pattern matching is dependent on the domain knowledge of experts, and the generalization ability of the model is not strong. Machine learning focuses on how to construct a suitable classifier to transform event recognition into classification problems. Syed A. Yusuf et al. [4] used hidden Markov models and conditional random field classifiers to train discretely label data to obtain better event recognition results. With the development of deep learning technology, neural networks have received more and more attention in the field of event extraction, and successfully applied in this field to solve practical problems. KUN LIU [5] proposed a support vector machine and radial basis neural network. The combined classifier improves the reliability of event recognition results. OKWUDILI M. EZEME et al. [6] proposed an abnormal event recognition model based on long-term and short-term memory networks, which solved the online abnormal event recognition detection in network-physical systems. Zhang Yajun et al. [7] adopted a hybrid supervised deep belief network to improve the recognition
effect of the event, and at the same time realized the identification of other relevant elements of the event.

The traditional event extraction method requires manual design of features and tagging corpora, which consumes a lot of time and manpower. The choice of features directly affects the final effect of the relational classifier, and the use of NLP (natural language processing) tagging tools easily leads to error propagation problems [8]. Therefore, this paper improves the DBN (Deep Belief Network), and proposes the Conv-DBN model. The convolution operation is incorporated into the visible layer to capture the local features of the word vector, and the word vector is used as the input of the next hidden layer. The greedy algorithm is pre-trained and finally fine-tuned by back-propagation at the top layer to make the model converge as a whole and finish training, and then recognize Chinese emergencies.

2. Related Work

2.1. Preproces

By using regular expressions to extract the text sentence that need to be recognized from the text corpus, the neural network is used to extract the features of the words in the sentence and find the existing trigger words and classify them, which are common ideas for English event recognition [9]. Because English itself can use spaces as word separators, the step of word segmentation is omitted. However, for some languages without natural separators, such as Chinese, you need to take the word segmentation operation first. Generally, word segmentation operations performed on text are implemented by word segmentation tools. Common word segmentation tools include jieba, Stanford, CRF++, thulac, etc. [10]. The text corpus processed by word segmentation is trained by the Word2Vec model proposed by MIKOLOV [11] and it can obtain word vectors after train. The training methods are Skip-Gram or CBOW (Continuous Bag of Words) [12].

2.2. DBN

Deep belief network is a neural network model in deep learning. Hinton et al. [13] proposed that the model uses a layer-by-layer greedy learning method, which can effectively avoid the problem that is poor training performance of deep neural networks with multiple hidden layers in the traditional gradient descent algorithm. The model has been successfully applied to images, Speech, event extraction and other fields. The DBN network is made up of multiple layers of RBMs (Restricted Boltzmann Machine). The output of the previous RBM hidden layer is used as the visible layer of the next RBM, which is the input of this layer. In the task of event recognition, the DBN can be regarded as MLP (Multi-Layer Perceptron), and the softmax function is added at the top to be activated.

2.3. Convolution Operation

In a convolutional neural network, the convolution operation can capture significant local features, but the RBM unit in the DBN does not extract information features, but it only re-represents the original data. If the RBM structure can be improved and the information features can be extracted without affecting the greedy algorithm of each RBM layer, the overall performance of the DBN will be better. The state of each unit in the visible layer in RBM is jointly determined by the units of all hidden layers. This paper improves the structure of RBM. The convolution operation is incorporated into the visualization layer of the first RBM layer, so that the model can handle two dimensional data information, and it can improve the recognition performance of the model through the extracted local information features.

3. Model

Considering that the RBM module in DBN only re-represents the information without extracting information features, and can only handle the defects of one-dimensional data, the convolution operations is incorporated in the visible layer to improve the DBN. A Conv-DBN model is proposed to
recognize emergencies. The input of the model can be a two-dimensional feature matrix, which changes the idea of inputting one-dimensional data from the original network. By using the convolution operation as the operating unit of the first-level RBM, local information features in the input data can be extracted and down-sampling is realized, which reduces the number of hidden layer neural units and weights, and improves the overall performance of the model. The process of event recognition is shown in Fig. 1.

3.1. **Input Layer**
This paper uses the jieba word segmentation tool to perform word segmentation on text corpus and remove stop words. Because the sentence length in each text is inconsistent, fixed-length processing is required. Therefore, the length of the word segmentation in this paper is m words. Segmentation text that exceeds m directly intercepts the first m words, and text that is less than m words is filled with zero vectors. The pre-processed text is trained using the Word2Vec model, and then word vectors are obtained. The training method is Skip-Gram to obtain the n-dimensional word vector \( v_t \in \mathbb{R}^n \) \( (t = 1, 2, \ldots, m) \). The text word vector matrix \( V = [v_1, v_2, \ldots, v_m] \), where \( V \in \mathbb{R}^{m \times n} \).

3.2. **Conv-RBM**
In Conv-RBM, the input text word vector (yunnan, ruili, earthquake...) is represented as a matrix \( V = [v_1, v_2, \ldots, v_m] \), and the visible layer uses \( k \times n \) convolution kernel for convolution operations. The word vector features are obtained by sliding the convolution kernel on the word sequence. The equation is:

\[
y_j = f^\prime \left( w \ast V^\prime + b \right)
\]

Where the "\( \ast \)" operator indicates that the corresponding elements of the matrix are multiplied, \( V^\prime \) indicates \( k \times n \) matrix composed of word sequence vectors \( [v_1, v_2, \ldots, v_k] \) in \( V \). The \( b \) is an offset vector of \( n \) dimension, and \( f \) is a non-linear activation function. Use the obtained feature vector as the input of the next layer.

3.3. **RBMs**
RBM is an energy-based model. The energy function is defined as:
\begin{equation}
E(v, h \mid \theta) = - \left( \sum_{ij} w_{ij} v_i h_j + \sum_i b_i v_i + \sum_j c_j h_j \right) \tag{2}
\end{equation}

Where \(v\) represents the hidden layer, \(h\) represents the visualization layer, \(w\) represents the weight value, \(b\) is the bias of the visible layer, \(c\) is the bias of the hidden layer, and \(\theta\) represents the parameters \(w, c,\) and \(b\).

In a specific state:
\begin{equation}
P_{\theta}(v, h) = \frac{1}{Z(\theta)} e^{-E(v, h; \theta)} \tag{3}
\end{equation}

Where \(Z(\theta)\) is a normalization factor, which indicates the probability that \(v\) and \(h\) fit each other when \(\theta\) is determined.

Each RBM layer of the DBN uses a greedy algorithm during training. In order to determine the values of \(w, c\) and \(b\), these values are obtained by maximizing the function \(P_{\theta}(v, h)\), but maximizing this function is equivalent to maximizing its logarithmic value:
\begin{equation}
L(\theta) = \frac{1}{N} \sum_{n=1}^{N} \ln P_{\theta}(v^n) \tag{4}
\end{equation}

Therefore, based on the contrast divergence function, the following two formulas are repeatedly used:
\begin{equation}
P(h_j = 1 \mid v) = \sigma \left( \sum_{i=1}^{n} w_{ij} \times v_i + c_j \right) \tag{5}
\end{equation}
\begin{equation}
P(v_j = 1 \mid h) = \sigma \left( \sum_{i=1}^{n} w_{ij} \times h_i + b_j \right) \tag{6}
\end{equation}

Equations (4) and (5) are equivalent to forward propagation in neural networks. The vector features of the previous layer continuously repeat the process of formulas (4) and (5). The data is forward propagated, and \(v\) and \(h\) are reconstructed. The weights \(w\), offset values \(b\) and \(c\) are updated, and the status of each RBM layer is obtained.

### 3.4. Output Layer

The specific operation of the output layer is to use the result \(S\) of the previous layer through the final softmax layer. This layer uses softmax as the activation function to obtain the event type probability distribution \(P\):
\begin{equation}
P = \frac{\exp \left( S_{w_{ic}} + b_{ic} \right)}{\sum \exp \left( S_{w_{ic}} + b_{ic} \right)} \tag{7}
\end{equation}

Where \(w_{ic}\) represents the weight and \(b_{ic}\) represents the offset; \(P = \{p_1, p_2, \ldots, p_i, \ldots, p_c\}(i \in [1, c], c\) is the number of emergency types). The fine-tuning of back-propagation uses cross-entropy cost.
function as the loss function to obtain the loss value. Let the learning rate be $LR$, and the optimizer Adam is used for gradient descent to minimize the loss, so that the entire model training iterates to convergence:

$$loss = \text{average}\left( -\sum \text{label} \ast \text{tf.log} \left( p_i \right) \right)$$  \hspace{1cm} (8)$$

$$\text{AdamOptimizer} \left( LR \right) . \text{minimize} \left( loss \right)$$  \hspace{1cm} (9)$$

The verification and test phase uses the arg_max function to obtain the index value of the maximum probability in the output result vector $p_i$, so that it can match the actual label of the event type:

$$label = \text{arg\_max} \left( p_i \right)$$  \hspace{1cm} (10)$$

4. Experiment

In order to verify the effectiveness of the Conv-DBN model proposed by this paper in emergency event recognition, and recognition results are superior to other model methods, a comparative experiment is performed in this paper.

The corpus data uses the Chinese Emergency Corpus (CEC) set up by the Semantic Intelligence Laboratory of Shanghai University, and data is obtained through web crawlers. The corpus data comes from news reports of emergencies on the Internet and we-media information. These data contain five types of emergencies: earthquakes, fires, traffic accidents, terrorist attacks and food poisoning. There are 332 CEC in total and 28267 data obtained by the crawler. The 25530 in total after removing duplication. This paper conducts experiments on the two data sets, randomly selects the training set and the test set, and divides them according to 8:2.

The analysis method of the emergency recognition results used common evaluation standards, including Precision, Recall, and F Value. The multiple experiments are performed and compared with different models of other documents in emergencies recognition effect. The experimental comparison results in the CEC dataset are shown in Table 1.

| Model    | P/% | R/% | F/% |
|----------|-----|-----|-----|
| SVM [14] | 79.30 | 59.90 | 63.70 |
| CNN      | 72.73 | 64.00 | 68.09 |
| GRU      | 69.70 | 66.67 | 68.15 |
| BiGRU [15] | 71.10 | 69.00 | 70.00 |
| DBN      | 74.24 | 68.06 | 71.02 |
| Conv-DBN | 77.27 | 67.11 | 71.83 |

Because the CEC corpus is a structured xml format language, the pre-processed text features are still more obvious and easier to extract and train, which will lose generality and small amount CEC data may make model training insufficient or overfitting. Therefore, the experiment was carried out on the unstructured information data crawled from the Internet and comparative experiments are performed by using various models. The experimental comparison results are shown in Table 2.
Table 2. Experimental comparison results in we-media data.

| Model  | 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|
| DBN    | 74.13| 70.84| 72.45|
| Conv-DBN | 76.37| 72.21| 74.23|

It can be seen from the experimental results that the paper [14] in Table 1 and Table 2 use the traditional machine learning method, SVM (Support Vector Machine), to recognize emergencies, and achieve accuracy rates of 79.30% and 77.26% respectively. Though recognition effect is good, recall rate is very low. Other five groups of network models based on deep learning are significantly better than SVM in F value; F value of CNN is slightly higher than CNN, indicating that context information captured by GRU is more conducive to recognition of emergencies; the paper [15] in Table 1 uses the BiGRU model to achieve F value of 70.00%, compared with GRU, it has increased by nearly 2%, which indicates that the bidirectional recurrent neural network is better than unidirectional recurrent neural network in recognizing emergency by capturing the feature of context information. The DBN uses multiple layers of RBM to form a neural network and uses greedy algorithm to train model layer by layer. Trained model has achieved good results in the event recognition task; Conv-DBN incorporates a convolution operation in the visible layer, which improves the performance compared to DBN and effectively increases the precision and recall. This shows that convolution operation can extract information features based on DBN's function that re-represent information feature, so that it can improve effect of Chinese emergency recognition.

5. Conclusion

Compared with other methods, the Conv-DBN model proposed in this paper has achieved good results in accuracy and recall, and the F value is better than other methods. This shows that the improved deep belief network can extract local information features in input data when the convolution operation is integrated into the visible layer. The down-sampling is realized, which changes situation that original model is only re-represent the data features, and improve the recognition effect of Chinese emergencies.

Considering that the model lacks global context semantic features in feature extraction, and has poor interpretability for features with different degrees of importance. The next step is to try to extract richer contextual semantic features in combination with other network structures, enhance interpretability, and improve the overall performance of the model.

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).

References

[1] QIN Yanxia, ZHANG Min and ZHEND Dequan, “A survey on neural network-based methods for event extraction”, Intelligent Computer and Event Extraction, 2018, 8 (3): 1-10.

[2] Zhong Z, Jin L, Feng Z, “Multi-font printed Chinese character recognition using multi-pooling convolutional neural network,” International Conference on Document Analysis & Recognition, 2015.

[3] Valery Solovyev and Vladimir Ivanov, “Knowledge-Driven Event Extraction in Russian: Corpus-Based Linguistic Resources,” Computational Intelligence and Neuroscience, vol. 2016, Article ID 4183760, 11 pages, 2016.
[4] Syed A. Yusuf, David J. Brown, Alan Mackinnon, “Application of acoustic directional data for audio event recognition via HMM/CRF in perimeter surveillance systems,” Robotics and Autonomous Systems 72: 15-28, 2015.

[5] KUN LIU, ZHENSHI SUN, JUNFENG JIANG, et al., “A Combined Events Recognition Scheme Using Hybrid Features in Distributed Optical Fiber Vibration Sensing System,” IEEE Access, 2019.

[6] O. M. Ezeme, Q. H. Mahmoud and A. Azim, “DReAM: Deep Recursive Attentive Model for Anomaly Detection in Kernel Events,” in IEEE Access, vol. 7, pp. 18860-18870, 2019. doi: 10.1109/ACCESS.2019.2897122.

[7] ZHANG Ya-jun, LIU Zong-tian, ZHOU Wen, “Event Recognition Based on Deep Belief Network,” Acta Electronica Sinica, 45 (06): 1415-1423, 2017.

[8] Wan Jing, Li Haoming, Yan Huanchun, et al. “Relation extraction based on recurrent convolutional neural network”, Application Research of Computers, 2018, 1-6 [2019-09-27].

[9] ongyu 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.

[10] Tu Min, Liu Xiang, Liu Shuchun, “Natural Language Processing Core Technology and Algorithm with Python”, China Machine Press, 2018.

[11] 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.

[12] 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.

[13] Hinton G E, Salakhutdinov R R. Reducing the dimensionality of data with neural networks [J]. Science, 2006, 313 (5786): 504-507.

[14] XUAN Xiaoxing, LIAO Tao, GAO Beibei, “Automatic Extraction of Chinese Event Trigger Word,” Computer & Digital Engineering, (3): 457-461, 2015.

[15] MA Chenxi, CHEN Xingshu, WANG Wenxian et al., “Chinese Event Detection Based on Recurrent Neural Network,” Netinfo Security, (5): 75-81, 2018.