

## Correlation temporal feature extraction network via residual network for English relation extraction

Ping Li<sup>1,\*</sup>

<sup>1</sup>Department of Public Instruction, Nanyang Medical College, Nanyang City 473000, China

### Abstract

In relation extraction, a major challenge is the absence of annotated samples. Relation extraction aims to extract the relationships between entity pairs from a large amount of unstructured data. To solve the above problems, this paper presents a new method for English relation extraction based on correlation temporal feature extraction network via residual network. Firstly, the attention mechanism and recurrent neural network are used to obtain the temporal features of English word correlation. Secondly, a multi-branch feature sensing convolutional neural network is constructed to obtain global and local temporal correlation features respectively. Residual network can dynamically reduce the influence of noise data and better extract the deep information of English text. Finally, the relation extraction is realized with Softmax classifier. Experimental results show that the proposed method can extract English relation effectively than other methods.

**Keywords:** English relation extraction, correlation temporal feature extraction network, residual network, Softmax.

Received on 09 November 2021, accepted on 19 November 2021, published on 19 November 2021

Copyright © 2021 Ping Li *et al.*, licensed to EAI. This is an open access article distributed under the terms of the [Creative Commons Attribution license](#), which permits unlimited use, distribution and reproduction in any medium so long as the original work is properly cited.

doi: 10.4108/eai.19-11-2021.172213

\*Corresponding author. Email: snowycry@qq.com

### 1. Introduction

With the rapid development of Internet technology, how to extract useful structured information from massive non-structured data has become a research hotspot in industry and academia at home and abroad. As the core branch of Information Extraction (IE), relation extraction (RE) aims to extract semantic relationships between given entity pairs from unstructured text, and has been widely used in knowledge graph, text summary, automatic question answering and other aspects [1-5]. The sentence in figure 1 contains entity 1: influenza [6] and entity 2: virus, whose goal is to predict the semantic relationship between the two entities *Cause-Effect*( $e_2, e_1$ ).

Sentence: "Avian <e1>influenza</e1> is an infectious disease of birds caused by type A strains of the influenza <e2>virus</e2>."		
Entity 1: influenza	Entity 2: virus	Relation: Cause-Effect( $e_2, e_1$ )

**Figure 1.** Example of relation extraction

In recent years, neural network based on Transformer [7,8], recurrent neural network (RNN) [9,10], convolutional neural network (CNN) [11-14], and their variants have been widely used in relation extraction tasks [15-17]. In relation extraction using neural networks, the word vector representation of each word input is fixed, but the actual

meaning of the word changes with the context. Some neural networks expect to learn contextual information when encoding words. However, without overall consideration of the sentence [18], it is difficult to achieve the purpose of learning context information. The Transformer structure's self attention mechanism (SAM) [19] can capture the correlation between any pair of words in a sentence. This method encodes words according to the correlation between words so that the word encoding contains contextual correlation information. RNN uses sequence units to store time information and has advantages in processing time sequence information. Bidirectional Long Short-term Memory (BiLSTM) [20], a typical RNN, uses both past and future time information of word sequences to encode words with contextual time information due to its bi-directional timing structure.

In this paper, SAM and BiLSTM are first used to encode words by taking the temporal characteristics of word correlation, so as to express the meaning of words in sentences more accurately. On the other hand, most neural networks only consider single-branch information flow. For example, Zhang et al. [21] used BiLSTM to encode time sequence of words to obtain contextual time sequence information. Nguyen et al. [22] used convolution kernels of different sizes to improve the local perception ability of CNN. However, the above methods are still difficult to obtain enough semantic features to extract relationships.

To solve this problem, this paper proposes a relation extraction method based on global and local feature-aware network (GLFN) via residual network. Based on the extracted temporal correlation features, the global and local feature-aware convolutional neural network (GLFCNN) is used to fully express the important semantic features of sentences. Specifically, aiming at GLFCNN, this paper firstly constructs multi-branch feature-aware convolutional neural networks (multi-branch feature-aware convolutional neural networks, (MFA-CNN), respectively. It senses the global and local correlation temporal features, splices and screens these two features. Finally, it combines with Softmax classifier, predicts the extracted semantic information. The relation extraction methods proposed in this paper based on global and local feature sensing networks mainly include correlated sequence feature extraction network (CSFEN), global and local feature sensing convolutional Neural Network (GLFCNN), and Prediction Output Network (PON).

(1) Correlation temporal feature extraction network (CSFEN): SAM and BiLSTM are used to obtain the correlation temporal feature of words. Specifically, the former is used to learn the correlation features of any word in a sentence, while the latter performs sequence encoding for words containing correlation features to obtain sequence information.

(2) Global and local feature sensing convolutional neural network (GLFCNN): Firstly, the multi-branch feature

sensing convolutional neural network (MFA-CNN) is constructed to respectively sense global and local temporal correlation features, and splicing and filtering these two features to comprehensively represent important semantic features of sentences.

(3) Predictive output network (PON): The full connection layer is used to map the output of GLFCNN to the classification space, and the relationship between entity pairs is predicted by combining with Softmax classifier.

In this paper, a relation extraction method based on global and local feature perception networks is proposed. Firstly, SAM and BiLSTM are used to extract the temporal features of word correlation, and GLFCNN is constructed to obtain the semantic features of different levels of words. The main contributions are as follows:

(1) SAM and BiLSTM are used to extract the temporal features of word correlation, which is more precise representing the meaning of words in a sentence.

(2) GLFCNN is constructed to respectively sense global and local correlation time sequence features to avoid the mutual influence of global and local perception. The global and local perceptual features are spliced and screened to fully represent the important semantic features of sentences.

(3) The validity of the proposed relation extraction method is verified on the standard SemEval-2010 Task 8 and KBP37 datasets, with F1 reaching 86.1% and 64.9%, respectively.

## 2. Related works

The research on relation extraction mainly includes the methods based on artificial rules in the early days and the artificial intelligence methods based on neural networks which have developed rapidly in recent years. The former mainly uses NLP tools or manually designs different kernels to select features [23,24]. Such methods mainly have the problems of error propagation caused by NLP tools [25] and limitations of manual experience.

With the rapid development of neural network, scholars at home and abroad try to use it for relation extraction. Zeng et al. [26] used CNN to extract the features of word and sentence level, and combined with the relative position information between each word and entity pair in the sentence to extract the relationship, which was more efficient than the method based on artificial rules. However, due to the limitations of CNN structure [27], this method cannot give full play to its performance in the face of remote entity pairs. RNN is good at processing temporal information and has advantages in acquiring word long-distance dependence [28]. Tymoshenko et al. [29] used RNN to sequence encode sentences and obtain semantic features including contextual sequence information. Guo et al. [30] captured the distance phase in sentences with the help of BiLSTM's powerful memory preservation ability. It

could time dependent information between distant words. However, the above methods only consider the information of short distance dependence or time series dependence of words, without considering the correlation between words, so it is difficult to describe the overall linguistic information of words in a sentence.

In this paper, SAM and BiLSTM are used to extract the temporal characteristics of word correlation to encode the word, so that it can more accurately represent the overall semantic information of the context of the word in the English sentence.

In terms of feature perception, Nguyen et al. [31] used CNN to extract local features, combined with maximum pooling operation to reduce redundancy and sense important features. However, as convolution kernels of fixed size would limit the perception range of CNN, Liu et al. [32] used convolution kernels of different sizes to improve CNN's perception ability of local features. Zhang et al. [33] used Gaussian attention to improve SAM's perception of local information, so that local information near the central word could get more attention, and further applied to natural language reasoning tasks. However, most of the above networks adopt single-branch structure, which can only perceive a single information flow feature, and lack comprehensive attention to the global and local features of sentences, so it is difficult to obtain enough semantic features to extract relationships.

Aiming at this problem, this paper proposes a network based on global and local characteristics of the awareness of relationship extraction method, based on the correlation of temporal feature extracting, further builds MFA-CNN, sense of global and local correlation sequence characteristics respectively, avoids global and local perception influence each other. The global and local perceptual features are spliced and screened to represent the important semantics of sentences.

### 3. Residual network (ResNet)

The learning objective of residual network is residual  $F(x) = H(x) - x$ . The skipping learning structure ignores that the middle layer directly connects the low-level

representation with the high-level representation, and greatly alleviates the problem of gradient disappearance that plights deep networks. In the designed model, we use shortcut connections to build the residual convolution structure. Each residual block contains two convolution layers, each convolution layer is followed by a nonlinear layer, and the activation function is set as ReLU activation function [34-36]. All convolution windows are  $l \times d$  in size, and the output remains the same size as the input after the padding. After the first layer of convolution, the output result of the  $i$ -th window is as follows:

$$\tilde{c}_i = f(w_1 c_{ii+f} + b_1) \quad (1)$$

After the second convolution, the output of the  $i$ -th window is as follows:

$$\tilde{c}_i = f(w_2 \tilde{c}_{ii+f} + b_2) \quad (2)$$

$b_1 \in R^{d_c}$  and  $b_2$  are the offsets of each layer. The objectives of residual block learning are as follows:

$$c = c + \hat{c} \quad (3)$$

The entire ResNet consists of four of the above residual blocks to form a cascade architecture, and the final output is  $c$ , the extracted more abstract relational information.

### 4. Relation extraction methods based on both global and local feature-sensing networks

This section introduces the methods of relation extraction based on global and local feature sensing networks. As shown in figure 2, CSFEN uses SAM and BiLSTM to extract the temporal features of correlation between words in sentences. GLFCNN performs global and local perception of correlation temporal features and obtains multi-level semantic information. Finally, PON is used to predict the extracted semantic information.

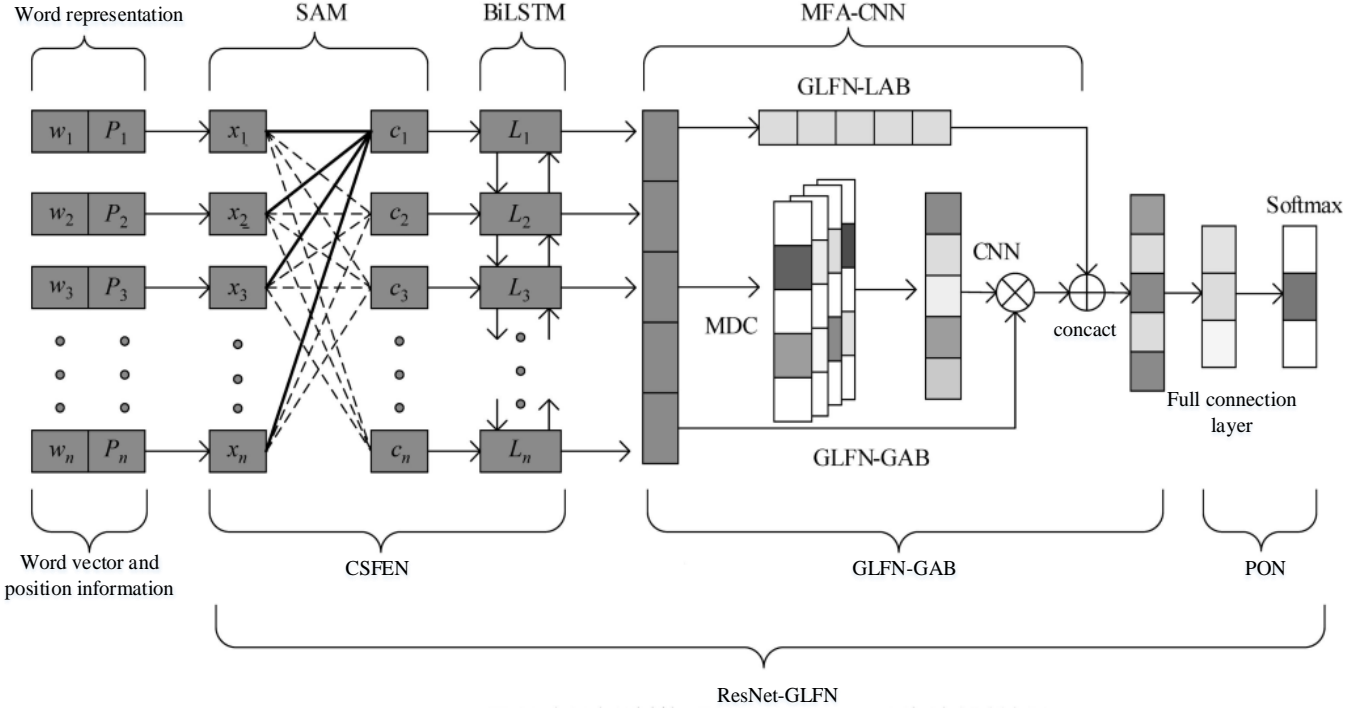


Figure 2. ResNet-GLFN network structure

### 4.1. Correlation temporal feature extraction network

CSFEN first uses the pre-trained GloVe word vector and the relative position information [37] containing each word and entity pair to represent words. The input of sentence is  $X = \{x_1, \dots, x_n\}$ . However, the actual meaning of a word tends to change from context to context, and the input sentence X is fixed. In order to make the encoding of the representation words contain contextual relevance information and describe the contextual semantics of the sentence, this paper uses SAM to learn the relevance features of each word in the sentence. SAM is mainly composed of two parts: dot attention and multiple-head attention. Dot product attention consists of three matrices: query matrix Q, key matrix K and value matrix V. The weights of the matrix are automatically updated by network training. The specific realization of dot attention is shown in formula (4).

$$Attention(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_x}}\right)V \quad (4)$$

In order to obtain the representation of attention weight in different sub-spaces, multiplex attention maps attention weight to multiple sub-spaces for learning, and then splices the information learned in different spaces. Multiple attention is shown in equations (5) and (6).

$$head_i = Attention(QW_i^Q, KW_i^K, VW_i^V) \quad (5)$$

$$MultiHead(Q, K, V) \quad (6)$$

$$= Concat(head_1, \dots, head_n)W^O$$

For the input sentence X, SAM is used in this paper to obtain the sentence representation  $C=SA(X)$  containing context-relevant information. SAM is used to learn the correlation between words in a sentence, but ignores the temporal information between words in a sentence. In order to extract the temporal sequence information in the positive and negative directions of context, BiLSTM is used to further represent the temporal sequence semantics of sentence correlation, that is, the temporal sequence features of the output sentence correlation  $L=BiLSTM(C)$ .

### 4.2. Global and local feature-aware convolutional neural network

In this paper, GLFCNN is used to comprehensively represent the important semantic features of sentences on the basis of obtaining the temporal features of sentence correlation. Specifically, aiming at GLFCNN, MFA-CNN was firstly constructed, including global feature perception branch (GLFN-GAB) and local feature perception branch (GLFN-LAB), respectively sensing global and local correlation time series features. GLFN-LAB uses multichannel dilated convolution (MDC) to

obtain the distribution information of global features to describe the importance of the temporal features of the correlation, so as to extract the temporal features of global correlation. GLFN-LAB uses convolution operation to obtain n-gram information of words and realize the perception of local correlation temporal features.

GLFN-LAB uses MDC to map correlation temporal features to multiple representation spaces to learn the distribution information of features in different spaces. The MDC is implemented as shown in formula (7) and (8).

$$Aware_i = DC(L) \quad (7)$$

$$\begin{aligned} Multichannel(L) = \\ Concat(Aware_1, \dots, Aware_t)W^A \end{aligned} \quad (8)$$

Where DC is the empty convolution operation. t is the number of MDC channels, which determines the number of representation spaces. This paper will discuss the influence of the value of t on network performance in the experimental part. In this paper, Gaussian Error Linear Units (GELU) is used to perform nonlinear transformation on distribution information after obtaining the distribution information of correlation time series features of different representation spaces to improve the nonlinear expression ability of the network. The calculation method is shown in formula (9).

$$GELU = xP(X \leq x) = x\Phi(x) \quad (9)$$

Where  $\Phi(x)$  represents the cumulative distribution of the Gaussian distribution of x. In order to obtain the distribution information of the global features, the global feature-aware weight matrix  $M_{aware}$  is shown as formula (10).

$$M_{aware}(L) = \sigma(GELU(Multichannel(L))W^a) \quad (10)$$

Where  $\sigma$  is the Sigmoid activation function that maps information to a range of 0 to 1.  $W^a$  represents the weight of GLFN-GAB. The obtained  $M_{aware}$  represents the importance distribution of global features to perceive the global correlation temporal features.

GLFN-GAB uses  $k$  convolution kernels to explore the local N-gram information N-gram(L) of order features for correlation, its calculation method is shown in equation (11).

$$N - gram(L) = W^s L + b^s \quad (11)$$

Where  $W^s$  represents the weight of the convolution kernel.  $b^s$  is the bias of the convolution kernel, and its value is automatically updated by network training. The obtained N-gram(L) is used to represent the local information of features, and further spliced with the global feature  $M_{aware} \Theta L$  to fuse the temporal sequence

features of global and local correlation. The calculation method is shown in formula (12).

$$V = M_{aware} \Theta L + N - gram(L) \quad (12)$$

Where  $\Theta$  is the element by element multiplication. The global and local correlation temporal features of splicing  $V \in R^{n \times k}$  can be refined as  $V = \{v_1, v_2, \dots, v_k\}$ . The  $v_i \in R^n$  represents the sentence representation of different feature channels.  $n$  and  $k$  represent the length of the sentence and the number of channels of the feature, respectively. In this paper, furthermore, the maximum value of  $v_i$  obtained is pooled in this paper to realize feature screening, and only the most important information is retained in each feature channel, so as to reduce feature dimension and remove redundancy. The obtained sequence features of global and local correlation in this paper are shown in formula (13).

$$O = [\max(v_1); \max(v_2); \dots; \max(v_k)] \quad (13)$$

### 4.3. Predicted output network (PON)

In this paper, the fully connected network is used to map the final perceived global and local correlation temporal features into  $|y|$  classification label space to obtain the prediction information  $\hat{G} \in R^{|y|}$ , as shown in equation (14).

$$\hat{G} = W_c O + b_c \quad (14)$$

Where  $W_c$  and  $b_c$  represent the weight and bias of the full connection layer, respectively. In this paper, Softmax classifier is used to make relation prediction of prediction information  $\hat{G}$ , and the calculation formula is shown in equation (15).

$$\hat{y}_i = \frac{\exp(\hat{G}_i)}{\sum_{j=1}^{|y|} \exp(\hat{G}_j)} \quad (15)$$

Where  $\hat{y}_i$  is the actual predicted output. In this paper, cross entropy is used as the loss function of GLFN and L2 regularization is used to punish the network parameters to improve the generalization performance of the network. The objective function of the network is shown in equation (16).

$$J(\theta) = -\frac{1}{|y|} \sum_{i=1}^{|y|} y_i \log(\hat{y}_i) + \lambda \|\theta\|_F^2 \quad (16)$$

Where  $y_i$  is the expected output.  $\lambda$  is the hyperparameter of L2 regularization.  $\theta$  is trainable parameter in the network, which is automatically updated



by training. The training goal of neural network is to minimize  $J(\theta)$ , so that the actual prediction output  $\hat{y}_i$  is close to the expected prediction output  $y_i$ .

## 5. Experiment

### 5.1. Experimental data and evaluation indexes

In this paper, the pre-trained GloVe word vector is used to represent words, and two standard data sets of SemEVAL-2010 Task 8 and KBP 37 are used to verify the validity of the proposed method. Semeval-2010 Task 8 data set includes 8000 training samples and 2717 test samples, including cause-effect, instrument-agency, product-producer, content-container, entity-Origin, Entity-Destination, Component-whole, member-collection, message-topic, and ‘‘Other’’ relationships. Since the first 9 relationships are directional, 19 relationships of this dataset are considered in this paper. KBP37 data set includes 15917 training samples and 3405 test samples, including 18 kinds of directional relations and a special relation ‘‘No\_Relation’’. Therefore, a total of 37 relationships of KBP 37 data set are considered. In addition, we use macro average F value to evaluate the effectiveness of the relation extraction method.

### 5.2. Experimental results and analysis

Since the channel number  $t$  of GLFN-GAB represents the number of space, selecting an appropriate  $t$  can effectively improve the perception ability of GLFN-GAB. In this paper, different number of channel numbers  $t$  are selected to evaluate the macro average F1 values of only global sensing branch (GLFN-GAB) and GLFCNN (including global sensing branch GLFN-GAB and local sensing branch GLFN-LAB) respectively. The performance of the above two methods on the SemEVAL-2010 Task 8 and KBP37 datasets is shown in table 1 and table 2 respectively.

Table 1. The effect of different  $t$  values on F1 of GLFN-GAB and GLFCNN on the SemEval-2010Task8 dataset

Method	Channel $t$					
	1	2	4	8	16	32
GLFN-	84.7	84.9	85.3	85.2	84.9	84.9

GAB						
GLFCNN	85.7	85.8	<b>86.2</b>	85.9	85.7	85.7

Table 2. The effect of different  $t$  values on F1 of GLFN-GAB and GLFCNN on the KBP37 dataset

Method	Channel $t$					
	1	2	4	8	16	32
GLFN-GAB	64.2	64.4	64.4	64.5	64.3	63.8
GLFCNN	64.5	64.6	<b>65.1</b>	64.5	64.4	63.9

As can be seen from Table 1, GLFCNN performs better than GLFN-GAB for all channel number  $t$ . As can be seen from Table 2, except for  $t=8$ , GLFCNN and GLFN-GAB have the same F1 values, GLFCNN also performs better than GLFN-GAB in other channels, and when  $t=4$ , GLFCNN achieves the highest F1 value in both data sets.

After obtaining the best number of channel  $t$ , this paper also compares GLFN-GAB, GLFN-LAB, GLFCNN, and the relation extraction method of the default GLFCNN (Non GLFCNN). Figures 3 and 4 show the training process on the SemEval-2010 Task8 and KBP37 data sets using the four methods above. Non GLFCNN directly takes the output of the Relevant Timing Speciality Network (CSFEN) as input to the Predictive Output Network (PON).

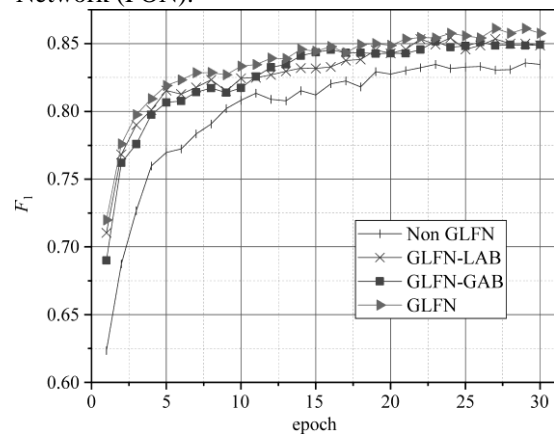
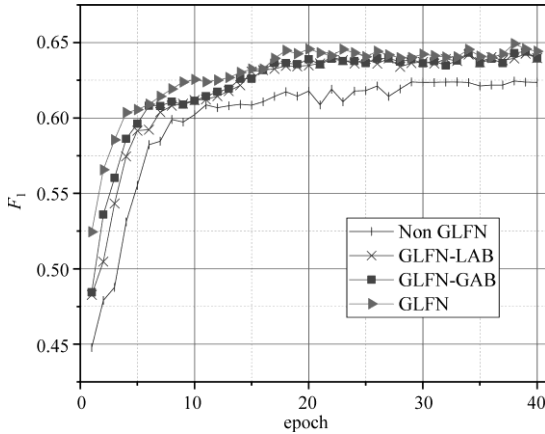


Figure 3. Training procedure for four methods on the SemEval-2010 Task 8 dataset.



**Figure 4.** Training procedure for four methods on the KBP37 dataset.

As shown in figures 3 and 4, from the general trend, F1 value of non-GLFN, GLFN-LAB, GLFN-GAB and GLFN increases rapidly with the number of iterations in the first several stages. In the later stages, the F1 values of these four methods gradually increase, but there are certain fluctuations.

Compared with non-GLFN, GLFN-GAB and GLFN-LAB, GLFN obtains the highest F1 value at round 27 and 38 on SemEval 2010 Task and KBP37 dataset, respectively. This is because GLFN uses global and local feature sensing convolutional neural network to obtain semantic features at different levels, comprehensively represent the important semantics of sentences, and improve the accuracy of relation extraction. Meanwhile, table 3 and table 4 are used to show the highest F1 values of the above four methods on SemEval2010 Task 8 and KBP37 data sets respectively.

**Table 3.** The highest F1 values of four methods on the Semeval-2010 Task 8 dataset

Method	F1/%
non-GLFN	83.7
GLFN-GAB	85.3
GLFN-LAB	85.5
GLFN	<b>86.2</b>

**Table 4.** The highest F1 values of four methods on the KBP37 dataset

Method	F1/%
non-GLFN	62.5
GLFN-GAB	64.5
GLFN-LAB	64.3
GLFN	<b>65.1</b>

As can be seen from Table 3 and Table 4, GLFN has the highest F1 value comparing with non-GLFN, GLFN-GAB and GLFN-LAB. On the SemEval-2010 Task 8 and KBP37 datasets, GLFN's F1 reaches 86.2% and 65.1%, respectively. In addition, Table 5 shows GLFN and 14 mainstream English relation extraction methods (CNN [38], CNN+PF [39], multi-CNN [40], CR-CNN, Attention-CNN [41], Bi-LSTM, BiLSTM+feature, ATT-BiLSTM, HierLSTM [42], BiLSTM-Attention, FORESTFT-DDCNN, LST-AGCN, GCN and S-Att) for the highest F1 values on the SemEval-2010 Task8 dataset.

**Table 5.** F1 values of 15 methods on the SemEval-2010 Task8 dataset

Method	F1/%
GCN	79.7
Bi-LSTM	80.2
CNN	80.3
S-Att	80.3
CNN+PF	82.8
multi-CNN	82.9
CR-CNN	84.2
Attention-CNN	84.4
BiLSTM+feature	84.4
ATT-BiLSTM	84.1
HierLSTM	84.4
BiLSTM-Attention	85.3
FORESTFT-DDCNN	85.6
LST-AGCN	86.1
Proposed	86.2

Finally, for KBP37 data set, this paper also compares CNN, RNN, BiLSTM-CNN, Block+CNN, Ranking CNN, Att-RCNN, Bi-SDP-Att and GLFN. Table 6 shows the highest F1 values of the 8 methods on KBP37 data set.

Table 6. F1 values of 8 methods on KBP37 data set

Method	F1/%
CNN	55.9
RNN	58.9
BiLSTM-CNN	60.2
Block+CNN	61.0
Ranking CNN	61.4
Att-RCNN	61.9
Bi-SDP-Att	64.5
GLFN	<b>65.2</b>

As shown in Table 5, in the SemEVAL-2010 Task 8 data set, the proposed GLFN in this paper has the highest F1 value comparing to the 14 mainstream relation extraction methods. This is because GLFN can respectively sense the global and local correlation temporal sequence features to avoid the mutual influence of global and local perceptions, and splicing and filtering the two features to filter out the features with low contribution, so as to comprehensively represent the important semantics of sentences and improve the accuracy of English relation extraction [43-46].

LST-AGCN based on grammar graph enriches sentence information from the perspective of syntax dependencies, and thus has a higher F1 value compared with other mainstream methods. However, this method requires NLP tool to build syntax dependency tree, and its performance is highly dependent on NLP tool. As shown in Table 6, on the KBP 37 data set, GLFN has the highest F1 value compared with CNN, RNN, BiLSTM-CNN, Ranking CNN, AttRCNN, Block+CNN and Bi-SDP-ATT. Compared with Bi-SDP-ATT with the second highest F1 value, the F1 value of GLFN increases by 0.7%. Moreover, Bi-SDP-ATT needs to construct bidirectional short-dependent path and corresponding attention mechanism, and its structure is relatively complex. In the other 6 methods, Att-RCNN has a higher F1 value, and the GLFN proposed in this paper has a higher F1 value than that by 3.3%.

## 6. Conclusion

At present, most neural networks only consider single-branch information flow, it is difficult to obtain enough semantic features for relation extraction. To solve this question, a GLFN-based English relation extraction method is proposed. The method first uses SAM and BiLSTM to obtain the relevance timing characteristics of the words. Second, build ResNet to obtain global and local correlation timing characteristics, respectively, to avoid the interaction between global and local perception. Further, the two features are stitched and filtered to show the essential semantic features of the sentence in a full-faceted table. Finally, it uses the Softmax to classify the relation extraction. To verify the validity of the proposed method, extensive experiments are conducted on the standard SemEval-2010Task8 and KBP 37 datasets. The experimental results show that the F1 value of the proposed method in this paper is 86.2% and 65.1% respectively, which is better than the mainstream relation extraction methods between convolution-based neural network and circulatory neural network. In the future, we will verify the validity of the proposed method on more language word vector models. At the same time, we will target the unstructured data in different fields, extract the relationship between entity pairs, and build the entity relationship triple, so as to establish a domain structured data base for further use in knowledge mapping, automatic question-and-answer and other tasks. In addition, we will also develop a convenient user interface to provide users with a better experience.

## Acknowledgements.

The author is grateful to the reviewers for their anonymous reviews.

## References

- [1] Juan Gómez-Romero, et al. Visualizing large knowledge graphs: A performance analysis - ScienceDirect[J]. Future Generation Computer Systems, 2018, 89:224-238.
- [2] Q. Xu, X. Wang, J. Li, Q. Zhang and L. Chai, "Distributed Subgraph Matching on Big Knowledge Graphs Using Pregel," in IEEE Access, vol. 7, pp. 116453-116464, 2019, doi: 10.1109/ACCESS.2019.2936465.
- [3] Shoulin Yin, Hang Li, Asif Ali Laghari, et al. A Bagging Strategy-Based Kernel Extreme Learning Machine for Complex Network Intrusion Detection[J]. EAI Endorsed Transactions on Scalable Information Systems. 21(33), e8, 2021. <http://dx.doi.org/10.4108/eai.6-10-2021.171247>
- [4] Li Yan, Xiaowei Wang, and Shoulin Yin. Campus Garbage Image Classification Algorithm Based on New Attention Mechanism [J]. International Journal of



- Electronics and Information Engineering. Vol. 13, No. 4, pp. 131-141, 2021.
- [5] Liu, J., Zhang, J. & Yin, S. Hybrid chaotic system-oriented artificial fish swarm neural network for image encryption. *Evolutionary Intelligence* (2021). <https://doi.org/10.1007/s12065-021-00643-5>
- [6] Wei Song, Xinfu Zhu. Global and Local Feature-Aware Network for Relation Extraction [J]. *Journal of Chinese Information Processing*. 34(11), 96-103, 2020. (In Chinese)
- [7] Liserre M, Buticchi G, Andresen M, et al. The Smart Transformer: Impact on the Electric Grid and Technology Challenges[J]. *IEEE Industrial Electronics Magazine*, 2016, 10(2):46-58.
- [8] C. Lin and S. Lucey, "Inverse Compositional Spatial Transformer Networks," 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, pp. 2252-2260, doi: 10.1109/CVPR.2017.242.
- [9] Su L, Zhou L. Exponential synchronization of memristor-based recurrent neural networks with multi-proportional delays[J]. *Neural Computing and Applications*, 2019, 31(5).
- [10] Deng H, Zhang L, Wang L. Global context-dependent recurrent neural network language model with sparse feature learning[J]. *Neural computing & applications*, 31, pp. 999-1011, 2019.
- [11] Qingwu Shi, Shoulin Yin, Kun Wang, Lin Teng and Hang Li. Multichannel convolutional neural network-based fuzzy active contour model for medical image segmentation. *Evolving Systems* (2021). <https://doi.org/10.1007/s12530-021-09392-3>
- [12] Shoulin Yin, Hang Li, Shahid Karim, and Yang Sun. ECID: Elliptic Curve Identity-based Blind Signature Scheme[J]. *International Journal of Network Security*, 23, No. 1, pp. 9-13, 2021.
- [13] Jisi A and Shoulin Yin. A New Feature Fusion Network for Student Behavior Recognition in Education [J]. *Journal of Applied Science and Engineering*. vol. 24, no. 2, pp.133-140, 2021.
- [14] Ting-Ting Gao, Hang Li, and Shou-Lin Yin. Adaptive Convolutional Neural Network-based Information Fusion for Facial Expression Recognition [J]. *International Journal of Electronics and Information Engineering*. Vol. 13, No. 1, pp. 17-23, 2021.
- [15] Xie C, Kumar A. Finger vein identification using Convolutional Neural Network and supervised discrete hashing[J]. *Pattern Recognition Letters*, 2019, 119:148-156.
- [16] J. Sulam, A. Aberdam, A. Beck and M. Elad, "On Multi-Layer Basis Pursuit, Efficient Algorithms and Convolutional Neural Networks," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 42, no. 8, pp. 1968-1980, 1 Aug. 2020, doi: 10.1109/TPAMI.2019.2904255.
- [17] S. Yin and H. Li. Hot Region Selection Based on Selective Search and Modified Fuzzy C-Means in Remote Sensing Images[J]. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 13, pp. 5862-5871, 2020, doi: 10.1109/JSTARS.2020.3025582.
- [18] Kim J, Kim H. Applying Recurrent Neural Network to Intrusion Detection with Hessian Free Optimization[C]// *International Workshop on Information Security Applications*. Springer International Publishing, 2015.
- [19] Stevens C, Lauinger B, Neville H. Differences in the neural mechanisms of selective attention in children from different socioeconomic backgrounds: an event-related brain potential study[J]. *Developmental Science*, 2010, 12(4):634-646.
- [20] Dhieb T, Boubaker H, Ouarda W, et al. Deep bidirectional long short-term memory for online multilingual writer identification based on an extended Beta-elliptic model and fuzzy elementary perceptual codes[J]. *Multimedia Tools and Applications*, 2021(1).
- [21] Zhang S, Zheng D, Hu X, et al. Bidirectional long short-term memory networks for relation classification[C]//*Proceedings of the 29th Pacific Asia Conference on Language, Information and Computation*, 2015: 73-78.
- [22] Nguyen T H, Grishman R. Relation extraction: Perspective from convolutional neural networks [C]// *Proceedings of the 1st Workshop on Vector Space Modeling for Natural Language Processing*, 2015: 3948
- [23] Shoulin Yin, Hang Li, Desheng Liu and Shahid Karim. Active Contour Modal Based on Density-oriented BIRCH Clustering Method for Medical Image Segmentation [J]. *Multimedia Tools and Applications*. Vol. 79, pp. 31049-31068, 2020.
- [24] Yin, S., Li, H. & Teng, L. Airport Detection Based on Improved Faster RCNN in Large Scale Remote Sensing Images [J]. *Sensing and Imaging*, vol. 21, 2020. <https://doi.org/10.1007/s11220-020-00314-2>
- [25] Shoulin Yin, Jie Liu, and Lin Teng. A Sequential Cipher Algorithm Based on Feedback Discrete Hopfield Neural Network and Logistic Chaotic Sequence [J]. *International Journal of Network Security*. Vol. 22, No. 5, pp. 869-873, 2020.
- [26] Zhang D, Wang D. Relation Classification via Recurrent Neural Network[J]. *Computer Science*, 2015. arXiv:1508.01006
- [27] B. Shuai, Z. Zuo, B. Wang and G. Wang, "Scene Segmentation with DAG-Recurrent Neural Networks," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 40, no. 6, pp. 1480-1493, 1 June 2018, doi: 10.1109/TPAMI.2017.2712691.
- [28] Wang X, Lu C, Wang C, et al. Methods for testing the performance of long-distance wireless power transmission systems[J]. *EURASIP Journal on Wireless Communications and Networking*, 2020, 2020(1):1-13.

- [29] Tymoshenko K, Moschitti A. Shallow and Deep Syntactic/Semantic Structures for Passage Reranking in Question-Answering Systems[J]. *ACM Transactions on Information Systems*, 2019, 37(1):8.1-8.38.
- [30] Guo X, Hao X, Tang Z, et al. ACE-ADP: Adversarial Contextual Embeddings Based Named Entity Recognition for Agricultural Diseases and Pests[J]. *Agriculture*, 2021, 11.
- [31] T. -N. Nguyen, C. -N. -N. Hoang, T. -S. Le and T. A. Tran, "A System for Text Extraction in Complex-Background Document Images," 2019 International Conference on Advanced Computing and Applications (ACOMP), 2019, pp. 65-69, doi: 10.1109/ACOMP.2019.00017.
- [32] Liu G, Zhang K, Lv M. ASKS: Convolution with Any-Shape Kernels for Efficient Neural Networks[J]. *Neurocomputing*, 446, 32-49, 2021.
- [33] Zhang X, Chen F, Huang R. A Combination of RNN and CNN for Attention-based Relation Classification[J]. *Procedia Computer Science*, 2018, 131:911-917.
- [34] Xiaowei Wang, Shoulin Yin, Hang Li. A Network Intrusion Detection Method Based on Deep Multi-scale Convolutional Neural Network[J]. *International Journal of Wireless Information Networks*. 27(4), 503-517, 2020.
- [35] Xiaowei Wang, Shoulin Yin, Ke Sun, Hang Li, Jie Liu and Shahid Karim. GKFC-CNN: Modified Gaussian Kernel Fuzzy C-means and Convolutional Neural Network for Apple Segmentation and Recognition [J]. *Journal of Applied Science and Engineering*, vol. 23, no. 3, pp. 555-561, 2020.
- [36] Yin, S., Li, H. GSAPSO-MQC:medical image encryption based on genetic simulated annealing particle swarm optimization and modified quantum chaos system. *Evolutionary Intelligence* (2020). doi: 10.1007/s12065-020-00440-6
- [37] A. Safaei and M. N. Mahyuddin, "Adaptive Cooperative Localization Using Relative Position Estimation for Networked Systems With Minimum Number of Communication Links," in *IEEE Access*, vol. 7, pp. 32368-32382, 2019, doi: 10.1109/ACCESS.2019.2903219.
- [38] Peng Y, Rios A, Kavuluru R, et al. Chemical-protein relation extraction with ensembles of SVM, CNN, and RNN models[J]. 2018. arXiv:1802.01255
- [39] X. Zhao, W. Li, Y. Zhang, S. Chang, Z. Feng and P. Zhang, "Aggregated Residual Dilation-Based Feature Pyramid Network for Object Detection," in *IEEE Access*, vol. 7, pp. 134014-134027, 2019, doi: 10.1109/ACCESS.2019.2941892.
- [40] R. Fan, R. Feng, L. Wang, J. Yan and X. Zhang, "Semi-MCNN: A Semisupervised Multi-CNN Ensemble Learning Method for Urban Land Cover Classification Using Submeter HRRS Images," in *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 13, pp. 4973-4987, 2020, doi: 10.1109/JSTARS.2020.3019410.
- [41] Nie W, K Wang, Liang Q, et al. Panorama based on multi-channel-attention CNN for 3D model recognition[J]. *Multimedia Systems*, 25, 655–662, 2019.
- [42] Xiao M, Liu C. Semantic relation classification via hierarchical recurrent neural network with attention [C]//*Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*, 2016: 1254-1263.
- [43] Laghari A A, Laghari M A. Quality of experience assessment of calling services in social network[J]. *ICT Express*, 2021(2).
- [44] Laghari A A, Laghari K, Memon K A, et al. Quality of Experience (QoE) Assessment of Games on workstations and Mobile[J]. *Entertainment Computing*, 2020, 34:100362.
- [45] A. A. Laghari, H. He, A. Khan, N. Kumar and R. Kharel, "Quality of Experience Framework for Cloud Computing (QoC)," in *IEEE Access*, vol. 6, pp. 64876-64890, 2018, doi: 10.1109/ACCESS.2018.2865967.
- [46] Laghari, A.A., Jumani, A.K. & Laghari, R.A. Review and State of Art of Fog Computing. *Arch Computat Methods Eng* 28, 3631–3643 (2021). <https://doi.org/10.1007/s11831-020-09517-y>