IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received August 12, 2019, accepted September 7, 2019, date of publication September 16, 2019, date of current version September 30, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2941653

Chinese Event Detection Based on Multi-Feature Fusion and BiLSTM

GUIXIAN XU[1,2], YUETING MENG[1], XIAOKAI ZHOU[1], ZIHENG YU[1], XU WU[1], AND LIJUN ZHANG[1]

[1]College of Information Engineering, Minzu University of China, Beijing 100081, China
[2]Department of Health Sciences Research, Mayo Clinic, Rochester, MN 55905, USA

Corresponding author: Guixian Xu (guixian_xu@muc.edu.cn)

This work was supported in part by the National Social Science Fund of China under Grant 19BGL241, in part by the Project of Humanities and Social Sciences, Ministry of Education of China, under Grant 18YJA740059, and in part by the China Scholarship Council under Grant (2018) 10038.

ABSTRACT With the rapid development of the Internet, the number of Internet users has grown rapidly, and the Internet has become more and more influential on people's lives. As a result, the amount of network text is increasing rapidly, and it is difficult to extract interested event information from it only by manual reading. Therefore, event extraction technique automatically extracting useful information from a large amount of unstructured texts becomes increasingly important. Event detection is the first step of event extraction task and plays a vital role in it. However, current event detection research lacks comprehensive consideration of the context of the trigger words. A Chinese event detection method based on multi-feature fusion and BiLSTM is proposed in this paper. The contextual information of word is divided into sentence-level and document-level in the method. The contextual information is captured based on BiLSTM model. At the same time, a word representation method suitable for trigger word classification tasks is proposed in this paper. The word representation incorporates semantic information, grammar information, and document-level context information of word. The word vectors in the sentence are sequentially inputted into BiLSTM model to obtain output vectors containing sentence-level contextual information. Finally, output vectors of BiLSTM are inputted into the Softmax classifier to realize the identification of the trigger words. The experimental results show that Chinese Event Detection Based on Multi-feature Fusion and BiLSTM method proposed in this paper has high accuracy.

INDEX TERMS Event detection, event extraction, word representation, BiLSTM.

I. INTRODUCTION

With the rapid growth of Internet data, there is an urgent need for a technology that can automatically find and extract useful information from massive data. Therefore, information extraction technology emerged and became a research focus in the field of natural language processing. The information extraction task is to automatically extract structured and semi-structured information from unstructured texts. The information can be queried and retrieved by the user, thus saving a lot of manpower and material resources.

Event extraction is one of the most important branches of information extraction. Events on the Internet related to public life, the state, and society are endless, such as fire incidents, earthquake events, terrorist attacks, and so on. Therefore, automatically and quickly extracting events and event-related information is a highly practical research. Currently, there is no uniform definition of event and event trigger word. Therefore, the definition of event and event trigger word in the literature [1] is cited in this paper. Event: Occurring at a specific time and place and showing several actions by participants. Event trigger word: A word that clearly indicates the event in the text, that is, the action element of the event. The event extraction mainly consists of two steps: the first step is the detection of the event, that is, the identification and classification of the trigger word; the second step is to analyze the identified event and extract the event elements. For example, in the sentence "北京时间2014年3月1日，一伙歹徒在昆明火车站持械伤人。(on March 1, 2014, a group of gangsters armed with injuries in Kunming Railway Station)", the event extraction technology determine the trigger word "伤(injury)" of the event

The associate editor coordinating the review of this manuscript and approving it for publication was Martin Gonzalez-Rodriguez.

TABLE 1. Example of event recognition.

Serial number	Sentence	Event type
Example 1	WTO 成立于 2000 年。(WTO was established in 2000.)	Commerce
Example 2	去年，国务院办公厅成立了国家科技领导小组。(Last year, the State Council General Office established the National Leading Group on Science and Technology.)	Politics
Example 3	它成立于 1994 年，现在是一支深受欢迎的乐队。(It was founded in 1994 and is now a popular band.)	Non-event
Example 4	......。虽然我很舍不得，但我知道我可能要离开了。1 因为作为一个销售经理，我这几个月的业务量远远没有达到标准。2......(....... Although I am reluctant, I know that I may have to leave.[1] Because as a sales manager, my business volume in the past few months is far from the standard.[2] ......)	Departure

and determine that this is a terrorist attack. Secondly, it is necessary to extract the event element information such as the time of the event "北京时间2014年3月1日(March 1, 2014)", the location "昆明火车站(Kunming Railway Station)", and the participants "一伙歹徒(a group of gangsters)".

Event detection is a key step in event extraction, which directly affects the effect of event element extraction. Currently, event detection faces certain challenges, mainly due to the complexity of the trigger words recognition tasks. The recognition of the trigger words is closely related to their context. The same trigger word in different contexts often represents different types of events. As shown in TABLE 1 of Example 1 - Example 3, all three sentences contain "成立(establishment)". Example 1 represents a "商业(commercial)" event, Example 2 represents a "政治(political)" event, and the main idea of example 3 is "它是一支乐队(it is a band)", it is not an event. In addition, the identification of the trigger words also needs to consider the semantic relationship between each sentence, that is, the context information at the document level. As shown in the first sentence of Example 4, the sentence contains the trigger word "离开(leave)", but it is difficult to derive the type of event. It can be seen from the semantics of the second sentence that the "离职(leave)" trigger word triggers the "地区(departure)" event.

In this paper, event detection is studied. The contextual features in the process of the trigger words recognition are fully considered, and a Chinese event detection method based on multi-feature fusion and BiLSTM is proposed. In this paper, the following contents will be divided into four aspects. Firstly, the relevant background of the event detection is expounded. Secondly, the Chinese event detection method based on multi-feature fusion and BiLSTM is described in detail. Thirdly, the experiments are carried out and the experimental results are analyzed and discussed. Finally, the proposed method is summarized and the next research direction is introduced.

II. BACKGROUND

A. EVENT DETECTION TECHNOLOGY

At present, the event detection research has got many achievements. Research methods are broadly divided into two categories: pattern-based matching methods and machine learning methods.

In the early studies of event detection, the methods based on pattern matching were widely used. The pattern-based matching method refers to discover trigger words according to some predefined patterns. Its core lies in the construction of patterns. Yangarber [2] proposed a seed-based ExDisco learning system. The system is based on artificially constructed seed templates and learns new templates by iteration. Wu *et al.* [3] constructed extraction rules by using manually determined sentence templates. The rule is used to extract event information from the processed text to populate the sentence template. Jiang [4] proposed a domain-independent event extraction model GenPAM. This method can automatically learn the template and reduce manual participation. Liang *et al.* [5] proposed a framework-based information extraction model and established a unified catastrophic event framework. This model uses the framework's inheritance-inductive features to simplify the system implementation process and summarize event information. Feng [6] used a predefined event framework to extract information from news elements. Pattern-based matching method relies on specific areas. It requires domain experts to build a large-scale knowledge base, which is labor intensive and the system is less portable.

With the construction of corpora and the continuous enrichment of Internet text resources, machine learning has become the mainstream research method of event detection. Using machine learning to identify events is to transform event recognition into trigger words classification. Its core is in the representation of the structure and features of the classifier. Li *et al.* [7] adopted a joint event extraction model of structured perceptron [8], which classifies the trigger word recognition and event elements as a whole sequence annotation task. Experiments proved that the joint model solved the problem of error propagation to a certain extent. Nguyen and Grishman [9] used CNN [10] model for event detection, which can automatically learn valid feature representations from trained word vectors, locations, and entity types. Experiments showed that compared with feature-based engineering, the method has improved trigger word classification and domain adaptability. Hou *et al.* [11] used the

LDA model to cluster words, which was based on the over-fitting problem of morphological features. Considering the inconsistency between the Chinese automatic segmentation and the triggering word boundary of the annotation, a trigger word recognition method based on the CRFs model was proposed. Chen *et al.* [12] introduced word vectors and convolutional neural networks into English event detection, which can automatically learn vocabulary level and sentence level features. In recent years, LSTM model has been widely used in event detection. Zeng *et al.* [13] proposed a Convolution-BiLSTM neural network model to detect Chinese events. Meanwhile, word-based vector construction method and character-based vector construction method were compared. Experiments showed that character-based vector construction method had better effect. Feng *et al.* [14] proposed an event detection method that did not depend on any natural language processing (NLP) tools. They incorporated both Bi-LSTM and convolutional neural networks to capture sequence and chunk information from specific contexts for event detection. Hong *et al.* [15] proposed a deep convolutional neural networks (DCNN) combined with long-short term memory (LSTM) to detect emergency event in Uyghur text. DCNN is used to extract the high-level local features of the event sentences, and LSTM is used to capture the sentence relations in the event. Then, a softmax classfier is trained to accomplish the event detection task. Hou and Ji [16] regarded the part of speech and the descriptive information of named entities of medical texts as additional attributes. Then, bidirectional LSTM neural network is utilized to learn the hidden feature representations. The method performed well in the task of clinic event detection from 2016 SemEval.

The current event detection task has achieved great results. However, related research lacks comprehensive consideration of contextual information. Based on this, the contextual information is divided into sentence-level context information and document-level context information in this paper. BiLSTM is used to extract context information.

B. WORD REPRESENTATION

In the field of natural language processing, words contained in sentences or documents are often used as features [17]–[20]. Therefore, learning a low-dimensional, real-numbered, non-sparse word representation method for a word is a critical step. Currently, there are two kinds of vectorized representations of words: One-hot representation and Distributed representation.

One-hot representation indicates that the length of the vector is the size of the dictionary. One component of the vector can be 1, which corresponds to the position of the word in the dictionary, and the others are 0. The dimension of this representation is larger and it does not describe well the similarities between words. To compensate for the flaws in the One-hot representation, Hinton [21] proposed the concept of Distributed representation. The distributed representation maps each word in the language into a fixed-length, low-dimensional real vector. All of these vectors form word vector space and each vector can be considered a point in the space. The advantage of the distributed representation is that the vector dimension is low and it can well characterize the semantic similarity between words. In 2013, Mikolov *et al.* [22] proposed the Word2vec technology, which greatly improved the training speed of distributed word vectors. It contains two new log-linear models: Continuous Bag-of-Words (CBOW) and Continuous Skip-gram (SG).

At present, word vectors have been applied in large scale in the field of event extraction. For example, Hou and Ji [16] introduced the word vector into medical event detection. They combined the distributed word vector, the part-of-speech feature POS of the medical text and the named entity feature NE as the input of the model. Wu and Zhang [23] used binary representation model to construct word vectors based on part of speech, dependencies and distance from core words. The BiLSTM-CRF model was used to implement Chinese event extraction and achieved good results. Zhang *et al.* [24] selected part of speech, dependency syntax, length of words, position of words, distance from core words, and word frequency as characteristics. And corresponding feature representation rules are formulated to convert words into word vectors. The deep semantic information of the word is then extracted through the deep belief network, and the event is identified by the Back-Propagation (BP) neural network.

However, the word vector used in existing event detection research lacks comprehensive consideration of semantic information, grammar information, and context information. Aiming at this problem, a word vector representation method suitable for trigger word classification task is proposed, which is based on Word2vec technology. The word representation method incorporates the semantic information, grammar information and context information of the words.

III. RESEARCH METHODS

A. LSTM

Traditional neural network models are less effective in dealing with sequence problems because they cannot describe the correlation between the input sequence before and after. RNN can effectively learn the sequence characteristics of data dynamically and has certain memory capabilities. A typical RNN structure is shown in FIGURE 1.

U is the weight from the input layer to the hidden layer, W is the weight of the hidden layer connected to itself, and V is the weight from the hidden layer to the output layer. The calculation process of RNN is as follows:

(1) At time t, input x_t to the hidden layer;

(2) s_t is the output of the t$^{\text{th}}$ step of the hidden layer, s_t is obtained from the output x_t of the current input layer and the state s_{t-1} of the hidden layer at the previous moment, $s_t = f(U^* x_t + w^* s_{t-1})$, Where f generally takes a nonlinear function such as tanh or ReLU;

(3) Finally, the output o_t is given, $o_t = soft\max(v^* s_t)\circ$

In theory, the neural network model can process infinitely long sequences. However, in the practical application,

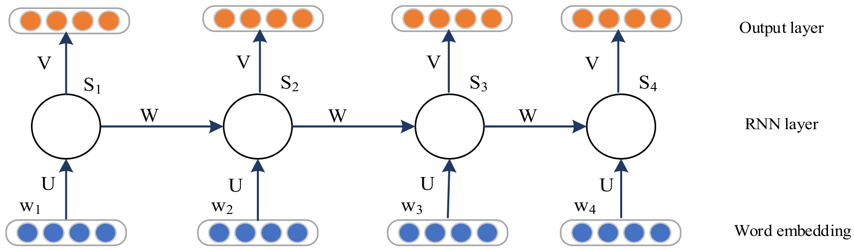

FIGURE 1. RNN model.

the problem of gradient explosion and gradient disappearing is easy to occur when training RNN. It makes the RNN unable to cope with the effects of long distances, resulting in an unsatisfactory result. A long-term and short-term memory network appears for the problem of the disappearance of the gradient in the basic RNN model.

The hidden state at each moment of the long-and-short memory model is calculated by the input gate i (Input gate), the output gate o (Output gate), the forget gate f (Forget gate), and the internal memory unit (Memorycell). The input, pass and forgetting operations of the internal memory are selected by the three thresholds i, o, and f. The LSTM network structure is shown in the FIGURE 2 [25].

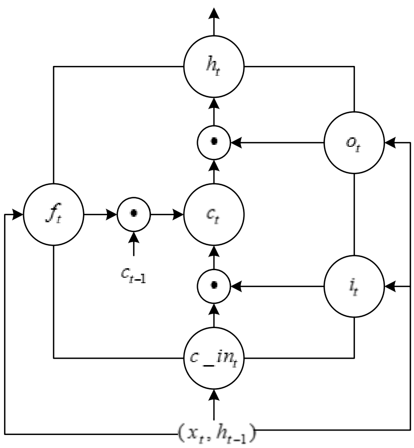

FIGURE 2. LSTM network structure.

Forget gate determines what information is discarded from the Memorycell at the previous moment. The input is h_{t-1} and x_t, and the output value is between 0 and 1. The calculation is as follows:

$$f_t = \sigma(W_f x_t + U_f h_{t-1} + V_f c_{t-1} + b_f) \tag{1}$$

Among them, h_{t-1} and x_t are the inputs of LSTM unit. W_f is the connecting weight of x_t and forget gate f. U_f is the connecting weight of h_{t-1} and forget gate f. c_{t-1} is the state of Memorycell at the last moment. V_f is the connecting weight of c_{t-1} and forget gate f. b_f is the bias term. $\sigma(\cdot)$ is sigmoid activation function.

Input gate determines which information is updated in the Memorycell at the current time. After two non-linear transformations, the content to be updated is selected, and then the Memorycell is updated. The calculation is as follows:

$$\begin{cases} i_t = \sigma(W_i x_t + U_i h_{t-1} + V_i c_{t-1} + b_i) \\ c\_in_t = \tanh(W_c x_t + U_c h_{t-1} + V_c c_{t-1} + b_c) \\ c_t = f_t \cdot c_{t-1} + i_t \cdot c\_in_t \end{cases} \tag{2}$$

Among them, W_i is the connecting weight of x_t and i_t. U_i is the connecting weight of h_{t-1} and i_t. V_i is the connecting weight of c_{t-1} and i_t. W_c. is the connecting weight of x_t and c_in_t. U_c is the connecting weight of c_in_t and h_{t-1}. tanh(.) is tanh activation function. f_t and i_t refer to weights of c_{t-1} and c_in_t. b_i and b_c are the bias terms.

Output gate determines the output value of the LSTM unit. The output is first determined, and then the final output is obtained by nonlinear transformation. The calculation is as follows:

$$\begin{cases} o_t = \sigma(W_o x_t + U_o h_{t-1} + V_o c_{t-1} + b_o) \\ h_t = o_t \cdot \tanh(c_t) \end{cases} \tag{3}$$

Among them, W_o is the connecting weight of x_t and o_t. U_o is the connecting weight of h_{t-1} and o_t.V_o is the connecting weight of c_{t-1} and o_t. b_o is the bias term.

B. CONSTRUCTION OF PART-OF-SPEECH, NAMED ENTITY, AND DEPENDENT GRAMMAR FEATURES

The LTP [26] tool of the Harbin Institute of Technology Semantic Lab is used in this paper to construct the features of part of speech, named entities, and dependent grammars, as follows:

Part of speech features: part of speech is the most basic grammatical feature of words. In sentences, grammatical components have strong restrictions on part of speech. For example, nouns and pronouns can act as subject components in sentences. Therefore, part of speech is used as an expression of the semantic information of the text and a basic understanding of human beings in the recognition of events. In the literature [24], the parts of speech of CEC are counted. It is found that the concentration of part of speech distribution of trigger words is very high, 84% of which are verbs and 14% are nouns. The corpus is preprocessed, and the word sequence obtained by the preprocessing is subjected to part-of-speech tagging. The part-of-speech sequence of the resulting sentence is shown in FIGURE 3. In the figure,

FIGURE 3. Diagram of part of speech analysis.

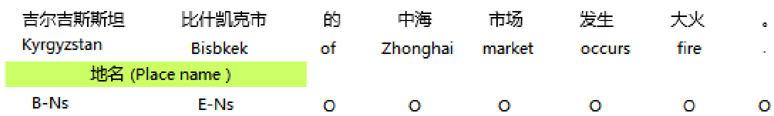

FIGURE 4. Diagram of named entity recognition.

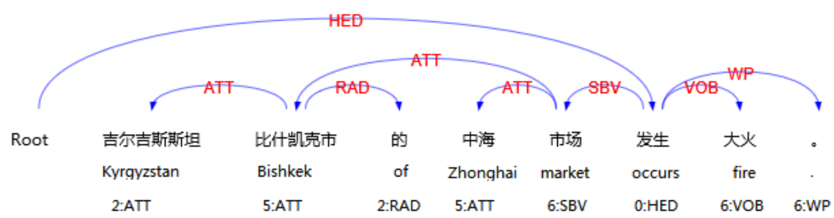

FIGURE 5. Diagram of dependency grammar role recognition.

the first line is words, the second line is English translation of words, and the third line is the part of speech.

863 part-of-speech tag sets is used in LTP, and there are 28 kinds of part-of-speech features, including adjectives, verbs, nouns, etc., which can construct a part-of-speech dictionary of length 28. Each word corresponds to a vector of length 28. Only one component of the vector can be 1, which corresponds to the position of the word in the dictionary of the part of speech, and the others are 0.

Named entity features: Named entity recognition is a very basic task in NLP. It refers to the identification of named referential items from the text, thus paving the way for tasks such as subsequent information extraction. The diagram of named entity recognition is shown in FIGURE 4. In FIGURE 4, the first line is words, the second line is English translation of words, and the third line is the named entities.

The LTP tool labels the entities in the word sequence and uses O-S-B-I-E annotation mode to identify the names of people, places, and institutions in the sentence. Each word corresponds to a vector of length 13. Only one component of the vector can be 1, which corresponds to the position of the word in the dictionary of name entity, and the others are 0.

Dependent grammar role features: Dependent grammar considers the verb in the sentence to be the center of other components. However, it is not subject to any other components, and all the dominating components are subordinate to their dominator in a certain dependency relationship. Dependent grammar directly describes the semantic role relationship between words and words, with very high semantic expression. Li *et al.* [7] have shown that dependent grammatical features play an important role in event extraction tasks. In this paper, the LTP tool is used to analyze the dependency relationship of a sentence, and the dependent grammar role of the word in the sentence is obtained, as shown in FIGURE 5. In FIGURE 5, the first line is words, the second line is English translation of words, and the third line is the dependent grammar roles.

There are 15 dependent grammar roles for LTP annotation. The words are converted into dependent grammar vectors, and each word corresponds to a 15-dimensional dependent grammar vector. Only one component of the vector can be 1, which corresponds to the position of the word in the dictionary of dependent grammar role and the others are 0.

An example of feature representation is shown in TABLE 2.

In TABLE 2, POS denotes Part of speech, DGR denotes Dependency grammar role, and NE denotes named entity.

TABLE 2. Example of feature representation.

Word	Type of Feature	Feature Vector Value
吉尔吉斯斯坦 (Kyrgyzstan)	POS	0000 0000 0000 0000 1000 0000 0000
	DGR	0000 0100 0000 000
	NE	0000 0000 0100 0
发生 (occurs)	POS	0000 0000 0000 0000 0000 0000 1000
	DGR	0000 0000 0000 001
	NE	0000 0000 0000 0
大火 (fire)	POS	0000 0000 0001 0000 0000 0000 0000
	DGR	0100 0000 0000 000
	NE	0000 0000 0000 0

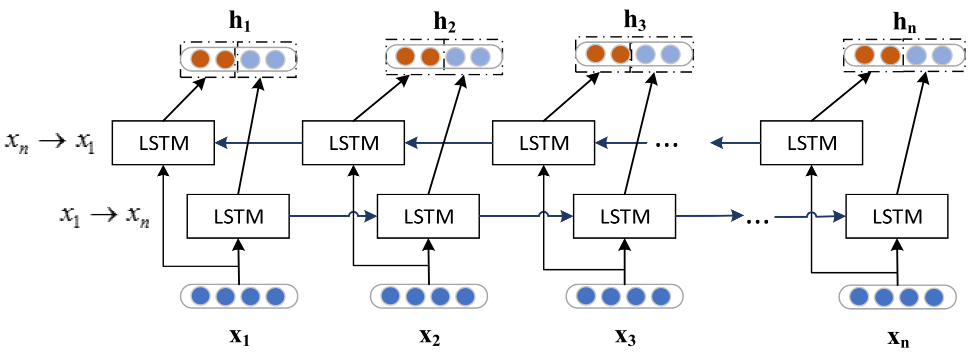

FIGURE 6. Basic structure of the BiLSTM network.

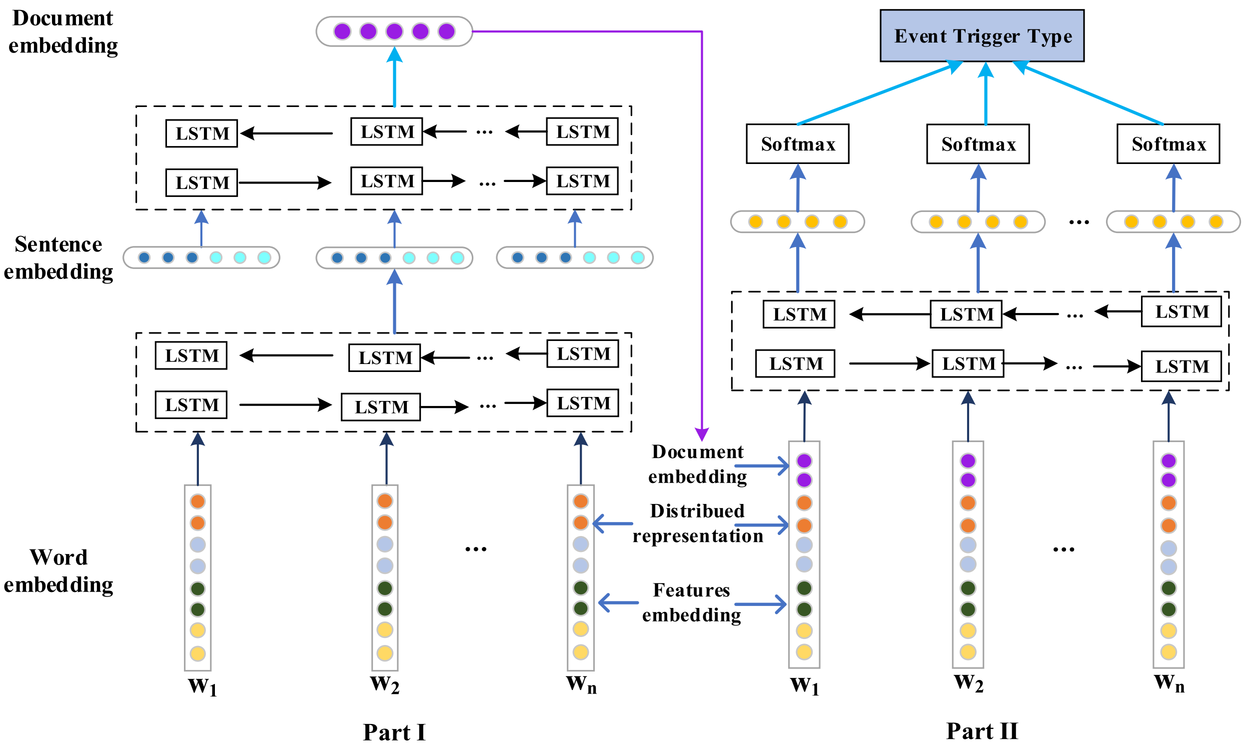

FIGURE 7. Chinese event detection model based on multi-feature fusion and BiLSTM.

C. CHINESE EVENT DETECTION BASED ON MULTI-FEATURE FUSION AND BILSTM

When using the traditional cyclic neural network model for training, the gradient will gradually decrease until the disappearance in the back-propagation process, which forms the problem of the gradient disappearing. Therefore, the length of the sequence is limited to a short range. LSTM model mentioned in Section A. overcomes the problem of gradient disappearance of RNN by introducing a threshold mechanism. In order to contain context information at every moment, BiLSTM is used to transfer semantic information between word vectors in this paper.

The basic structure of the BiLSTM network is shown in FIGURE 6. In FIGURE 6, x_i denotes i^{th} word in sentence, $x_1 \rightarrow x_n$ denotes forward propagation, and $x_n \rightarrow x_1$ denotes backward propagation. When the words in the sentence are input into the BiLSTM, the vector h_i concated by $\vec{h}_i$ (the output of forward LSTM) and $\overleftarrow{h}_i$ (the output of reverse LSTM) will be obtained. LSTM model can only be transmitted in order from front to back. However, BiLSTM can contain context information and understand the semantic information of the current word.

$$h_i = \vec{h}_i || \overleftarrow{h}_i \tag{4}$$

The schematic diagram of the Chinese event detection model proposed in this paper is shown in FIGURE 7.

The model is divided into Part I and Part II. Part I is the learning process of doucument embedding, and Part II is the process of Chinese event detection. In FIGURE 7, The Word embedding layer consists of Features embedding,

```
<Paragraph>
    <Sentence...>
    <Sentence>
        <Event eid="e16">
            <Time type="timeInterval" tid="t16">截至9日上午10时(by 10 a.m. on Day 9)</Time>,
            <Location lid="l16">该地区(the area)</Location>又发生了近1500次
            (happened nearly 1500 times)
            <Denoter type="emergency" did="d16">余震(aftershock)</Denoter>。
        </Event>
    </Sentence>
</Paragraph>
```

FIGURE 8. Diagram of CEC corpus.

TABLE 3. Types of events in CEC.

Event type	Trigger word	Quantity
action	相撞、超载、带走、讯问、侧翻...... (Collide, overload, take away, interrogate, rollover......)	598
emergency	交通事故、事故、追尾、车祸、起火...... (Traffic accident, accident, rear-end collision, car accident, fire......)	106
movement	赶往、前往、赶赴、赶到、送往...... (Rush, go, rush, arrive, send......)	129
operation	封锁、疏散、调查、启动、会议...... (Blockade, evacuate, investigate, start, meeting......)	416
perception	看见、谴责、慰问、认为、希望...... (See, condemn, condolence, think, hope......)	103
stateChange	死亡、受伤、失踪、损失、倒塌...... (Death, injury, missing, lose, collapse......）	224
statement	报道、提出、说、告诉、声明...... (Report, propose, speak, tell, declare......）	118

Distributed representation, and Document embedding. Distributed representation is distributed word vector trained by Word2vec. Features embedding refers to the part-of-speech vector, named entity vector, and dependency vector described in Section B. Document embedding refers to a document-level context feature vector. In part I, the word vectors consisting of Features embeddings and Distributed representations are inputted into the underlying BiLSTM model. Then, the sentence embedding is obtained. Sentence embeddings of all sentences in the document are inputted into the upper BiLSTM model, and the document-level context features are obtained. In part II, the multi-feature fusion word vectors containing the semantic information, the grammar information, and the context features are inputted into the BiLSTM layer, so that the word vectors that fully capture the sentence-level context information could be obtained. Then, the output word vectors of BiLSTM layer are inputted into the Softmax layer. Through the mapping of the Softmax function, the class probability distribution of the words is obtained, thereby realizing the recognition and classification of the trigger words. In order to prevent over-fitting in the training process, the Dropout mechanism is introduced. Partially trained parameters are discarded during each iteration, so that the updated weights no longer depend on some inherent features, which increases the applicability of the model. The Dropout discard rate is set to 0.5.

IV. EXPERIMENT

A. DATA SET CONSTRUCTION

The dataset of this paper is CEC corpus [27] constructed by Shanghai University (Semantic Intelligence Lab). Although the size of the CEC corpus is small, the annotation of events and event elements is comprehensive. CEC uses the XML language as the annotation format, which contains the six most important data structures (tags): Event, Trigger word, Time, Location, Participant, and Object. Among them, Event is used to describe a event. Trigger word, Time, Location, Participant, and Object are used to describe the indicators and elements of the event. The schematic diagram of the corpus is shown in FIGURE 8:

The CEC corpus divides the event type into seven major categories: action, stateChange, emergency, movement, operation, statement, and perception. The trigger words that are common in each category are shown in TABLE 3.

CEC consists of 332 documents. In this paper, the training set and the test set are divided according to the ratio of 4:1, 266 and 66, respectively. Firstly, the CEC original XML format documents are processed to the TXT format documents. Then, the word segmentation system of LTP is used to segment the words. The original tagged trigger words in XML format documents are compared with the words after segmentation in TXT format documents. As a result, the words could be labelled as categories.

TABLE 4. Preprocessed CEC corpus.

Word	Type	Word	Type
截至(up to)	0	发生(happen)	0
9日(9th)	0	了(auxiliary verb)	0
上午(morning)	0	近(approach)	0
10时(10:00)	0	1500	0
，	0	次(times)	0
该(this)	0	余震(aftershock)	emergency
地区(region)	0	。	0

The schematic diagram of the preprocessed corpus is shown in TABLE 4.

In TABLE 4, each line is composed of a word and its type. The types of words are divided into eight types: seven types of trigger words (action, emergency, movement, operation, perception, stateChange, statement) and non-trigger words (0).

The preprocessed corpus is inputted into the analysis module of LTP to get the parts of speech, named entities and dependent grammar roles of words. As for the word "地区 (region)" in TABLE 4, the results of analysis are "n (general noun)", "O (not a named entity)" and "SBV (subject-verb)", which would be used as the basis of feature vector generation.

The 300-dimensional word vector trained by the Skip-Gram model in Word2vec provided by DataScience (https://mlln.cn) is used as distributed word vector. The specific parameter settings are shown in TABLE 5.

B. EVALUATION INDICATORS

In this paper, the accuracy of Precision (P), Recall (R) and F1 are used to evaluate the results. The calculation formula is as follows:

$$P = \frac{TP}{TP + FP} \tag{5}$$

$$R = \frac{TP}{TP + FN} \tag{6}$$

$$F1 = \frac{2 \times P \times R}{P + R} \tag{7}$$

TP is the number of trigger words correctly identified. FP is the number of triggers identified by mistake. FN is the number of unrecognized trigger words. P is the proportion of correctly recognized trigger words to recognized trigger words, R is the proportion of correctly recognized trigger words to total trigger words. F1 is the harmonic mean of P and R. The trigger word recognition correctly indicates that the trigger word is detected and the type is correctly identified.

C. EXPERIMENT

1) PARAMETER SETTINGS

The hyperparameters of the Chinese event detection model proposed in this paper mainly include Epochs, Learning rate, Hidden Size, Embedding Number and so on. The following is a study of the parameters that are optimal for model classification.

a: EMBEDDING NUMBER

Embedding Number is the number of vectors in the input layer of the model. If the Embedding Size is too large, the input layer will be filled with too many zero vectors. If the Embedding Number is too small, too many words will be discarded. The sentence length of the dataset is counted, and its distribution is shown in FIGURE 9.

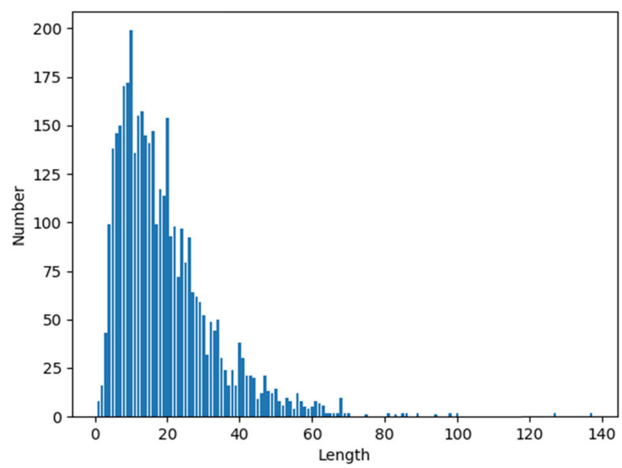

FIGURE 9. Length distribution of data set sentences.

As can be seen from the FIGURE 9, the number of sentences less than 100 is the largest. The essence of event detection is to detect and classify trigger words. The words in the sentence should be retained as much as possible in the experiment to improve the reliability of the experiment. Therefore, the value of the Embedding Number is set to 100.

b: EPOCHS

Epochs is the number of iterations of a complete training set. As Epochs increased, the model learned more. However, if the number of epochs is too large, over-fitting problem is easily generated, and the generalization ability of the model is degraded. Therefore, it is important to choose the right Epochs. FIGURE 10 is a model classification effect by using different Epochs.

TABLE 5. Training parameters for word2vec.

Window size	Dynamic window	Sub-sampling	Low-Frequency word	Iteration	Negative Sampling*
5	Yes	1e-5	10	5	5

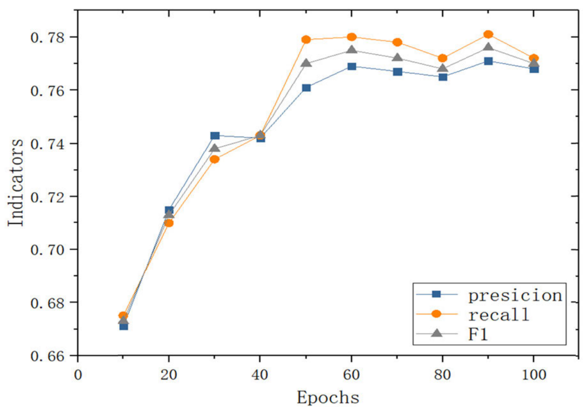

FIGURE 10. **Relationship between epochs and indicators.**

It can be seen from FIGURE 10 that with the increase of Epochs, the classification precision, recall and F1 of the model gradually increases. It tends to be stable when Epochs is 60, achieving a satisfactory effect. Therefore, Epoch is set to 60.

c: LEARNING RATE

The appropriate choice of learning rate is important for the optimization of weights and offsets. If the learning rate is too large, it is easy to exceed the extreme point, making the system unstable; if the learning rate is too small, the training time is too long. FIGURE 11 is a model classification effect by using different learning rates.

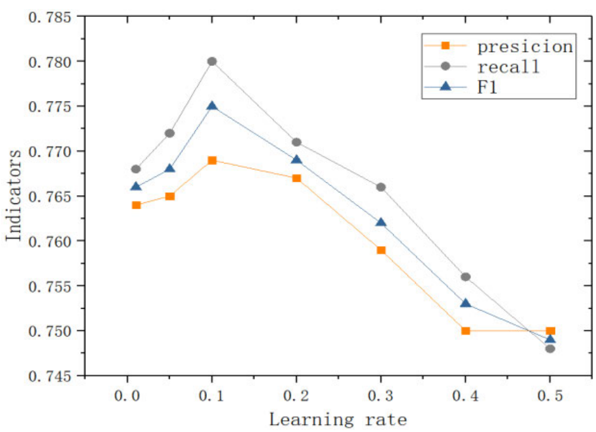

FIGURE 11. **Relationship between learning rate and indicators.**

It can be seen from FIGURE 11 that as the learning rate changes, the F1 value of the model remains around 75%-78%, reaching a maximum at 0.1. Therefore, the learning rate is taken as 0.1.

d: HIDDEN SIZE

The number of hidden layer nodes has a certain influence on the complexity and effect of the model. If the number of nodes is too small, the network learning ability will be very limited; if the number of nodes is too large, it will not only increase the complexity of the network structure, but also make it easier to fall into local minimum points during the training process, and the network learning speed will decrease. FIGURE 12 is a model classification effect by using different hidden size.

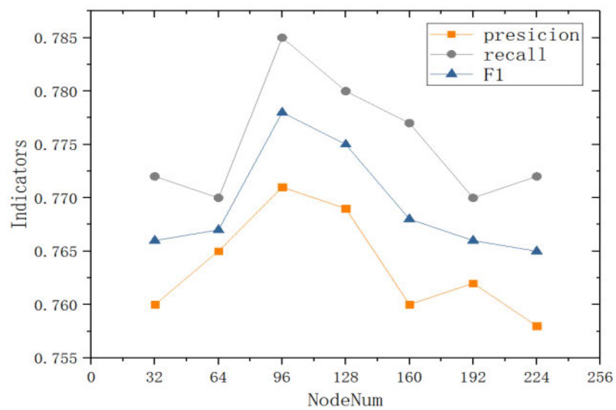

FIGURE 12. **Relationship between hidden size and each indicator.**

According to the experimental results, it can be seen that when the number of nodes increases from 32 to 96, the F1 value shows an upward trend. When the number of nodes exceeds 96, the F1 value begins to show a downward trend. Therefore, the Hidden Size is chosen to be 96.

2) COMPARATIVE EXPERIMENT

Word vector with multi-feature fusion is proposed. It is composed of distributed word vector, part of speech vector, named entity vector, dependent grammar role vector, and document-level context vector. In order to verify the validity of the multi-feature fusion word vector proposed in this paper, the same data set CEC is adopted. Word representation containing different features is inputted into the Chinese event detection model shown in FIGURE 6. The effects of event detection are compared by experiments. The parameter settings of the Chinese event detection model are shown in TABLE 6.

TABLE 6. **List of hyparameters.**

Hyperparameter	Value
Epochs	60
Learning rate	0.1
Optimization function	Mini-Batch Gradient Descent
loss function	Cross Entropy Loss Function
Embedding Number	100
Dropout	0.5
BatchSize	64
NodeNum	96

The Word2vec trained distributed word vector that incorporates the semantic features of the word is recorded as C1. Then, based on the previous feature, Part of Speech C2, Named Entity C3, Dependency Relation C4, and Document-Level Context C5 are added. Ten experiments were repeated, and the F1 values of each experiment are shown in FIGURE 13.

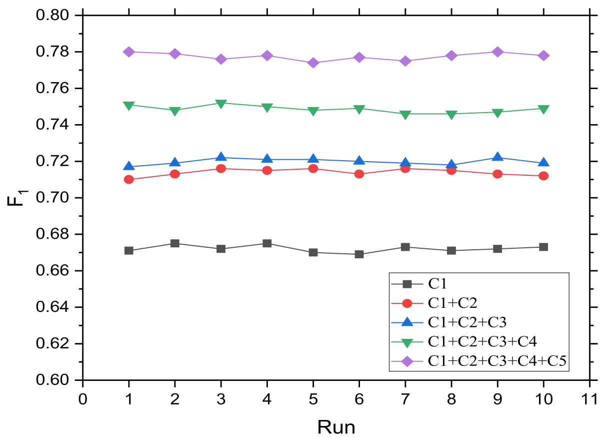

FIGURE 13. **Experimental results of word vectors incorporating different features.**

As can be seen from FIGURE 13, the F1 values of the various word vector representation methods are relatively stable. Except for the distributed word vector containing only the semantic information of the words, the F1 values after the introduction of each feature are all above 0.70. However, the F1 value of the distributed word vector is maintained at around 0.67. After introducing the part of speech, named entity, and dependent semantic role features, the F1 value can reach around 0.75. After introducing all the features, the F1 value is further increased to around 0.78. At the same time, the experimental results also show that the features introduced in the word representation layer have a good recognition effect. The introduced features are all suitable for Chinese event detection tasks. After each feature is introduced, the F1 value is raised.

The average accuracy, recall, and F1 values of the experiments is shown in TABLE 7 and FIGURE 14.

TABLE 7. **Comparison of event detection effects of different word representation methods.**

Word vector representation	Precision	Recall	F1
Base Feature	0.694	0.661	0.672
Part of Speech	0.715	0.710	0.713
Named Entity	0.725	0.709	0.717
Dependency Relation	0.754	0.743	0.749
Document-Level Context	0.783	0.772	0.778

It can be seen from the experimental results that the word vector representation method of multi-feature fusion proposed in this paper has achieved good results. The precision, recall, and F1 values reached 0.783, 0.772, and 0.778, respectively. Compared with the distributed word vector (C1) of Word2vec training commonly used in the field of natural language processing, the representation method proposed in this paper improves the precision of event recognition

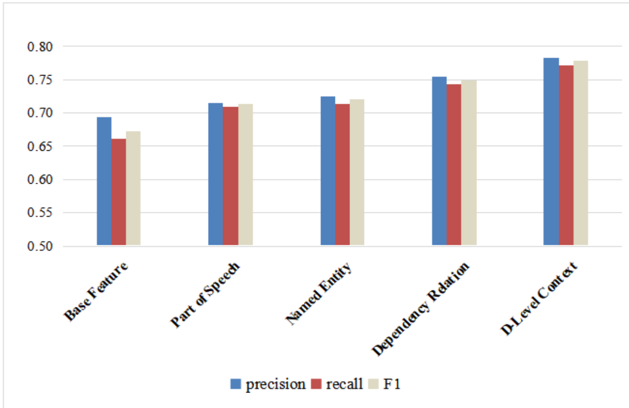

FIGURE 14. **Comparison of event detection effects of different word representation methods.**

by 0.089, the recall by 0.111, and the F1 value by 0.106. The experimental results show that the contribution of each feature to the recognition performance improvement is not the same. The addition of part-of-speech features, dependent syntactic features and contextual features makes the recognition effect more obvious. F1 values increased by 0.041, 0.032, and 0.029, respectively. The addition of named entity features has a limited improvement in recognition. F1 value increased by 0.004. In summary, the experiment proves the validity of each feature introduced in the word vector representation layer. Part of speech, Named entity, Dependent grammar role, and Document-Level Context all play an important role in the recognition and classification of the trigger word.

In order to further prove the validity of the Chinese event detection method proposed in this paper, the proposed method is compared with the existing Chinese event detection method based on the same corpus (CEC corpus). The comparison methods are as follows:

Extended Trigger lexicon (ET lexion): The extended trigger vocabulary for Chinese event detection is used in the literature [28]. Through the research and statistics of the CEC corpus, the trigger words with higher frequency events in the corpus are sorted out as the original trigger vocabulary. The synonym table of Harbin Institute of Technology is used in the paper. The original trigger vocabulary is expanded by the word triggering clustering method. Finally, event detection is implemented by matching the extended trigger vocabulary.

BiGRU: The word vector of the input layer of the literature [29] consists of distributed word vector, named entity vector, part of speech vector, and dependent grammar role vector. BiGRU is selected as the presentation layer. Event detection is implemented by using the Softmax classifier for trigger word class identification on the CEC dataset.

SVM: A word vector representation method that combines distributed word vector, named entity vector, part of speech vector, and dependent grammar role vector is used in feature representation. The SVM classifier is trained on the CEC data set and the classifier is used to identify the trigger words.

The method proposed in this paper: The word vector of the multi-feature fusion proposed in this paper is input into the BiLSTM model to obtain output vector containing sentence-level context information. Then, output vectors of BiLSTM are input into the Softmax classifier to realize the recognition of the trigger words.

LSTM: LSTM model is replaced to BiLSTM model in this paper. Other structures and parameters are consistent with the event detection methods in this paper.

The comparison results of the various methods are shown in TABLE 8 and FIGURE 15.

TABLE 8. Results of each chinese event detection method.

Method	Precision	Recall	F1
ET lexion	0.732	0.674	0.697
BiGRU	0.711	0.690	0.700
SVM	0.712	0.659	0.684
LSTM	0.756	0.712	0.733
Proposed method	0.783	0.772	0.778

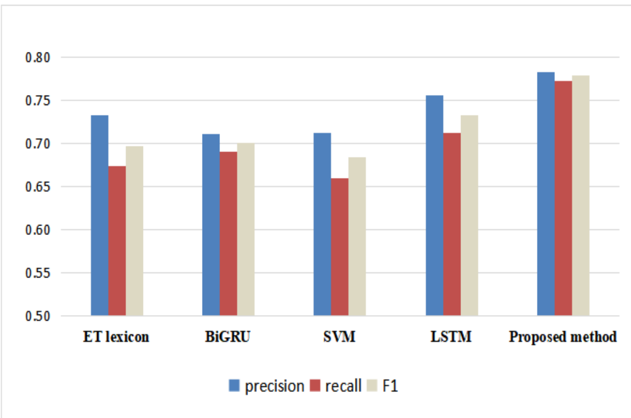

FIGURE 15. Comparison of the effects of various chinese event detection methods.

It can be seen from the experimental results that the event detection method proposed in this paper is better than other methods in precision, recall and F1, reaching 0.783, 0.772 and 0.778 respectively. In contrast, the effect of triggering dictionary matching and SVM is worse than the other three methods. The F1 values of both methods are below 0.70. This is because the trigger dictionary matching method is extremely dependent on the built trigger dictionary. If the trigger dictionary is too large, a large number of non-trigger words will be introduced, resulting in a low precision. If the trigger dictionary is too small, it will result in the omission of a large number of trigger words, resulting in a low recall. More importantly, some trigger words have a polysemy, and the types of triggers are often different with different contexts. The trigger dictionary matching method and the SVM method didn't consider the contextual information between words, so this problem can't be solved. For example, in CEC corpus, the sentence "云南省盈江县 20 日早晨发生地震 (An earthquake occurred on the morning of 20thin Yingjiang County, Yunnan Province)", "地震 (Earthquake)" as a trigger word triggered "Emergency" events. However, In the another sentence "地震是地壳快速释放能量过程中造成的震动。(Earthquakes are shocks caused by the rapid release of energy from the earth's crust.) "," 地震 (Earthquake)" is not a trigger word. The sentence is only a general definition of an earthquake, and does not mean that an event has occurred. The method based on SVM and trigger dictionary cannot distinguish the different types of trigger words in different contexts, but the proposed method can distinguish it. Compared with the BiGRU and unidirectional LSTM methods, the F1 value of the event detection method in this paper is increased by 0.078 and 0.045, respectively. By analyzing the reason, BiGRU method can effectively transfer contextual information between words by introducing GRU units. However, the contextual information is only sentence-level, ignoring the context information at the document-level. For example, in the CEC corpus, the sentence "2002年12月5日上午, 江苏镇江丹阳市皇塘镇吴塘初级中学发生严重事故。(On the morning of December 5, 2002, a serious accident occurred in Wutang Junior Middle School, Huangtang Town, Zhenjiang City, Jiangsu Province.)" The sentence contains the trigger word "发生(occurred)", but the type of trigger word is difficult to determine. According to the following sentence "近百名学生吃了学校食堂早餐后,集体食物中毒。(Nearly 100 students suffered from collective food poisoning after eating breakfast in the school canteen).", it can be concluded that the trigger word "发生(occurred)" triggered the "Emergence" event. BiGRU method cannot recognize the type of trigger words. The proposed method combines the context information at the document level and improves the accuracy of trigger word recognition. In LSTM, information is only transmitted in the order from front to back, so it is not possible to include contextual information at every moment. The Chinese event detection method based on multi-feature fusion and BiLSTM proposed in this paper uses the word vector suitable for the trigger word classification task as input. Contextual information at the sentence-level and at the document-level is obtained through BiLSTM model. The effect of event detection is better than other methods.

V. CONCLUSION

A Chinese event detection method based on multi-feature fusion and BiLSTM is proposed in this paper. Since current event detection research lacks comprehensive consideration of the contextual information, BiLSTM model is used to capture sentence-level and document-level contextual information in this paper. At the same time, a word vector representation method suitable for trigger word classification tasks is proposed in the paper. The word representation incorporates semantic information, grammar information, and document-level contextual information of the word. The word vectors in the sentence is sequentially inputted into the

BiLSTM model to obtain output vectors containing sentence-level contextual information. Finally, all output vectors of BiLSTM are inputted into the Softmax classifier to realize the identification of the trigger words. Thus, the detection of the event is completed. In this paper, the experiment of word representation with different features is used to prove the validity of the proposed word vector representation method of multi-feature fusion. The accuracy of the event detection method proposed in this paper is proved by comparison with other Chinese event detection methods. In the future work, an attention mechanism will be considered. Meanwhile, the event extraction task will be further studied, that is, to complete event detection and event element extraction at the same time.

REFERENCES

[1] H. Wang, and H. Gao, ''Reseasrch on event recognition based on word2vec and dependency analysis,'' *Comput. Eng. Softw.*, vol. 38, no. 6, pp. 62–65, Jun. 2017.

[2] R. Yangarber, ''Scenario customization for information extraction,'' *Diss. Abstr. Int., B, Sci. Eng.*, vol. 61, no. 12, pp. 6574–6574, Dec. 2001.

[3] P. Wu, Q. Chen, and L. Ma, ''Study on intelligent retrieval of event relevant documents based on event frame,'' *J. Chin. Inf. Process.*, vol. 17, no. 6, pp. 25–30, Mar. 2003.

[4] J. Jiang, ''A research about the pattern acquisition for free text IE,'' Ph.D. dissertation, Inst. Comput., Chin. Acad. Sci., Beijing, China, 2004.

[5] H. Liang, Q. Chen, and P. Wu, ''Information extraction system based on event frame,'' *J. Chin. Inf. Process.*, vol. 20, no. 2, pp. 40–46, Apr. 2006.

[6] L. Feng, ''Breaking events information extraction based on event frame,'' M.S. thesis, Software College, Shanghai JiaoTong Univ., Shanghai, China,2008.

[7] Q. Li, H. Ji, and L. Huang, ''Joint event extraction via structured prediction with global features,'' in *Proc. 51st Annu. Meeting Assoc. Comput. Linguistics*, Sofia, Bulgaria, Aug. 2013, pp. 73–82.

[8] H. Liang, S. Fayong, and G. Yang, ''Structured perceptron with inexact search,'' in *Proc. Conf. North Amer. Chapter Assoc. Comput. Linguistics: Hum. Lang. Technol.*, Montreal, PQ,, Canada, Jun. 2012, pp. 142–151.

[9] T. H. Nguyen and R. Grishman, ''Event detection and domain adaptation with convolutional neural networks,'' in *Proc. 53rd Annu. Meeting Assoc. Comput. Linguistics 7th Int. Joint Conf. Natural Lang. Process.*, Beijing, China, Jul. 2015, pp. 365–371.

[10] N. Kalchbrenner, E. Grefenstette, and P. Blunsom, ''A convolutional neural network for modelling sentences,'' in *Proc. 52nd Annu. Meeting Assoc. Comput. Linguistics*, Baltimore, MD, USA, Apr. 2014, pp. 655–665.

[11] L. Hou, P. Li, and Q. Zhu, ''Study of event recognition based on CRFs and cross-event,'' *Comput. Eng.*, vol. 38, no. 24, pp. 191–195, Jan. 2012.

[12] Y. Chen, L. Xu, K. Liu, D. Zeng, and J. Zhao, ''Event extraction via dynamic multi-pooling convolutional neural networks,'' in *Proc. 53rd Annu. Meeting Assoc. Comput. Linguistics 7th Int. Joint Conf. Natural Lang. Process.*, Beijing, China, Jul. 2015, pp. 26–31.

[13] Y. Zeng, H. Yang, Y. Feng, Z. Wang, and D. Zhao, ''A convolution BiLSTM neural network model for Chinese event extraction,'' in *Proc. Int. Conf. Comput. Process. Oriental Lang.*, Kunming, China, Dec. 2016, pp. 275–287.

[14] X. Feng, B. Qin, and T. Liu, ''A language-independent neural network for event detection,'' *Sci. China Inf. Sci.*, vol. 61, no. 9, pp. 81–92, Aug. 2018.

[15] H. Li, L. Yu, S. Tian, I. Turgun, and J. Zhao, ''Uyghur emergency event extracton based on DCNNs-LSTM model,'' *J. Chin. Inf. Process.*, vol. 32, no. 6, pp. 52–61, Jun. 2018.

[16] W. Hou, and D. Ji, ''Research on clinic event recognition based Bi-LSTM,'' *Appl. Res. Comput.*, vol. 35, no. 7, pp. 1974–1977, Jul. 2018.

[17] X. Jiang, H. Zhang, R. Liu, and Y. Zuo, ''A diversifying hidden units method based on NMF for document representation,'' in *Proc. IEEE Int. Conf. Knowl. Eng. Appl.*, Sep. 2016, pp. 103–107.

[18] Y. Liu, W. Song, L. Liu, and H. Wang, ''Document representation based on semantic smoothed topic model,'' in *Proc. 17th IEEE/ACIS Int. Conf. Softw. Eng., Artif. Intell., Netw. Parallel/Distrib. Comput.*, Shanghai, China, May/Jun. 2016, pp. 65–69.

[19] L. Zhu, G. Wang, and X. Zou, ''A study of chinese document representation and classification with word2vec,'' in *Proc. 9th Int. Symp. Comput. Intell. Design*, Hangzhou, China, Dec. 2016, pp. 298–302.

[20] R. Socher, A. Perelygin, J. Wu, J. Chuang, C. D. Manning, A. Ng, and C. Potts, ''Recursive deep models for semantic compositionality over a sentiment treebank,'' in *Proc. Conf. Empirical Methods Natural Lang. Process.*, Seattle, WA, USA, Oct. 2013, pp. 1631–1642.

[21] G. E. Hinton, ''Learning distributed representations of concepts,'' in *Proc. 8th Annu. Conf. Cogn. Sci. Soc.*, Amherst, MA, USA, 1989, pp. 46–61.

[22] T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado, and J. Dean, ''Distributed representations of words and phrases and their compositionality,'' *Comput. Sci.*, vol. 26, pp. 3111–3119, Oct. 2013.

[23] Y. Wu and J. Zhang, ''Chinese event extraction based on attention and semantic features: A bidirectional circular neural network,'' *Future Internet*, vol. 10, no. 10, p. 95, Sep. 2018.

[24] Y. Zhang, Z. Liu, and W. Zhou ''Event recognition based on deep belief network,'' *Acta Electron. Sinica*, vol. 45, no. 6, pp. 1415–1423, Jul. 2017.

[25] S. Wang, I. Turgun, A. Kahaerjiang, W. Aishan, and A. Gulinigeer, ''Sentiment classfication of Uyghur text based on BLSTM,'' *Comput. Eng. Des.*, vol. 38, no. 10, pp. 2879–2886, Nov. 2017.

[26] [Online]. Available: http://ltp.ai/.Language Technology Platform. 2019.4.20

[27] [Online]. Available: https://github.com/shijiebei2009/CEC-Corpus. Chinese Emergency Corpus.2019.4.21

[28] X. Xuan, T. Liao, and B. Gao, ''Automatic extraction of Chinese event trigger word,'' *Comput. Digit. Eng.*, vol. 43, no. 3, pp. 457–461, Apr. 2015.

[29] C. Ma, X. Chen, W. Wang, and H. Wang, ''Chinese event detection based on recurrent neural network,'' *Netinfo Secur.*, no. 5, pp. 75–81, May. 2018.

GUIXIAN XU was born in Changchun, Jilin, China, in 1974. She received the B.S. and M.S. degrees from the Changchun University of Technology, in 1998 and 2002, respectively, and the Ph.D. degree in computer software and theory from the Beijing Institute of Technology, in 2010.

Since 2002, she has been a Teacher with the Information Engineering College, Minzu University of China, where she is currently a Professor with the College of Information Engineering. Her research interests include data mining and machine learning.

YUETING MENG was born in Shijiazhuang, Hebei, China, in 1996. She received the B.S. degree in computer science and technology from the Hebei University of Science and Technology, in 2018. She is currently pursuing the master's degree in software engineering with the Minzu University of China. Her research interests include artificial intelligence, natural language processing, and data mining.

XIAOKAI ZHOU was born in Luan, Anhui, China, in 1999. He is currently pursuing the bachelor's degree in computer science and technology with the Minzu University of China. His research interests include data mining and machine learning.

ZIHENG YU was born in Taizhou, Zhejiang, China, in 1994. He received the B.S. degree in software engineering from Beijing Union University, in 2017. He is currently pursuing the master's degree in software engineering with the Minzu University of China. His research interests include data mining, natural language processing, and artificial intelligence.

XU WU was born in Fenghuang, Hunan, China, in 1993. He received the B.S. degree in software engineering from the Chongqing University of Posts and Telecommunications, in 2017. He is currently pursuing the master's degree in modern education technology with the Minzu University of China. His research interests include data mining, natural language processing, and artificial intelligence.

LIJUN ZHANG was born in Taian, Shandong, China, in 1997. She received the B.S. degree in intelligent science and technology from Qingdao University, in 2019. She is currently pursuing the master's degree in computer science and technology with the Minzu University of China. Her research interests include data mining, natural language processing, and artificial intelligence.