Computing and Informatics, Vol. 41, 2022, 1625–1647, doi: 10.31577/cai_2022_6_1625

RESEARCH ON EVENT EXTRACTION MODEL BASED ON SEMANTIC FEATURES OF CHINESE WORDS

Shaowu Zhu, Haichun Sun*

College of Information and Network Security
People's Public Security University of China
Beijing, China
e-mail: shaowuzhu@163.com, sunhaichun@ppsuc.edu.cn

Hanying Jian

College of Public Security Management
People's Public Security University of China
Beijing, China
e-mail: jianhanying@126.com

Abstract. Event Extraction (EE) is an important task in Natural Language Understanding (NLU). As the complexity of Chinese structure, Chinese EE is more difficult than English EE. According to the characteristics of Chinese, this paper designed a Semantic-GRU (Sem-GRU) model, which integrates Chinese word context semantics, Chinese word glyph semantics and Chinese word structure semantics. And this paper uses the model for Chinese Event Trigger Extraction (ETE) task. The experiment is compared in two tasks: ETE and Named Entity Recognition (NER). In ETE, the paper uses ACE 2005 Chinese event dataset to compare the existing research, the effect reaches 75.8 %. In NER, the paper uses MSRA dataset, which reaches 90.3 %, better than other models.

Keywords: BERT, five strokes, Chinese glyph, event extraction, trigger extraction, named entity recognition

Mathematics Subject Classification 2010: 68-T10

* Corresponding author

1 INTRODUCTION

In recent years, the task of Information Extraction (IE) usually involves embedding words in sentences into semantic space and using the vector of semantic space to represent the meaning of words in natural language. Word2vec [1] uses CBOW and Skip-gram methods to combine the semantic information of the context and form the semantic vector of the text. BERT Pre-trained Language Model (PLM) [2] uses the attention method in sentences to make the word vectors contain the semantics of sentences, effectively solving the problem of polysemy. Word semantic vectors can be trained from a large amount of generic domain text, which can be transferred to other tasks, making the model achieve better results through fine-tuning with lower resource consumption.

Text usually consists of a set of events, each containing different actions (trigger words) and different roles (event elements). By extracting events from the text, the information contained in the text can be presented in a structured way. This is very important for information processing. However, different natural languages have different ways of semantic expression. The arrangement of letters in English has special meanings. The prefixes 'un-' and 'in-' carry the meaning of 'no'. The radical and the glyph of Chinese have special semantics, '扌' carries the meaning of 'hand', and '日' carries the shape of the sun. The existing LM focuses on the contextual logic features of language but ignore the characteristics of language itself.

The research on Chinese LM is becoming more and more important. NLP tasks on Chinese have gained broad interest. Compared with English, Chinese EE is more difficult. This is mainly because the entity boundary of Chinese text is difficult to determine and Chinese grammatical structure is complex [3]. The entity boundary and character boundary of Chinese are not unified. And the word segmentation leads to the accumulation of error propagation, which is also a major challenge for Chinese recognition at present. At the same time, Chinese text and English text have different features. As a pictograph with a long history, Chinese glyphs contain rich semantic information. The general LM is designed based on the English text, which cannot extract the glyph semantic information in a single word. It can lead to the loss of partial Chinese semantic information, which is difficult to obtain through model training. Therefore, constructing the appropriate type of feature information based on Chinese characters can enrich the semantics and improve the task effect in Chinese domain.

Contributions. The main contributions of this paper are as follows:

1. A Semantic Gated Recurrent Unit (Sem-GRU) model is proposed to study Chinese ETE. The accuracy of Chinese ETE is improved by integrating three types of Chinese text semantic information, including word context, word glyph and word structure.

2. Sem-GRU model is applied to two tasks, ETE and NER. GRU is used to get deep features between characters in Language Model (LM).
3. In ETE, ACE 2005 Chinese event dataset is used to improve F1-score to 75.8 %; In NER, MSRA dataset is used to test, F1-score reached 90.3 %.

2 RELATED WORK

Because different natural languages have different semantic features, it is necessary to distinguish languages for study. Different natural languages have different linguistic features. According to features, adopting different algorithm structures can get a better effect.

In the general domain, most studies are based on PLM, such as BERT. PLM has been able to represent the basic meaning of words with a large amount of text training in the general domain. After input text data, the language text is transformed into a semantic vector through pre-training model encoding. In text-related tasks, combining the semantic features of the text can help the machine to better identify the text. Event-based text segmentation (EVENTSEG) [4] is proposed as an auxiliary task to improve sub-event detection. A learning and constraint method is proposed to improve the extraction effect of the relation between sub-event detection and EVENTSEG prediction. Compared with the baseline method, experimental results show that the proposed method improves the sub-event detection by 2.3 % and 2.5 % on HiEve and IC, respectively, and achieves good performance in EVENTSEG prediction. Lai et al. [5] proposed to model the relations between training tasks in episodic few-shot learning by introducing a cross-task prototype. Experimental results show that the effect is consistent on three few-shot learning datasets and the model is more robust when the labeled data of new event types are limited. Chen et al. [6] identified and solved the trigger curse problem in few-shot event detection (FSED). They suggest intervention through backdoor adjustments during training. Experimental results show that the method improves the performance on MAVEN, ACE 2005 and KBP17 datasets. Li et al. [7] introduced a new concept of temporal complex event mode: a graph-based schema representation, including events, arguments, temporal connections and relations. Yu et al. [8] proposed a new lifelong learning framework to deal with the prediction of dynamic events. They take lifetime event detection (ED) as an example and propose a new problem formula that can be extended to other Information Extraction (IE) tasks. Experimental results show that the framework outperforms the competitive baseline, improving F1 by 5.1 %. In addition, on some new long-tail rare event types, the framework can improve F1-score by more than 30 %. Le and Nguyen [9] presented a study on fine-grained event detection (FED), in which the event types of datasets reached 449. They propose a method to transform the Semcor dataset for word sense disambiguation into a large and high-quality dataset for fine-grained ED. Paolini et al. [10] focused on multi-task learning and proposed a Translation between Augmented Natural Language (TANL) model, which added annotations to output text. In entity extraction, relation classification, and semantic role labeling have reached new heights. Lu et al. [11] proposed the Text2Event model to extract events into the form of Seq2Seq, input sentences and output structured events. Text2Event achieves relatively competitive performance compared to the other methods. Hsu et al. [12] adopted a generative approach to generate predefined templates with Prompt. An end-to-end design model can incorporate knowledge of tags that capture the dependencies between triggers and arguments. Experimental results on low-resource end-to-end EE tasks show that the model is more efficient and exceeds all baseline. Lin and Chen [13] designed an event template for EE task, which can be divided into two prompt modes: Single Argument Prompt and Joint Argument Prompt. Experiments show that the model can achieve good results.

Many people study EE at different levels. Usually, EE is word-level, but sentence-level and script-level semantic are indispensable. Bai et al. [14] proposed a model called MCPredictor, which integrates deep event level and script level information for script event prediction. Experiments on the New York Times corpus verify the superiority and effectiveness. Li et al. [15] studied document-level EE and modeled the task into a form of event template generation and a keyword-based ETE method. The experimental results significantly outperform the previous approach in the cross-sentence RAMS dataset and WIKIEVENTS dataset. Previous studies on document-level EE mainly focused on building argument chains in the form of autoregressive. Zhu et al. [16] designed a non-autoregressive decoding algorithm to extract event parameter combinations from pruned complete graphs under the guidance of automatically selected pseudo triggers, which only took 3.6 % GPU time to train. And reasoning up to 8.5 times faster. Du et al. [17] introduced an encoder-decoder framework based on generating transformer (GRIT), which is designed for modeling in document-level context: it can make extraction decisions across sentence boundaries; is implicitly aware of noun phrase coreference structure and can respect cross-role dependencies in the template structure. Pouran Ben Veyseh et al. [18] proposed a method that used BERT to conduct document-level context modeling for EE. This method dynamically selects relevant sentences in the document for event prediction of the target sentence. The selected sentence will be enhanced with the target sentence and the presentation learning of EE will be improved by BERT. The validity of the model is demonstrated by experiments with multiple reference datasets.

In the domain of Chinese, there are also many algorithms for Natural Language Processing (NLP). Hou et al. [19] proposed a BERT-based Chinese relation extraction algorithm. The algorithm can effectively extract relevant security information when applied to the domain of public security. Liu et al. [20] use Convolutional Neural Network (CNN) and BERT to study the problem of NER, and use CNN to extract the semantic features of the context. Wang et al. [21] proposed ERNIE-Joint, a joint model based on ERNIE. ERNIE-Joint can make use of sentence-level and word-level functions through joint training of NER and text classification task. It has a good recognition effect on MSRA-NER and Microblog data. Yu et al. [22] used BERT and dichotomous fine-tuning methods to break sentences of ancient Chinese prose without punctuation, and compared with BiLSTM-CRF, the effect is better. Xie et al. [23] applied the BERT-BiLSTM-CRF model to Chinese NER, and the recognition effects in MSRA corpus and People's Daily corpus reached 94 % and 95 %. Gan and Zhang [24] studied the influence of contextual character embedding with BERT and proposed a method to integrate word information into self-attentional network word segmentation, which combined with BERT-CRF method performed better than LSTM method. Zhang et al. [25] proposed a BERT-based entity type information model, BERT-KGC, which is used to complete knowledge graph of Chinese cultural relics texts, to reduce ambiguity between entities and relations to a certain extent and achieve better effects.

In Chinese ETE, Li et al. [26] proposed an inference mechanism to infer unknown triggers through compound semantics within Chinese words and another inference mechanism to recover trigger mentions through textual consistency among Chinese trigger mentions.

Most studies focus on the application of the model but fail to integrate the characteristics of Chinese text into the model. Zhao et al. [27] proposed a character vector generation model based on a Chinese stroke sequence. But Chinese stroke sequence features are not well-matched in semantics. These studies have improved the model with characteristics of Chinese, but they still lack image semantics, which is important semantic information of Chinese. Chinese belongs to pictographs, and the difference between Chinese and English lies in its image semantics rather than simple symbolic representation. Therefore, the study of Chinese must be integrated with Chinese image semantics.

3 CHINESE SEMANTICS EXTRACTION FRAMEWORK

This chapter proposes and formalizes three semantic structures based on Chinese, extracting semantic features implied in Chinese symbols from three dimensions respectively. It includes word vector feature, glyph feature and structure feature of Chinese characters.

This paper proposes and constructs a Sem-GRU model for ETE and NER in text. Sem-GRU model comprehensively considers the semantics of Chinese word contexts, Chinese word glyphs and Chinese word structures, and extracts the three kinds of Chinese semantics, to obtain a relatively complete semantic recognition of Chinese. The overall structure of the proposed model is shown in Figure 1.

3.1 Extracting Word Context Semantics Based on BERT

The BERT model has a good representation of text semantics after a large number of public general domain text training. The word vector output by the model can represent the basic semantics of the text. Specifically, semantically related synonyms

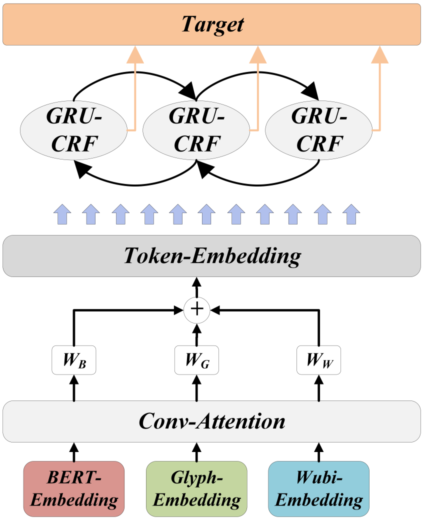

Figure 1. Architecture of Sem-GRU Model

have highly similar directions in vector space, such as "wood" and "tree", and their semantic vectors can be approximately parallel in vector space.

In addition, natural language usually shows a sequential flow structure, and there is a strong correlation in context. BERT's semantic vectors enable the model to learn the semantic meaning represented by words in context structure. This kind of semantics conforms to the law of natural language and is the most important part of the expression structure of natural language.

The input word context semantics of Chinese text is to extract the meaning of each word in the language context. This kind of semantic is obtained through the training of a large number of texts and can express good contextual semantic information.

BERT is an encode-decode model based on transformer, which can obtain a wide semantic model through pre-training encode structures. Then fine-tuning of specific tasks is carried out by decoding structure, to accurately identify and classify specific tasks. It emphasizes that the traditional one-way LM or the method of shallow splicing of two one-way LM is not used for pre-training, but the new Masked Language Model (MLM) is used to generate in-depth two-way language representation and performs well in many general domains.

The attention mechanism in BERT structure [28] enables the model to focus on the relationship between words in the language, excavate the implied information in language, and deal with language problems with good results. The attention method is shown in formulas (1), (2), (3) and (4):

$$Q_{seq} = W_Q \cdot X, \tag{1}$$

$$K_{seq} = W_K \cdot X, \tag{2}$$

$$V_{seq} = W_V \cdot X, \tag{3}$$

$$Attention_{Emb} = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{4}$$

where X is the input word vector, 768 dimensions. QKV is the parameter matrix, 768 dimensions. *Attention* is the token vector of the word, 768 dimensions.

BERT first segmented a language sequence. The initial input is marked with [CLS] and the sentence interval marked with [SEP]. The input text is converted into a vector after embedding layer. Each embedding is composed of three parts: segment, word, and position. Only word and position are used in the current embedding task. Word embeddings represented the word vector in the vector space, and position embeddings represented the position information of the word in the sentence. The input word vector information part is shown in formula (5):

$$Infor_{BERT} = transformer(token_{emb} + pos_{emb}) \tag{5}$$

where *emb* is the 768 dimensions vector, *token* is the word semantic vector, *pos* is the word position vector. *Infor* is the final BERT semantic output vector.

3.2 Extracting Word Glyph Semantics Based on CNN

English is phonetic while Chinese is ideogrammatic like hieroglyphs. The sources of the two languages are not unified, and the ways of expressing semantics are therefore different. The glyphs of ideograms contain pictorial expressions of things, such as "日" and "月", the initial image form of the two characters is formed according to the shape of the sun and moon. Therefore, it is a new attempt to extract the semantic information contained in Chinese characters to supplement the construction of Chinese word glyph semantics model, which is conducive to the machine to further distinguish the semantic meaning information of Chinese characters. With the help of the ideogrammatic features of Chinese characters, the model helps the machine to understand the semantics of the characters themselves.

Chinese characters have a long history. In the process of continuous evolution, Chinese characters have undergone many changes and eventually formed the current simplified Chinese characters. In the history of Chinese language development, the oracle bone script is the earliest, relatively mature and complete Chinese character, which later developed into Chinese bronze inscriptions. After the First Emperor of Qin unified China, he implemented the principle of "train on the same track, book on the same script", which unified the characters into the lesser seal characters.

Among the ancient characters of the oracle bone script, about 2 500 characters have been deciphered and confirmed. In the “金文编”, 2 420 Chinese bronze inscriptions can be identified, “说文解字” included 9 353 the lesser seal characters. The collection situation of three words sources is shown in Table 1.

Script	Number	Source
Oracle bone script	About 2 500	Existing research
Bronze inscriptions	2 420	“金文编”
The lesser seal	9 353	“说文解字”

Table 1. Oracle, Bronze inscriptions and The lesser seal script collection situation

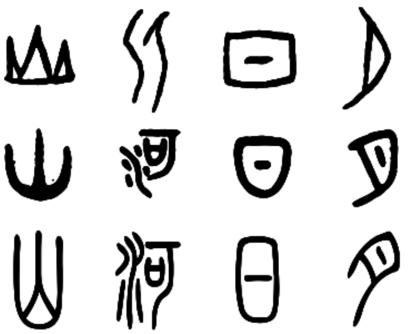

Figure 2. Examples of three types of script images

Figure 2 shows an example. It shows three kinds of Chinese words from different word sources in Table 1. The four words contain “mountain”, “river”, “sun” and “moon”. Among them, are from top to bottom fonts for oracle bone script, Chinese bronze inscriptions and the lesser seal characters.

The simplified Chinese characters commonly used have lost some character information in the process of evolution. After analyzing the evolution process of Chinese fonts and comparing different font types, by integrating the character information and character quantity, we choose “the lesser seal characters” font to extract the ideogrammatic features of characters. Compared with the oracle bone script and Chinese bronze inscriptions, the lesser seal characters are simplified in shape, but their volume is far more than the other two characters, which is conducive to the machine extraction of its characteristics.

There is abundant glyph semantic information in Chinese, which can be extracted by using text and images. Firstly, the mapping relationship between simplified Chinese characters and the lesser seal characters is established. Simplified Chinese text is usually used in a modern font, and the model encodes simplified Chinese text as the lesser seal characters images. Each simplified Chinese character is related to its picture of the lesser seal characters. For the ideogrammatic semantic extraction of a single word, the text pixel image of "the lesser seal characters" font is formed, and the standard format is 50×50 pixels gray image, as shown in Figure 3.

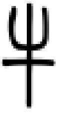

Figure 3. The lesser seal characters of Chinese character – '牛' (cow)

CNN is constructed by imitating the biological visual perception mechanism, and the convolutional kernel in the network can better extract the features of the lattice, which has a better fitting effect for image processing problems. Taking the image information of the text as the input of the Neural Network (NN), the image features of the ideogrammatic can be extracted, to get better results for the text classification problem.

In the frame design of text glyph information extraction, two convolution operations are done. Specifically, for each convolution, 3×3 convolution is used to check each text image for convolution operation.

3.3 Extraction Word Structure Semantics Based on GRU

The glyph of Chinese characters is a strong feature in semantic information. Meng et al. [29] used Chinese glyphs to enhance semantic representation. The five-stroke representation of Chinese characters can also reflect the structural features of Chinese characters to a certain extent.

Five-stroke character type is a Chinese input method invented by Yongmin Wang in 1983, which is called "Wang encode five-stroke". Five-stroke character type is a typical font code input method, which encodes Chinese characters completely according to strokes and characters. Five-stroke codes complex Chinese characters in the way of stroke structure by encoding and combining stroke structure of Chinese characters. At the same time, five-stroke coding also takes into account the complexity of input, disassembles and codes complex characters in the way of word, and preserves the semantic information of characters and glyphs, as shown in Figure 4.

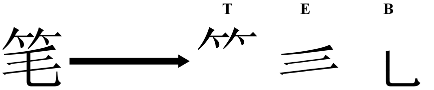

Figure 4. Five-stroke coding architecture of "笔"

The five-stroke characters show the basic components of Chinese characters. The radicals of Chinese characters also contain the meaning of Chinese characters, especially in verbs. By the five-stroke character type, the semantic information of Chinese characters is further supplemented by dismantling the radicals of Chinese characters.

The model uses five-stroke coding to vectorize Chinese characters and designs a set of five-stroke coding vectorization methods. The machine can process vectorized coding, and the model can learn the semantics represented by each coded word from a large amount of training data.

First of all, a set of mapping dictionaries of common Chinese characters and five-stroke codes is established, $\{\text{‘谓’} : \text{‘}yleg\text{’}, \text{‘语’} : \text{‘}ygkg\text{’}, \dots\}$. On input, the text is first mapped to a five-stroke code. Five-stroke code uses 26 letters from A to Z as the code element, and the code element of each group is no more than 4 dimensions. Therefore, when coding vectorization, the integer value of 4-dimensional 1-26 is used as the coding vector, $\{\text{‘}yleg\text{’} : [25, 12, 5, 7], \text{‘}ygkg\text{’} : [25, 7, 11, 7], \dots\}$.

And then the Gated Recurrent Unit (GRU) algorithm is introduced to extract the deep features of the five-stroke code element vector. The five-stroke vector $e^{wubi}(c_{(i,1)})$, $\dots$, $e^{wubi}(c_{(i,k)})$, $i \in (1, 4)$, serves as the input of GRU, we can further learn that the hidden state of five-stroke characters is h_1, $\dots$, h_k, then the last feature of five-strokes v^{wk} can be expressed as formulas (6), (7), (8), (9), and (10):

$$z_t = \sigma(W_z \cdot e^{wubi}(c_{i,k})_t + GRU_z \cdot h_{t-1}), \tag{6}$$

$$r_t = \sigma(W_t \cdot e^{wubi}(c_{i,k})_t + GRU_z \cdot h_{t-1}), \tag{7}$$

$$\tilde{h_t} = \tanh(W_t \cdot e^{wubi}(c_{i,k})_t + GRU(r_t \times h_{t-1})), \tag{8}$$

$$h_t = (1 - z_t) \times h_{t-1} + z_t \times \tilde{h_t}, \tag{9}$$

$$Infor_{wubi} = GRU(h_{k-1}, input_k), \tag{10}$$

where e^{wubi} is the five-stroke input vector, h_{t-1} is the last time output, $\tilde{h_t}$ is the current time information. *Infor* is the final GRU semantic output vector.

3.4 Word Semantic Fusion Based on Context Semantics, Glyph Semantics and Structure Semantics

Word semantic fusion is to synthesize the features of the three kinds of semantics, comprehensively extract the features and form the specific coded semantic information.

In the process of semantic fusion of the model, the semantic vector is regarded as the image vector, and the RGB three channels correspond to Chinese word context semantics, Chinese word glyph semantics and Chinese word structure semantics, respectively. Each channel information contains a set of 128×768 vectors, namely

128 × 768 pixels semantic image. The semantic information is convolved in layers, and the three semantics are complementary to each other. 128 represents the length of input text, and 768 represents the semantic vector corresponding to each word. Formulas (11), (12), (13), and (14) represent the operations in semantic fusion:

$$X(u,v) = \begin{bmatrix} a_{u-1,v-1} & a_{u-1,v} & a_{u,v+1} \\ a_{u,v-1} & a_{u,v} & a_{u,v+1} \\ a_{u+1,v-1} & a_{u+1,v} & a_{u+1,v+1} \end{bmatrix}, \tag{11}$$

$$Conv(u,v)_{R,G,B} = \sum_{i=1}^{k}\sum_{j=1}^{k} X_{i,j} \cdot K_{i,j}, \tag{12}$$

$$Multiconv(u,v) = \sum_{i=1}^{n} Conv(u,v)_i, \tag{13}$$

$$O_{conv}(u,v) = \max(Relu(Multiconv(u,v)_{i,j|i,j\in(1,k)})) \tag{14}$$

where K represents the convolution kernel, 3 × 3 matrix, k represents the shape of the convolution kernel, n is the channel number of convolution, u and v are the positions of the matrix. *Conv* and *multiconv* are convolution operations.

4 CHINESE SEMANTIC VECTOR INTERPRETATION

In encode-decode model, the encoder section encodes the input sentences to form the semantic vector of sentences. The decoder section decodes semantic vectors to accomplish specific tasks.

Chinese semantic vector integrates characteristics of Chinese glyph structure and can represent the deeper semantics of Chinese in the encoding part and form a better semantic vector. Secondly, in the decoding part, different decoders are designed for fine-tuning according to different tasks.

In this paper, we complete the task of ETE and NER. These tasks all belong to Sequence Labeling (SL) tasks. In SL tasks, models identify each input sequence in a sentence, which is an important task of NLP. To fit such tasks, the following methods are used:

Semantic interpretation decoder design includes two parts, Bi-directional Gated Recurrent Unit (BiGRU) and Conditional Random Field (CRF). In line with the characteristics of BiGRU, component relations in text sentences are extracted and analyzed, and the transfer probability of text labels is modified by CRF, as shown in Figure 5.

BiGRU belongs to the sequential NN, which is improved from LSTM. LSTM is used to solve the problem of long-term dependent gradient explosion which cannot be solved by Recurrent Neural Network (RNN) algorithm. Set up two propagation streams, long-term memory and current output. GRU simplifies the structure of

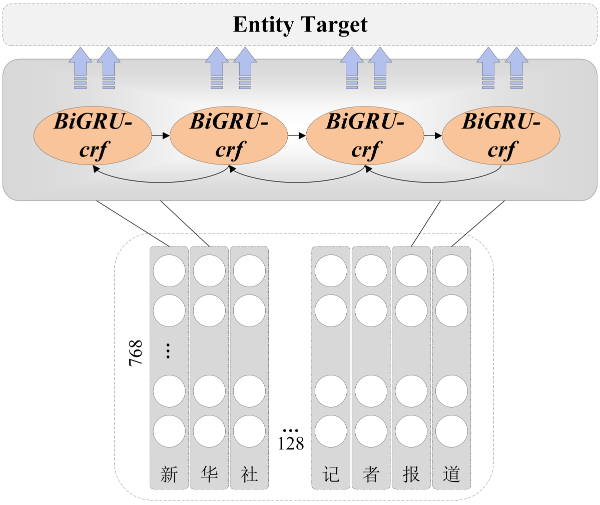

Figure 5. Architecture of decoding layer BiGRU-CRF

LSTM, eliminates the propagation flow of long-term memory, adopts the update unit and the reset unit, directly controls the current input and output, reduces parameters, and makes the model convergence easier while maintaining the performance.

The bidirectional GRU algorithm can better extract the relevant information in sentences. Forward GRU trains the association information of the target word above, while reverse GRU can train the association information of the target word below. For the information at each time, there is contextual information associated with it.

After semantic extraction of text vector by Sem-GRU model, BiGRU is introduced to extract semantic fusion between words in sentences. The number of BiGRU layers is set to 3, and the number of hidden layer neurons is 128.

As a common model for solving sequence problems, CRF can control the causal sequence relationship in the decoding layer and provide a set of transition probabilities. In the sequence problem, CRF can comprehensively consider the rationality of sequence labels. For example, in the **BIO** annotation method, **B** is the beginning of the positive label and **I** is the rest of the positive label. Then **I** will not appear in the head alone. Only when the previous label is **B**, will the **I** appear in probability. CRF controls the transfer probability, so as to learn the rationality rules of labels and improve the identification accuracy. Formulas (15), (16), (17), and (18) represent the calculation process.

$$s(X,Y) = \sum_{i=0}^{n} A_{y_i,y_{i+1}} + \sum_{i=1}^{n} P_{i,y_i} \tag{15}$$

where A represents the transfer score matrix, and A_{ij} represents the transfer score of the i label to the j label. P_{ij} represents the j^{th} label fraction of the i^{th} word. The probability of generating the predicted sequence Y is:

$$p(Y|X) = \frac{e^{s(X,Y)}}{\sum_{Y' \in Y_X} s(X,Y')}. \tag{16}$$

Take the logarithm of both sides of the equation to get the likelihood function of prediction:

$$\ln(p(Y|X)) = s(X,Y) - \ln\left(\sum_{Y' \in Y_X} s(X,Y')\right). \tag{17}$$

Y' represents the real annotation, and Y_X represents all possible annotations. The output of the maximum score obtained after decoding is:

$$Y^* = \arg\max(s(X,Y)_{Y' \in Y_X}). \tag{18}$$

In the decoding layer, BiGRU-CRF takes into account the context semantic and text structure information, calculates the corresponding loss value, and transmits it back to the preceding parameter.

5 EXPERIMENT

The experiment environment used Intel(R) Xeon(R) Gold 6 240 CPU @ 2.60 GHz, 64 GB RAM, Ubuntu 5.4.0-58-Generic, NVIDIA TITAN V video card, 12 GB video memory, CUDA 11.0. TensorFlow 1.15.0.

Experiment parameter settings are shown in Table 2.

Parameter	Value
batch_size	256
droprate	0.5
max_seq_length	128
learning_rate	0.001
Information fusion channel	3
five-stroke vector	$(1, 4)$
image pixel	$(50, 50)$, 8 bit

Table 2. Experimental model parameter setting

5.1 Event Trigger Extraction

5.1.1 Data Base

We use the ACE 2005 Chinese event data [30] in our experiments, following standard EE literature. We use the split in Chinese with 521 documents for training, 64 for validation and 64 for testing. The method represents the triggers with only one trigger in one sample. So, we combined different triggers in the same sentence. Thus, the training samples are 2 745 and the validation and testing samples are 246, since the majority of triggers and their arguments are within the same sentence.

5.1.2 Experiment Results of Trigger Extraction

In the experiment part, the paper compares the effect of the existing EE algorithm and selects three experimental states of Chinese word context semantics (BERT), Chinese word glyph semantics (CNN) and Chinese word structure semantics (GRU) for comparison. We perform the EE task only at the sentence level and fine-tune our model for 100 epochs and 0.001 learning rate on this dataset.

In terms of evaluation indicators, F1-score was used for evaluation. Because there are many non-entity samples in NER task, it is easier to identify non-entity samples. Therefore, if the accuracy rate is used for model evaluation, the results obtained are not convincing. F1-score can eliminate the influence of negative samples on the rationality of evaluation indicators. Specifically, non-entity samples in the dataset are taken as negative examples and all entity samples are taken as positive examples.

Firstly, we compare results on the standard ACE 2005 Chinese dataset. Secondly, we carry out ETE experiments on the Sem-GRU constructed in this paper. We set up these groups of comparative experiments:

1. Li et al. [26] propose an inference mechanism through the compositional semantics inside Chinese words and the discourse consistency between Chinese trigger mentions.
2. Three-Layer Joint Model (3JM), proposed by Li and Zhou [31], a baseline of three components, i.e., trigger identification, event type determination and event subtype determination.
3. Convolutional Bi-LSTM model (C-BiLSTM), proposed by Zeng et al. [32], a baseline of sentence and lexical method, using BiLSTM to encode the semantics of words in the whole sentence, and then CNN is used to capture salient local lexical features.
4. Hybrid Neural Network (HNN), proposed by Feng et al. [33], a baseline of word-based method, which employs Bi-LSTM and CNN to capture sequence and chunk information from specific contexts, respectively.
5. Nugget Proposal Networks (NPNs), proposed by Lin et al. [34], a baseline of character and word-based method, learning a hybrid representation which can summarize structural and semantic information from characters and words, respectively.
6. Hybrid Character Representation (HCR), proposed by Xi et al. [35], a baseline of character-based method, which incorporates word information and LM

representation into Chinese character representation to capture inner structure feature of event triggers and sentence-level context semantics.

7. Residual and Gated-based Atrous Convolution Neural Network (RG-ACNN), proposed by Wang et al. [36], a baseline to propose the head and tail of the potential trigger, as well as to identify its corresponding event type at each character.

In the experiment, the BERT model uses the BERT-based Chinese. BERT model is based on Sem-GRU standard, only Chinese word context semantics are input. Chinese word glyph semantics and Chinese word structure semantics are reduced. GRU adds Chinese word structure semantics based on BERT. Similar to GRU, CNN uses Chinese word context semantics and Chinese word glyph semantics.

Table 3 shows the performance of Chinese trigger word classification model on ACE 2005.

System	Trigger Extraction		
	Precision/%	Recall/%	F1-score/%
Li	70.2	50.1	58.5
3JM	73.5	65.7	69.4
C-BiLSTM	69.8	59.9	64.5
HNN	77.1	53.1	63.0
NPNs	60.9	**69.3**	64.8
HCR	66.4	76.0	70.9
RG-ACNN	65.5	69.0	67.2
Our BERT	64.3	63.4	63.8
Our CNN	80.9	65.1	72.2
Our GRN	71.7	57.9	64.1
Our Sem-GRU	**84.0**	69.0	**75.8**

Table 3. Experiment results on the standard ACE 2005 Chinese dataset

Through the comparison of the effect of the ETE task, it can be seen that Sem-GRU model has outstanding performance in recognition effect. F1-score of ETE reaches 75.8 % and Precision reaches 84 %, while Recall does not have the best performance. But Recall is still high. The relative accuracy of recall rate is always relatively low, because under the regulation of CRF layer, it is usually possible to learn the rules of BIO annotation method – starting with 'B' and remaining with 'I'. Therefore, compared with negative examples, positive examples have a better recognition effect, and the model can better deal with the problem of target extraction. It is difficult to determine the boundary of the target. This is also a difficult problem for Chinese SL tasks.

The relationship between the training set and the test set as well as the predicted results were observed. In Ace 2005 dataset, the data annotation of some triggers is incomplete, as shown in Figure 6. In training set, the event "发动军事政变" is not marked. This may be because this event type is not one of the 33 predefined event subtypes. But in test-label, the dataset is not marked with the trigger “搜捕”, while in training set, the trigger “搜捕” was marked. This indicates that the Ace 2005 dataset still has some defects.

The predictions of the Sem-GRU model designed in this paper are shown in Figure 6. Test-predict did not identify the trigger word “落网”, indicating that the model still has room for improvement. At the same time, the model identified the trigger “搜捕”. Because the trigger is not labeled in the dataset, this prediction is false for the model, but it does not mean that the model performs badly. On the contrary, it suggests a certain capacity for self-correction. The model can selectively learn from the data that is labeled incorrectly.

Figure 6. Sem-GRU model prediction results analysis

To describe the learning ability of the model designed in this paper, under the standard of F1-score, the experiment showed the training effects of the four models under 1-100 epochs, as shown in Figure 7.

As shown in Figure 7, Sem-GRU model can effectively improve the effect of the model on the ETE task after integrating Chinese word context semantics, Chinese word glyph semantics and Chinese word structure semantics. And it has a good effect after model convergence.

5.2 Named Entity Recognition

5.2.1 Data Base

MSRA dataset is a Chinese NER dataset released by SIGHAN in 2006. It consists of simplified Chinese news, including people’s names, place names and organization names, with a total of 46 365 corpora. In the experiment, 20 % of the data were randomly selected as the test set and 80 % as the training set.

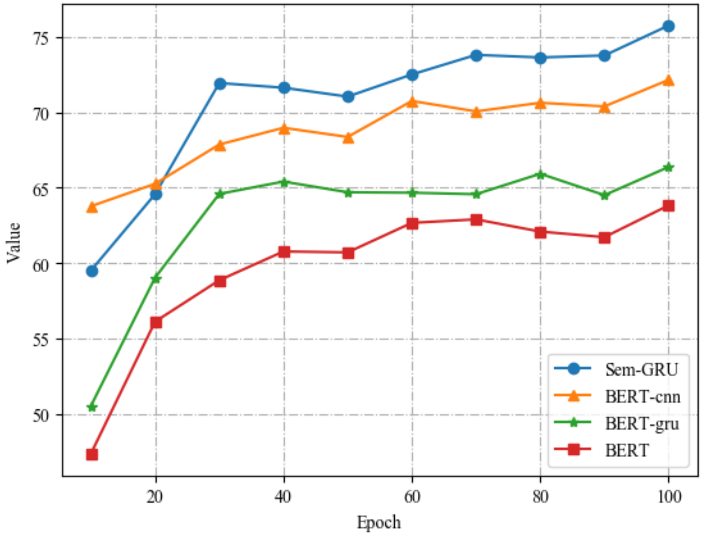

Figure 7. F1-score training effect on ACE 2005

5.2.2 Experiment Results of NER

The public MSRA dataset was used to test the universality of the model. MSRA dataset is an open and high-quality NER task dataset. Using MSRA dataset to test the model can more objectively show the superiority and efficiency of the model itself.

Algorithm	Precision/%	Recall/%	F1-score/%
MSRA BERT	88.0	75.5	81.2
MSRA GRU	90.1	76.9	83.0
MSRA CNN	94.2	86.2	90.0
MSRA Sem-GRU	**94.6**	**86.3**	**90.3**

Table 4. NER experiment effects comparison on MSRA dataset

As shown in Table 4, in MSRA dataset, GRU model does not improve the recognition effect significantly. In NER, there is more use of some proper nouns, which do not have stroke representation in the glyph. The verb strokes mostly contain action image semantics, such as "扌" meaning with hand, which is mostly a verb.

Moreover, the precision and recall rate of Sem-GRU model are greatly improved compared with other models. In terms of precision, the improvement of precision indicates that the model can better distinguish the features between entities and non-entities, and carry out deeper learning in high-frequency features. In terms of recall rate, the improvement of recall rate indicates that the Sem-GRU model can find more entities existing in the text, and the model can better learn the characteristics of the existing entity.

The epoch in the training was set as 20 and the learning rate was 0.001. The experimental results clearly show the difference in effect between different models. Sem-GRU model integrates Chinese word context semantics, Chinese word glyph semantics and Chinese word structure semantics. The lack of any Chinese information will affect Chinese recognition task, and the task effect will be reduced.

At the same time, GRU model and CNN model have some improvement in test results compared with the BERT model. This shows that the recognition efficiency of the model is gradually improved with the fusion of different information. This Sem-GRU model, which integrates Chinese semantics, can effectively find the existing Chinese information. The feature extraction of Chinese characters can effectively improve the recognition effect of generic domain tasks.

6 CONCLUSIONS

In this paper, a Sem-GRU model is proposed under the framework of EE, which combines the semantic information of Chinese word context semantics, Chinese word glyph semantics and Chinese word structure semantics. Secondly, we test the model in ETE and NER task. Experimental results show that Sem-GRU model can effectively improve the accuracy in Chinese domain.

The model uses convolution layers, which are not very friendly. Too many convolutional layers will reduce the training efficiency of the model, and there is a certain increase in the training time.

Acknowledgment

The authors contributed equally to this study. This work was supported by the Technical Research Project of the Ministry of Public Security (Grant No. 2020JSYJC22), the National Natural Science Foundation of China (Grant No. 41971367) and the Beijing Natural Science Foundation Program (Grant No. 4184099).

REFERENCES

[1] Mikolov, T.—Chen, K.—Corrado, G.—Dean, J.: Efficient Estimation of Word Representations in Vector Space. In: Bengio, Y., LeCun, Y. (Eds.): 1st International

Conference on Learning Representations (ICLR 2013), Workshop Track Proceedings. 2013, doi: 10.48550/arXiv.1301.3781.

[2] Devlin, J.—Chang, M. W.—Lee, K.—Toutanova, K.: BERT: Pre-Training of Deep Bidirectional Transformers for Language Understanding. In: Burstein, J., Doran, C., Solorio, T. (Eds.): Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers). ACL, 2019, pp. 4171–4186, doi: 10.18653/v1/N19-1423.

[3] Zhao, S.—Luo, R.—Cai, Z. P.: Survey of Chinese Named Entity Recognition. Journal of Frontiers of Computer Science and Technology, Vol. 16, 2022, No. 2, pp. 296–304, doi: 10.3778/j.issn.1673-9418.2107031 (in Chinese).

[4] Wang, H. Y.—Zhang, H. M.—Chen, M. H.—Roth, D.: Learning Constraints and Descriptive Segmentation for Subevent Detection. In: Moens, M. F., Huang, X. J., Specia, L., Yih, S. W. (Eds.): Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing (EMNLP 2021). ACL, 2021, pp. 5216–5226, doi: 10.18653/v1/2021.emnlp-main.423.

[5] Lai, V. D.—Dernoncourt, F.—Nguyen, T. H.: Learning Prototype Representations Across Few-Shot Tasks for Event Detection. In: Moens, M. F., Huang, X., Specia, L., Yih, S. W. (Eds.): Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing (EMNLP 2021). ACL, 2021, pp. 5270–5277, doi: 10.18653/v1/2021.emnlp-main.427.

[6] Chen, J. W.—Lin, H. Y.—Han, X. P.—Sun, L.: Honey or Poison? Solving the Trigger Curse in Few-Shot Event Detection via Causal Intervention. In: Moens, M. F., Huang, X., Specia, L., Yih, S. W. (Eds.): Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing (EMNLP 2021). ACL, 2021, pp. 8078–8088, doi: 10.18653/v1/2021.emnlp-main.637.

[7] Li, M. L.—Li, S.—Wang, Z. H. L.—Huang, L. F.—Cho, K.—Ji, H.—Han, J. W.—Voss, C. R.: The Future Is Not One-Dimensional: Complex Event Schema Induction by Graph Modeling for Event Prediction. In: Moens, M. F., Huang, X., Specia, L., Yih, S. W. (Eds.): Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing (EMNLP 2021). ACL, 2021, pp. 5203–5215, doi: 10.18653/v1/2021.emnlp-main.422.

[8] Yu, P.—Ji, H.—Natarajan, P.: Lifelong Event Detection with Knowledge Transfer. In: Moens, M. F., Huang, X., Specia, L., Yih, S. W. (Eds.): Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing (EMNLP 2021). ACL, 2021, pp. 5278–5290, doi: 10.18653/v1/2021.emnlp-main.428.

[9] Le, D.—Nguyen, T. H.: Fine-Grained Event Trigger Detection. In: Merlo, P., Tiedemann, J., Tsarfaty, R. (Eds.): Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume (EACL 2021). ACL, 2021, pp. 2745–2752, doi: 10.18653/v1/2021.eacl-main.237.

[10] Paolini, G.—Athiwaratkun, B.—Krone, J.—Ma, J.—Achille, A.—Anubhai, R.—dos Santos, C. N.—Xiang, B.—Soatto, S.: Structured Prediction as Translation Between Augmented Natural Languages. 9th International Conference on Learning Representations (ICLR), 2021, doi: 10.48550/arXiv.2101.05779.

[11] Lu, Y. J.—Lin, H. Y.—Xu, J.—Han, X. P.—Tang, J. L.—Li, A.—Sun, L.—

Liao, M.—Chen, S. Y.: Text2Event: Controllable Sequence-to-Structure Generation for End-to-End Event Extraction. In: Zong, C. Q., Xia, F., Li, W. J., Navigli, R. (Eds.): Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers) (ACL-IJCNLP 2021). ACL, 2021, pp. 2795–2806, doi: 10.18653/v1/2021.acl-long.217.

[12] Hsu, I. H.—Huang, K. H.—Boschee, E.—Miller, S.—Natarajan, P.—Chang, K. W.—Peng, N.: Text2Event: Controllable Sequence-to-Structure Generation for End-to-End Event Extraction. In: Carpuat, M., de Marneffe, M. C., Meza Ruiz, I. V. (Eds.): Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL 2022). ACL, 2022, pp. 1890–1908, doi: 10.18653/v1/2022.naacl-main.138.

[13] Lin, J. J.—Chen, Q.: PoKE: A Prompt-Based Knowledge Eliciting Approach for Event Argument Extraction. 2021, doi: 10.48550/arXiv.2109.05190.

[14] Bai, L.—Guan, S. P.—Guo, J. F.—Li, Z. X.—Jin, X. L.—Cheng, X. Q.: Integrating Deep Event-Level and Script-Level Information for Script Event Prediction. In: Moens, M. F., Huang, X., Specia, L., Yih, S. W. (Eds.): Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing (EMNLP 2021). ACL, 2021, pp. 9869–9878, doi: 10.18653/v1/2021.emnlp-main.777.

[15] Li, S.—Ji, H.—Han, J. W.: Document-Level Event Argument Extraction by Conditional Generation. In: Toutanova, K., Rumshisky, A., Zettlemoyer, L. et al. (Eds.): Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL 2021). ACL, 2021, pp. 894–908, doi: 10.18653/v1/2021.naacl-main.69.

[16] Zhu, T.—Qu, X. Y.—Chen, W. L.—Wang, Z. F.—Huai, B. X.—Yuan, N. J.—Zhang, M.: Efficient Document-Level Event Extraction via Pseudo-Trigger-Aware Pruned Complete Graph. 2021, doi: 10.48550/arXiv.2112.06013.

[17] Du, X. Y.—Rush, A.—Cardie, C.: GRIT: Generative Role-Filler Transformers for Document-Level Event Entity Extraction. In: Merlo, P., Tiedemann, J., Tsarfaty, R. (Eds.): Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume (EACL 2021). ACL, 2021, pp. 634–644, doi: 10.18653/v1/2021.eacl-main.52.

[18] Pouran Ben Veyseh, A.—Nguyen, M. V.—Ngo Trung, N.—Min, B.—Nguyen, T. H.: Modeling Document-Level Context for Event Detection via Important Context Selection. In: Moens, M. F., Huang, X., Specia, L., Yih, S. W. (Eds.): Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing (EMNLP 2021). ACL, 2021, pp. 5403–5413, doi: 10.18653/v1/2021.emnlp-main.439.

[19] Hou, J. Q.—Li, X.—Yao, H. P.—Sun, H. C.—Mai, T. L.—Zhu, R. C.: BERT-Based Chinese Relation Extraction for Public Security. IEEE Access, Vol. 8, 2020, pp. 132367–132375, doi: 10.1109/ACCESS.2020.3002863.

[20] Liu, Y. Y.—Yu, Z. T.—Gao, S. X.—Guo, J. J.—Zhang, Y. F.—Nie, B. G.: Chinese Named Entity Recognition Method Based on Machine Reading Comprehension. Pattern Recognition and Artificial Intelligence, Vol. 33, 2020, No. 7, pp. 653–659,

doi: 10.16451/j.cnki.issn1003-6059.202007007 (in Chinese).

[21] Wang, Y.—Sun, Y. N.—Ma, Z. C.—Gao, L. S.—Xu, Y.: An ERNIE-Based Joint Model for Chinese Named Entity Recognition. Applied Sciences, Vol. 10, 2020, No. 16, Art. No. 5711, doi: 10.3390/app10165711.

[22] Yu, J. S.—Wei, Y.—Zhang, Y. W.: Automatic Ancient Chinese Texts Segmentation Based on BERT. Journal of Chinese Information Processing, Vol. 33, 2019, No. 11, pp. 57–63 (in Chinese).

[23] Xie, T.—Yang, J. A.—Liu, H.: Chinese Entity Recognition Based on BERT-Bilstm-CRF Model. Computer Systems and Applications, Vol. 29, 2020, No. 7, pp. 48–55 (in Chinese).

[24] Gan, L. L.—Zhang, Y.: Investigating Self-Attention Network for Chinese Word Segmentation. IEEE/ACM Transactions on Audio, Speech, and Language Processing, Vol. 28, 2020, pp. 2933–2941, doi: 10.1109/TASLP.2020.3030487.

[25] Zhang, M.—Geng, G. H.—Zeng, S.—Jia, H. P.: Knowledge Graph Completion for the Chinese Text of Cultural Relics Based on Bidirectional Encoder Representations from Transformers with Entity-Type Information. Entropy, Vol. 22, 2020, No. 10, Art. No. 1168, doi: 10.3390/e22101168.

[26] Li, P. F.—Zhu, Q. M.—Zhou, G. D.: Using Compositional Semantics and Discourse Consistency to Improve Chinese Trigger Identification. Information Processing and Management, Vol. 50, 2014, No. 2, pp. 399–415, doi: 10.1016/j.ipm.2013.10.003.

[27] Zhao, H. X.—Yu, J. S.—Lin, J.: Design and Research on Chinese Word Embedding Model Based on Strokes. Journal of Chinese Information Processing, Vol. 33, 2019, No. 5, pp. 17–23 (in Chinese).

[28] Vaswani, A.—Shazeer, N.—Parmar, N.—Uszkoreit, J.—Jones, L.—Gomez, A. N.—Kaiser, L.—Polosukhin, I.: Attention Is All You Need. In: Guyon, I., Von Luxburg, U., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., Garnett, R. (Eds.): Advances in Neural Information Processing Systems 30 (NIPS 2017). Curran Associates, Inc., 2017, pp. 5998–6008.

[29] Meng, Y. X.—Wu, W.—Wang, F.—Li, X. Y.—Nie, P.—Yin, F.—Li, M. Y.—Han, Q. H.—Sun, X. F.—Li, J. W.: Glyce: Glyph-Vectors for Chinese Character Representations. In: Wallach, H., Larochelle, H., Beygelzimer, A., d'Alché-Buc, F., Fox, E., Garnett, R. (Eds.): Advances in Neural Information Processing Systems 32 (NeurIPS 2019). Curran Associates, Inc., 2019, pp. 2746–2757.

[30] Walker, C.—Strassel, S.—Medero, J.—Maeda, K.: ACE 2005 Multilingual Training Corpus LDC2006T06. Web Download. 2006, doi: 10.35111/mwxc-vh88.

[31] Li, P. F.—Zhou, G. D.: Three-Layer Joint Modeling of Chinese Trigger Extraction with Constraints on Trigger and Argument Semantics. Journal of Computer Science and Technology, Vol. 32, 2017, No. 5, pp. 1044–1056, doi: 10.1007/s11390-017-1780-5.

[32] Zeng, Y.—Yang, H. H.—Feng, Y. S.—Wang, Z.—Zhao, D. Y.: A Convolution BiLSTM Neural Network Model for Chinese Event Extraction. In: Lin, C. Y., Xue, N. W., Zhao, D. Y., Huang, X. J., Feng, Y. S. (Eds.): Natural Language Understanding and Intelligent Applications. Springer, Cham, Lecture Notes in Computer Science, Vol. 10102, 2016, pp. 275–287, doi: 10.1007/978-3-319-50496-4_23.

[33] Feng, X. C.—Qin, B.—Liu, T.: A Language-Independent Neural Network

for Event Detection. Science China Information Sciences, Vol. 61, 2018, No. 9, Art. No. 092106, doi: 10.1007/s11432-017-9359-x.

[34] Lin, H. Y.—Lu, Y. J.—Han, X. P.—Sun, L.: Nugget Proposal Networks for Chinese Event Detection. In: Gurevych, I., Miyao, Y. (Eds.): Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers) (ACL 2018). ACL, 2018, pp. 1565–1574, doi: 10.18653/v1/P18-1145.

[35] Xi, X. Y.—Zhang, T.—Ye, W.—Zhang, J. L.—Xie, R.—Zhang, S. K.: A Hybrid Character Representation for Chinese Event Detection. 2019 International Joint Conference on Neural Networks (IJCNN), 2019, pp. 1–8, doi: 10.1109/IJCNN.2019.8851786.

[36] Wang, Z. H.—Guo, Y.—Wang, J. H.: Empower Chinese Event Detection with Improved Atrous Convolution Neural Networks. Neural Computing and Applications, Vol. 33, 2021, No. 11, pp. 5805–5820, doi: 10.1007/s00521-020-05360-1.

Shaowu Zhu is a postgraduate at the College of Information and Network Security, People's Public Security University of China, Beijing, China. His current research interests include natural language processing, knowledge graph and deep learning.

Haichun Sun received her Ph.D. degree in computer software and theory from the Tongji University, Shanghai, China, in 2015. She is currently Associate Professor at the College of Information and Network Security, People's Public Security University of China, Beijing, China. She is Member of Professional Committee of the Internet Information Service of Chinese Association of Automation. Her current research interests include information service, Petri nets, and service-oriented computing. She has published over 10 papers on journals and conferences such as the IEEE Transactions on Systems, Man, and Cybernetics, the WISE 2014.

Hanying Jian is a postgraduate at the College of Public Security Management, People's Public Security University of China, Beijing, China. She is working on analysis of statistical characteristics in the public security domain.