Capsule Networks With Word-Attention Dynamic Routing for Cultural Relics Relation Extraction

MIN ZHANG\textsuperscript{1,2} AND GUOHUA GENG\textsuperscript{1}

\textsuperscript{1}School of Information Science and Technology, Northwest University, Xi’an 710127, China
\textsuperscript{2}College of Computer, Weinan Normal University, Weinan 714099, China

Corresponding author: Guohua Geng (ghgeng@nwu.edu.cn)

This work was supported in part by the National Nature Science Foundation of China under Grant 616731015, Grant 61731015, and Grant 61802311, and in part by the National Key Research and Development Program of China under Grant 2017YFB1402103.

ABSTRACT

Online museums and online cultural relic information provide abundant data for relation extraction research. However, in the relation extraction task of modelling space information, spatially insensitive methods of convolutional neural networks and long short term memory network in most current works still remain challenging in rich text structures, which makes models difficult to encode effectively and lacks the ability of text expression. To address this issue, we propose a framework named WAtt-Capsnet (the capsule network with word-attention dynamic routing), which is based on capsule networks with word-attention dynamic routing for the relation extraction task of online cultural relic data for capturing richer instantiation features. We further present combination embedding for capturing the characteristic information of Chinese sentences by considering the contribution of word embedding, parts of speech, character embedding and the position of words to capture rich internal structure information of sentences. More importantly, to reduce the decay of useful information in long sentences, we propose a routing algorithm based on a word-attention mechanism to focus on informative words. The experimental results demonstrate that the proposed method achieves significant performance for the relation extraction task of online cultural relic data.

INDEX TERMS

Capsule networks, cultural relics, dynamic routing, relation extraction, word-attention mechanism.

I. INTRODUCTION

Online museums and online cultural relic information provide better resources to help people understand cultural relics and provide data for information extraction research, which has great significance in identifying the semantic relation between the cultural relic entities in online cultural relic data for various applications related to cultural relics, such as cultural relic knowledge graphs and cultural relic retrieval.

Relation extraction (RE) is one of the important steps in knowledge extraction, the main task of which is to extract the semantic relation between entities in the text to obtain a triple (entity 1, relation, entity 2). These relations exist between cultural relic entities, such as, cultural relic name (CRN) and cultural relic dynasty (CRD), CRN and unearthed date (UD), CRN and unearthed location (UL), CRN and museum collection (MC).

Examples of semantic relations are presented in Table 1. The one of the existed relations between two entities in sentence S1 is the cultural relic unearth date between CRN and UD, e1 = “Three Sheep Bronze Lei” and e2 = “1977”, and the second relation is the cultural relic unearth location between CRN and UL, e1 = “Three Sheep Bronze Lei” and e2 = “Liu Jiahe, Pinggu, Beijing”. In sentence S2, the existed relation between CRN and CRD is the cultural relic dynasty, e1 = “the Lei” and e2 = “Shang dynasty”. In the last sentence, the existed relation is the cultural relic museum collection between CRN and MC, e1 = “Three Sheep Bronze Lei” and e2 = “the capital museum”.

In previous relation extraction studies, researchers have applied many different traditional models (e.g., feature-based methods [1] and kernel function-based methods [2]). Although these methods have an effect on the relation extraction task, the feature-based approach does not take full advantage of the context structure information of the entities, and the training data in the model mainly rely on manual labelling, leading to many incorrectly labelled data in the
The training speed and testing speed of the kernel function-based methods are slow and not suitable for processing large scale data [3].

Deep neural networks can automatically learn the underlying feature information and alleviate the laborious task of manual annotation [4]. In recent years, many researchers have applied neural network models to relation extraction tasks, such as, convolutional neural networks (CNNs) [5], recurrent neural networks [6], and bidirectional long-term short-term memory (BiLSTM) [7]. These systems predict one relation for a single entity pair in a sentence. However, multiple entity pairs may commonly appear in a sentence and describe more than one relation, so it is also beneficial to focus on other relations when extracting relations in the context. Nevertheless, when modelling space information, spatially insensitive methods of CNN and LSTM are inevitably limited by rich text structures, which makes them difficult to encode effectively and lack the ability of text expression.

Hinton et al. [8] proposed a capsule network, which replaced the single neuron node of the traditional neural network with a neuron vector to avoid the representational limitations of CNNs and RNNs. Sabour et al. [9] proposed a dynamic routing algorithm to replace the max-pooling of CNN or RNN. The length of a capsule indicates the significance of features, and the direction of a capsule expresses the specific property of a feature. In this paper, we utilize the capsule network with dynamic routing to extract one or multiple relations between multiple entity pairs in a sentence. Word representation is considered to be an important factor influencing the performance of relation extraction. Recent researches have shown that word embedding induced from a large corpus could effectively capture the semantic and syntactic information of words [10]. Character-based word embedding can improve the performance of natural language processing tasks by capturing unknown words [11]. Because Chinese cultural relic data often contain unknown words and consider the importance of context characteristics, we combine word embedding with character embedding as part of the word representations. In addition, the part of speech (POS) of words and the position of words are also incorporated into our model because they carry critical context information for key characters and highlight the impact of entities.

In the task of relationship extraction in the field of cultural relics, the sparsity of relation and irrelevant words in the cultural relics data are existed. It is pivotal to extract contextual information of long sentence. In addition, irrelevant words in a sentence have a great influence on the performance of the model for the task of relation extraction [12]. Effectively removing noise from sentences remains a key issue in existing relation extraction research. Half of the sentences in original data of cultural relics are longer than 50 words. Even though multiple relations might occur in a sentence, this still means that there are many irrelevant words. Since deep neural networks with attention mechanisms have achieved remarkable results in natural language processing tasks, some people have introduced attention mechanisms into relational extraction tasks for mitigating noise data [13], [14]. However, these attention mechanism models are all based on neural networks of different depths, adding a single-layer attention mechanism of different feature levels into the input vector. To further capture the relevance of context and reduce the decay of useful information in long sentences, we devise a word-attention mechanism in a dynamic routing algorithm to effectively search for capsules containing correlation characteristics.

We propose a relation extraction model named WAtt-Capsnet (the capsule network with word-attention dynamic routing), based on capsule networks with word-attention dynamic routing for the relation extraction task of online cultural relic data. In our method, the capsule network can capture richer instantiation features, while the combination embedding as the model input are used to capture critical context information. To reduce the decay of useful information in long sentences and further capture the relevance of context, we propose a routing algorithm based on a word-attention mechanism to focus on informative words. We conduct comparative experiments between our model and the baselines, and experimental results demonstrate that our method achieves significant performance for the relation extraction task of online cultural relic data.

The main contributions of our work are summarized as follows:

1. We propose a relation extraction model named WAtt-Capsnet, which uses capsule networks with word-attention dynamic routing to extract instantiation features for the relation extraction task of the online cultural relic data.
2. We also present combination embedding to capture the characteristic information of Chinese sentences by considering the contribution of words and parts of speech and characters as well as the position of words, which captures rich internal structure information of the sentence and highlights the impact of entities.
3. We also import word-attention in dynamic routing by considering the different contributions of the words, which iteratively amend the connection strength to solve the problem of long-term memory by increasing the weight coefficient of the informative words.
The remaining sections of this paper are described as follows. Prior relevant works about traditional and deep learning models for relation extraction and word representation are organized in Section 2. In Section 3, we represent our framework for cultural relic relation extraction and introduce the working principle of each component of our model in detail. In Section 4, we introduce the dataset, evaluation, baselines, and describe the experimental process and experimental results. In Section 5, we objectively discuss the performance of our model in relation extraction tasks through experimental results. Section 6 mainly describes our major research findings and future research priorities.

II. RELATED WORK

A. TRADITIONAL METHODS FOR RELATION EXTRACTION

Relational extraction has important significance for the research of information retrieval, text understanding and knowledge graph construction [15]. Traditional methods for relation extraction are feature engineering-based and kernel function-based.

Early methods performed on relation extraction tasks used handcrafted features. A set of designed features and a classifier are trained to classify other new relation instances. Kambhatla [16] employed maximum entropy models that combined diverse lexical, syntactic and semantic features exploited from the text to solve the problem of the scarcity of labelled data and the errors produced by entity detection modules in the task of extracting semantic relations between entities. Based on Wang’s work, Zhou et al. [17] incorporated additional features and systematically explored the method of feature-based relational extraction. They combined the base phrase chunking information at the syntactic aspect and used the semantic information (e.g., WordNet and name list) to improve the performance.

The overall performance of the approach based on features largely depends on the validity of the features generated. The methods applying high-level lexical and syntactic features suffer from error propagation [18]. The disadvantage of explicit feature engineering is avoided by kernel-based methods that design kernel functions to compute the similarities of two relation instances [19]. Mooney and Bunescu [20] proposed a kernel method that extracted semantic relations between entities. The method was based on a generalization of three types of subsequence kernels to maintain entity relations. The authors proposed that each word instead of each sequence was generalized to a feature vector (i.e., one feature vector for each word), then a sequence of feature vectors was represented for each relation instance. Khayyamian et al. [21] proposed a generalized convolution tree kernel and using syntactic parse tree kernels was proposed by Zhang et al. [22]. The convolutional tree kernel produced various subkernels from different definitions of the two weighting functions for relation extraction.

B. DEEP LEARNING MODELS FOR RELATION EXTRACTION

In kernel-based methods, elaborately designed kernels utilized pre-existing systems, which caused errors of the individual modules gathering downstream. Fortunately, this problem has been fortunately removed by popular deep learning [4]. The early deep learning model treated the task of relation extraction as a multi-class classification problem that added a relation class label to each sentence. Santos, et al. [23] proposed a ranking CNN for relation classification tasks. The model generates a distributed vector representation for the text using a convolutional layer and creates a score for each class comparing the distributed vector representation with the class representations. CNN cannot well represent the precise spatial relations between the high-level parts. Recurrent neural networks are another popular deep learning model for relation extraction. Sorokin et al. [6] utilized the LSTM-based encoder to jointly capture the representations for all relations in the sentential context. The final prediction was produced by combining representations of the context relations and the representation of the target relation.

To better capture important features, Lin, et al. [24] proposed an attention-based CNN model, and Zhang, et al. [25] proposed an attention-based LSTM model for RE. However, it is difficult for CNN and LSTM to obtain richer instantiation features.

C. CAPSULE NETWORKS FOR RELATION EXTRACTION

Recently, research has proposed the capsule network to improve the representation imperfection of CNNs and RNNs. Wang, et al. [26] proposed a framework based on the capsule network for relation extraction. They used a capsule network in the sentence encoder, and the involvement of the capsule layer made encoding spatial patterns more powerful, which helped to determine the relation expressed in sentences. Zhang, et al. [27] presented a framework based on capsule networks for relation extraction. In their method, the multi-label relation extraction problem is converted into a multiple binary classification problem by the capsule networks. They treated aggregation as a routing problem so that complex and interleaved features could be identified by the capsule network. Zhang et al. [28] proposed capsule networks with dynamic routing for multi-labelled relation extraction to address the challenge of having multiple relations in a sentence. To better cluster features, they devised a dynamic routing algorithm based on an attention mechanism. They further designed a sliding-margin loss function to extract the relations precisely.

Because of the effectiveness of the capsule network with attention-based dynamic routing in relation extraction, we use the capsule network with word-attention dynamic routing to extract instantiation features for the task of extracting cultural relic text relations.

D. WORD REPRESENTATION

Liu [5] used a simple convolutional neural network model to automatically capture features instead of handcrafted features. This model used word vectors and lexical features to encode the input sentence and then utilized a convolutional layer and a softmax layer given a probability
distribution of the relation classes. The model assigned a vector to each synonym class. However, it did not take full advantage of the real representational ability of word embedding. Zhou, et al. [29] converted a word into word embedding by utilizing the matrix-vector product (i.e., every word is transformed into a real-valued vector). They looked up the embedding matrix for each word when inputting a sentence consisting of some words and used 50-dimensional and 100-dimensional pre-trained word vectors to represent words as the input of their long short-term memory (LSTM) layer. Word embedding effectively captures semantic and syntactic features of words but underuses features contained in a character. Zhang, et al. [30] used character embedding initialized with a vector pre-trained by word2vec and character position embedding as the vector representation. In a sentence containing two entities, each character is mapped to a low-dimensional dense vector, which combines the character vector and the position vector of the word in the sentence. The position embedding combines the previous entity and the latter entity to the relative positions of the current character, each of which corresponds to a position embedding.

To make better use of character characteristics and effectively capture semantic and syntactic information of words, we propose a combination of word embedding and character embedding to improve the performance of relation extraction.

III. METHODOLOGY
In this section, we describe the architecture of WAtt-Capsnet based on a capsule network with word-attention dynamic routing for cultural relic relation extraction. Our proposed network is a variant of the original capsule network. As depicted in Figure 1, our relation extractor comprises four primary layers: the input representation layer, bidirectional LSTM layer, capsule network with a word-attention dynamic routing layer, and relation prediction layer.

1) Input Representation Layer. The input representations of our model incorporate four components to obtain the critical context information for each word.

2) Bidirectional LSTM Layer. The forward and backward networks of the BiLSTM model are used to capture the low-level features of the sentences.

3) Capsule Network with a Word-Attention Dynamic Routing Layer. Capsule networks with word-attention dynamic routing are applied to extract instantiation features. This layer contains the capsule network and word-attention dynamic routing algorithm.

4) Relation Prediction Layer. A separate-margin loss function similar to Zhang [28] is used to predict possible relations.

A. INPUT REPRESENTATION LAYER
In the input representation layer, we incorporate four components as the input representations of our model for each sentence: word embedding, character embedding, part-of-speech (POS) vectors and position vectors because they contain critical context information for each sentence.

1) WORD EMBEDDING
We adopt word2vec-based embedding [31], an open source tool developed by Google, which learns low-dimensional continuous vector representations of words. The word embedding pre-trained in the skip-gram setting of word2vec are distributed representations of words that map each word \( w_i \) of the sentence \( s_i \) to the \( d_w \)-dimensional real-valued vector.

2) CHARACTER EMBEDDING
We also use word2vec to train character embedding to address the unsatisfactory performance of word segmentation for relation extraction tasks [32]. Character embedding are representations mapping each character \( c_j \) of the word \( w_i \) to the \( d_c \)-dimensional vector.

3) POS EMBEDDINGS
To represent the POS features of the words in \( s_i \), POS tags are predefined according to the scheme of Jieba, and then we map each POS tag of \( w_i \) to a \( d_{POS} \)-dimensional vector.

4) POSITION EMBEDDINGS
To present the position features, we import the relative distances between the current word and \( M \) entities. When \( s_i \) contains only a single entity pair relation, \( M = 2 \); when \( s_i \) contains multiple entity pairs relation, we limit the maximum
number of entity pairs in a sentence to two, \( M = 4 \). When the distance between the word and entity1 is 2 and the distance between the word and entity2 is -4, this word has two positional characteristics \([d_1, d_2]\), which are represented as a \( d_p \)-dimensional vector. Position embedding highlight the impact of entities.

Finally, we concatenate the word embedding, character embedding, POS vectors and position vectors and then obtain a combinatorial representation for each word as input for our model.

B. BIDIRECTIONAL LSTM LAYER

Bidirectional LSTM is effective for modelling sequence data in relation extraction tasks [33]. Consequently, we leveraged BiLSTM to deeply exploit the low-level features of the sentence. The combination of embedding from the input representation layer is input into the BiLSTM model. Then, the global sequence features are captured by both the forward and backward networks of the BiLSTM model. Finally, the hidden state vector \( h_t \) is expressed as \( h_t = [\tilde{h}_t \oplus \bar{h}_t] \) as the output, where \( \oplus \) is the element-wise sum.

C. CAPSULE NETWORK WITH WORD-ATTENTION DYNAMIC ROUTING

1) CAPSULE NETWORK

The output of the capsule is a vector that can more richly express the features extracted by nodes, and different dimensions in the vector are used to record different attributes of the same feature. In our work, we used capsule networks to extract instantiation features. A capsule is a group of neurons whose activity vector represents the instantiation parameters of a specific type of relation feature. We use the length of the output vector of a capsule to represent the probability that the relation features exist, and the orientation of the vector expresses the specific property of one kind of feature. We disperse all the semantic information captured by the BiLSTM into the first level capsules. The representation of each word token is expressed by capsules. Then, a non-linear “squash” function shown as follows is used to compress the vector length,

\[
v_j = \text{squash}(s_j) = \frac{\|s_j\|^2}{1 + \|s_j\|^2} s_j,
\]

where \( v_j \) is the output of capsule \( j \) whose input is \( s_j \). Except for the first layer of capsules, the total input \( s_j \) to a capsule \( j \) is a linear weighted sum over all the prediction vectors \( \hat{u}_{ij} \) from the capsules in the low-level layer and is calculated by multiplying the instantiated parameters of a capsule \( u_i \) in the layer below by a weight matrix \( W_{ij} \) as follows:

\[
s_j = \sum_i c_{ij} \hat{u}_{ij}, \quad \hat{u}_{ij} = W_{ij} u_i,
\]

where \( c_{ij} \) is the coupling coefficient that is iteratively obtained by dynamic routing during the testing stage, the initial logit \( b_{ij} \) coupled through capsule \( i \) and capsule \( j \) is the log prior probability, which determines the coupling coefficients \( c_{ij} \) through a routing softmax. The coupling coefficient \( c_{ij} \) can be calculated as follows:

\[
c_{ij} = \frac{\exp(b_{ij})}{\sum_j \exp(b_{ij})}
\]

2) WORD-ATTENTION DYNAMIC ROUTING ALGORITHM

With the capsule network, we can obtain high-level capsules that represent relation features. The high-level capsule network can more richly express relation features. However, the long sequence of source sentences in the data of cultural relic results in a large attenuation in the gradient update and failures to effectively update the parameters at the head of the sequence.

Therefore, we propose a word-attention dynamic routing algorithm that iteratively amends the connection strength. Our attention mechanism allows the decoder to obtain important input information to solve the problem of long-term memory, which is an extension of the standard attention mechanism. The goal of word-attention is to ignore irrelevant words and focus on informative words, and the weight coefficient of informative words is greater than that of irrelevant words. The word-attention mechanism distributes different weights to different words; that is, informative words have high weights, while irrelevant words have low weights to increase the intensity of informative words.

Let \( P \) be a matrix of the output vectors \([p_1, p_2, \ldots, p_L]\) generated by the BiLSTM layer, where \( L \) denotes the sentence length. In detail, the weight \( \beta_{ij} \) of relational word \( p_{ij} \) as described by Zhou et al. [29] is computed by the following equation.

\[
\beta_{ij} = \frac{\exp(p_{ij} w_{ij} q_{ij})}{\sum_{k=1}^{L} \exp(p_{ik} w_{ik} q_{ik})},
\]

\[
p_{ij} = (p_i)^T p_j,
\]

where \( w_{ij} \) is a weighted matrix, and vector \( q_{ij} \) is an informative word.

The pseudocode for the word-attention routing algorithm is summarized in Algorithm 1.

\begin{algorithm}
\caption{Word-Attention Dynamic Routing Algorithm}
\begin{algorithmic}[1]
\STATE 1: Input to Routing Procedure: \((\hat{u}_{ij1}, \hat{a}_{ji1}, r)\), iterative number \( r \)
\STATE 2: initialize the logits of coupling coefficients \( b_{jii} = 0 \).
\STATE 3: For \( r = 0 \) iterations do:
\STATE 4: \( c_{ji} = \hat{a}_{ji} \text{softmax}(b_{jii}) \)
\STATE 5: \( s_j = \sum_i c_{ij} \hat{u}_{ij}, \quad \hat{u}_{ij} = W_{ij} u_i \)
\STATE 6: \( \beta_{ij} = \text{softmax}(p_{ij} q_{ij}) \)
\STATE 7: \( v_j = \text{squash}(s_j \cdot c_{ij} \beta_{ij}), \quad a_j = |v_j| \)
\STATE 8: \( b_{jii} = b_{jii} + v_j \cdot \hat{u}_{jj} \)
\STATE 9: Return \( v_j, a_j \)
\STATE 10: End procedure
\end{algorithmic}
\end{algorithm}
We compare the performance of our model on the evaluation dataset. It can be seen that our model WAtt-Capsnet improves more than the baselines in precision, recall and F1-score. Overall, our model WAtt-Capsnet exhibits the best performances in precision, recall and F1-score.

2) EFFECT OF WORD REPRESENTATION ON OUR METHOD

To evaluate the effect of word representation on our model, we compared our method to the model-applied capsule network with word-attention dynamic routing, which uses four kinds of word representations (word embedding, character embedding, the combination of word and character embedding, the combination of word, character embedding and POS) as input. Based on the combination of word, character embedding and POS, we consider the importance of position and add position vector in the word representation to ensure that our model performs effectively in capturing features of the entities. The experimental results are shown in Figure 2. Our model WAtt-Capsnet achieves maximum precision, recall and F1-scores in relations DUR, LUR, DMR, MCR, and the mean precision, recall and F1-scores of the four relations.

In Figure 2, DUR is the relation of cultural relic/date_unearth, LUR is the relation of cultural relic/location_unearth, DMR is the relation of cultural relic/dynasty_manufactured, MCR is the relation of cultural relic/museum_collection, and All is the mean of the DUR, LUR, DMR and MCR. Word is utilizing the word embeddings as the input representations of our model, character is utilizing the character embeddings as the input representations of our model, W_C is utilizing the combination of words and character embeddings as the input representations of our model, W_C_POS is utilizing the combination of word embeddings, character embeddings and POS vectors as the input representations of our model, Watt-CapsNet is our method that utilizes the four components as the input representations: word embeddings, character embeddings, part-of-speech (POS) vectors and position vectors.

3) EFFECT OF WORD-ATTENTION DYNAMIC ROUTING ON OUR METHOD

To better validate the performance of word-attention dynamic routing, we compared our model with the variant model NWatt-CapNet, which removes the word-attention from dynamic routing. The experimental results in Figure 3 show that our model WAtt-Capsnet using a capsule network with word-attention dynamic routing outperforms the variant model NWatt-CapNet, although the variant model still proves effective when used for other tasks. Word-attention is useful for capsule networks and greatly improves the performance of cultural relic relation extraction.

V. DISCUSSION

The overall results of the performance comparison shown in Table 1 have higher precision, recall and F-measure over
the baseline models, which demonstrates that the capsule network with word-attention dynamic routing performs better than the baseline model, which implies the substantial appropriateness and the advantage of our proposed model based on the capsule network capturing richer instantiation features over existing approaches in the task of cultural relic relation extraction.

The further word representation constituted by word embedding, character embedding, POS vector and position vector evaluation reveals that using pre-trained embedding combined with POS and word position can dramatically improve the performance of relation extraction. Most likely, the same words appear in different positions in a sentence, which has different POSs and positions, so they are assigned to different representation vectors as the embedding of the model input. The POS of the word can further effectively express the feature of the word, and the position of the word highlights the impact of entities. This could be one reason why our model generally improves the overall relation extraction performance and outperforms the other approaches.

The evaluation of word-attention dynamic routing on our method also suggests that word-attention dynamic routing is useful for capsule networks and greatly improves the performance of cultural relic relation extraction. This may be because word-attention dynamic routing focuses on informative words and the weight coefficient of informative words
is greater than that of irrelevant words to increase the intensity of informative words, which enhances the identification information of different types of relations. This could be another reason why our model generally improves the overall relation extraction performance.

VI. CONCLUSIONS

In this paper, we propose a model named WAtt-Capsnet based on capsule networks with word-attention dynamic routing for the relation extraction task of online cultural relic data. The capsule network captures richer instantiation features and enhances relation discrimination features with word-attention dynamic routing to reduce the decay of useful information in long sentences so that informative word features can be combined effectively in relation extraction. We also present the combination embedding of word embedding, character embedding, POS and the position of words as the model input to capture the features and rich internal structure information of sentences. Comparative experiment results demonstrate that our method achieves significant performance for the relation extraction task of online cultural relic data.

At present, most deep learning methods for relation extraction are supervised methods and require a large quantity of labelled data. In the future, we will study the semi-supervised and unsupervised learning methods of relation extraction to further improve the efficiency of relation recognition and reduce dependency on labelled data simultaneously.

REFERENCES

[1] W. Li, P. Zhang, F. Wei, Y. Hou, and Q. Lu, “A novel feature-based approach to chinese entity relation extraction,” in Proc. 46th Annu. Meet. Assoc. Comput. Linguistics Hum. Lang. Technol. Short Papers (HLT), 2008, pp. 89–92.

[2] S. Zhao and R. Grishman, “Extracting relations with integrated information using kernel methods,” in Proc. 43rd Annual Meeting Assoc. Comput. Linguistics, Jun. 2005, pp. 419–426.

[3] D. Zeng, K. Liu, S. Lai, G. Zhou, and J. Zhao, “Relation classification via convolutional deep neural network,” in Proc. COLING, 2014, pp. 2335–2344.

[4] S. Kumar, “A survey of deep learning methods for relation extraction,” 2017, arXiv:1705.03645. [Online]. Available: http://arxiv.org/abs/1705.03645

[5] C. Liu, W. Sun, W. Chao, and W. Che, “Convolution neural network for relation extraction,” in Proc. Int. Conf. ADMA, Berlin, Germany: Springer, 2013, pp. 231–242.

[6] D. Sorokin and I. Gurevych, “Context-aware representations for knowledge base relation extraction,” in Proc. Conf. Empirical Methods Natural Lang. Process., 2017, pp. 1784–1789.

[7] B. Yu, Z. Zhang, T. Liu, B. Wang, S. Li, and Q. Li, “Beyond word attention: Using segment attention in neural relation extraction,” in Proc. 28th Int. Joint Conf. Artif. Intell., Aug. 2019, pp. 5401–5407.

[8] G. E. Hinton, A. Krizhevsky, and S. D. Wang, “Transforming auto-encoders,” in Proc. Int. Conf. Artif. Neural Netw. Berlin, Germany: Springer, 2011, pp. 44–51.

[9] S. Sabour, N. Frosst, and G. E. Hinton, “Dynamic routing between capsules,” in Proc. NIPS, Long Beach, CA, USA, 2017, pp. 3856–3866.

[10] X. Ma and E. Hovy, “End-to-end sequence labeling via bi-directional LSTM-CNNs-CRF,” 2016, arXiv:1605.01354. [Online]. Available: http://arxiv.org/abs/1605.01354

[11] G. Lample, M. Ballesteros, S. Subramanian, K. Kawakami, and C. Dyer, “Neural architectures for named entity recognition,” in Proc. Conf. North Amer. Chapter Assoc. Comput. Linguistics, Hum. Lang. Technol., San Diego, CA, USA, 2016, pp. 260–270.

[12] R. Hoffmann, C. Zhang, X. Ling, L. S. Zettlemoyer, and D. S. Weld, “Knowledge-based weak supervision for information extraction of overlapping relations,” in Proc. ACL, 2011, pp. 541–550.

[13] Y. Shen and X. J. Huang, “Attention-based convolutional neural network for semantic relation extraction,” in Proc. 26th Int. Conf. Comput. Linguistics, Tech. Papers (COLING), 2016, pp. 2526–2536.

[14] G. Ji, K. Liu, S. He, and J. Zhao, “Distant supervision for relation extraction with sentence-level attention and entity descriptions,” in Proc. AAAI, 2017, pp. 3060–3066.

[15] L. Duan, Q. Xu, and A. Li, “Research on effect of entities semantic information on Chinese entity relation extraction,” Appl. Res. Comput., vol. 34, no. 1, pp. 141–146, 2017.

[16] M. Kambhala, “Combining lexical, syntactic, and semantic features with maximum entropy models for extracting relations,” in Proc. ACL Interact. Poster Demonstration Session, 2004, pp. 22–27.

[17] Z. GuoDong, S. Jian, Z. Jie, and Z. Min, “Exploring various knowledge in relation extraction,” in Proc. 43rd Annu. Meeting ACL. Ann Arbor, MI, USA: Univ. of Michigan, 2005, pp. 427–434.

[18] M. Miwa and M. Bansal, “End-to-end relation extraction using LSTMs on sequences and tree structures,” 2016, arXiv:1601.00770. [Online]. Available: http://arxiv.org/abs/1601.00770

[19] S. Pawar, G. K. Palanikar, and P. Bhattacharyya, “Relation extraction: A survey,” 2017, arXiv:1712.05191. [Online]. Available: http://arxiv.org/abs/1712.05191

[20] R. J. Mooney and R. C. Bunescu, “Subsequence kernels for relation extraction,” in Proc. Adv. Neural Inf. Process. Syst., 2006, pp. 171–178.

[21] M. Khayamian, S. A. Mirroshandel, and H. Abolhasani, “Syntactic tree-based relation extraction using a generalization of collins and dutty convolution tree kernel,” in Proc. Hum. Lang. Technol., Annu. Conf. North Amer. Chapter Assoc. Comput. Linguistics. Companion Volume, Student Res. Workshop Doctoral Consortium (NAACL), 2009, pp. 66–71.

[22] M. Zhang, G. Zhou, and A. Aw, “Exploring syntactic structured features over parse trees for relation extraction using kernel methods,” Inf. Process. Manage., vol. 44, no. 2, pp. 687–701, Mar. 2008.

[23] C. Nogueira dos Santos, B. Xiang, and B. Zhou, “Classifying relations by ranking with convolutional neural networks,” 2015, arXiv:1504.06580. [Online]. Available: http://arxiv.org/abs/1504.06580

[24] Y. Lin, S. Shen, Z. Liu, H. Luan, and M. Sun, “Neural relation extraction with selective attention over instances,” in Proc. 54th Annu. Meeting Assoc. Comput. Linguistics, vol. 1, 2016, pp. 2124–2133.

[25] R. Zhang, F. Meng, Y. Zhou, and B. Liu, “Relation classification via recurrent neural network with attention and tensor layers,” Big Data Mining Anal., vol. 1, no. 3, pp. 234–244, 2018.

[26] Z. Wang, Y. Zhang, and C. Xing, “Reducing wrong labels for distant supervision relation extraction with selective capsule network,” in Proc. APWeb WAIM Joint Int. Conf. Web Big Data Cham, Switzerland: Springer, 2019, pp. 77–92.

[27] N. Zhang, S. Deng, Z. Sun, X. Chen, W. Zhang, and H. Chen, “Attention-based capsule networks with dynamic routing for relation extraction,” in Proc. Conf. Empirical Methods Natural Lang. Process., 2018, pp. 1–8.

[28] X. Zhang, P. Li, W. Jia, and H. Zhao, “Multi-labeled relation extraction with attentive capsule network,” in Proc. AAAI Conf. Artif. Intell., Jul. 2019, pp. 7484–7491.

[29] P. Zhou, W. Shi, J. Tian, Z. Qi, B. Li, H. Hao, and B. Xu, “Attention-based bidirectional long short-term memory networks for relation classification,” in Proc. 54th Annu. Meeting Assoc. Comput. Linguistics, Berlin, Germany, vol. 2, 2016, pp. 207–212.

[30] Z. Zhang, T. Zhou, Y. Zhang, and Y. Pang, “Attention-based deep residual learning network for entity relation extraction in chinese EMR,” BMC Med. Inform. Decis. Making, vol. 19, no. S2, pp. 55–65, Apr. 2019.

[31] T. Mikolov, K. Chen, G. Corrado, and J. Dean, “Efficient estimation of word representations in vector space,” 2013, arXiv:1301.3781. [Online]. Available: http://arxiv.org/abs/1301.3781

[32] T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado, and J. Dean, “Distributed representations of words and phrases and their compositionality, in Proc. Adv. Neural Inf. Process. Syst., Lake Tahoe, NV, USA, 2013, pp. 3111–3119.

[33] L. Song, Y. Zhang, Z. Wang, and D. Giklda, “N-ary relation extraction using graph state LSTM,” 2018, arXiv:1808.09101. [Online]. Available: http://arxiv.org/abs/1808.09101
MIN ZHANG received the B.S. degree from Xi’an University, Xi’an, China, in 2004, and the M.S. degree in computer software and theory from Northwest University, Xi’an, in 2010. She is currently pursuing the Ph.D. degree in software engineering at Northwest University. She is also an Associate Professor. Her research interests include information processing, machine learning, and knowledge graph. She is a member of China Computer Federation.

GUOHUA GENG is currently a Professor at Northwest University, Xi’an, China. She is also a Ph.D. Supervisor with the School of Information Science and Technology, Northwest University, China. Her major research interests include information processing and image processing. She is the Director of the National-Local Joint Engineering Research Center of Cultural Heritage Digitization and a reviewer in many international conferences and journals. She also received several academic rewards, including three National Science and Technology Progress Awards (second class).