Proceedings of Machine Learning Research vol 245:1–9, 2024 2024 International Conference on Machine Learning and Intelligent Computing

Hybrid Neural Network Model for Extracting Character Relationships that Integrates Multi-level Information

Xichao Zhu* 2021107252@ZUT.EDU.CN
Zhongyuan University of Technology, Zhengzhou, 451100, China

Zhanbin Chen
Zhongyuan University of Technology, Zhengzhou, 451100, China

Editors: Nianyin Zeng and Ram Bilas Pachori

Abstract

In the realm of natural language processing, extracting entity relationships in Chinese holds paramount importance, serving as the bedrock for various downstream tasks. However, the complexity entrenched within Chinese linguistic structures and semantic nuances presents formidable challenges. To surmount these obstacles, our study proposes a sophisticated neural network model meticulously designed to integrate multi-level information for relationship extraction. By segmenting sentences into distinct components, entity-adjacent words and the sentence. The model capitalizes on the strengths of BERT pre-trained models for dynamic word embedding. Furthermore, leveraging Bidirectional Gated Recurrent Units (BiGRU) and Convolutional Neural Networks (CNN) at the sentence level enables the capture of both sequential and structural features. At the entity-adjacent word level, a fusion of two fully connected neural networks extracts intricate associations between entities and neighboring words. Rigorous ablation and comparative experiments conducted on a comprehensive corpus underscore the efficacy of our proposed approach. Remarkably, compared to benchmark methods, our model exhibits a substantial 6.61% increase in recall rate and a noteworthy 5.31% improvement in the F1 value.

Keywords: Character relationship extraction, Multilevel information, Hybrid neural network, Pretrained model

1. Introduction

The rise of the digital era and internet technology has resulted in a vast amount of textual data needing processing. Extracting useful information quickly from this data is crucial. Relation Extraction (RE) is key in this, aiming to extract relationships between entities from unstructured text, represented as relation triplets <e1, r, e2>. These triplets can be applied in tasks like knowledge base construction and question-answering systems, boosting information discovery efficiency.

Relation extraction, a core task in natural language processing, has diverse applications across domains. It aims to identify relationships between entities in unstructured text. Given a sentence with two annotated entities, the task is to determine the appropriate relationship from a predefined set.

In recent years, the popularity of deep learning has risen significantly, with many researchers applying deep learning to the field of relation extraction and achieving notable results. The literature (Lin et al., 2016) uses convolutional neural networks to automatically extract lexical and sentence-level features for relation extraction, effectively mitigating the shortcomings of traditional features. The literature (Wang et al., 2016) introduce attention mechanisms to CNN, dynamically reducing the weight of noisy instances. However, CNN are limited in extracting features only from local information in the text, struggling to capture long-distance dependencies. The literature (Zhang and Wang, 2015) introduces recurrent neural networks (RNN) for relation extrection, but RNNs suffer from the issues of vanishing and exploding gradients, making it challenging to handle long sequences and maintain long-term memory during training. The literature (Hochreiter and Schmidhuber, 1997) designs Long Short-Term Memory (LSTM), effectively addressing the longterm dependency problem in traditional RNN models. The literature (Zhang et al., 2015) by using bidirectional LSTM, extracts context information more effectively from long texts, improving the performance of relation extraction. The literature (Zhou et al., 2016) combines attention mechanisms with bidirectional LSTM, merging wordlevel features at each time step of LSTM by multiplying with weight vectors for sentence-level feature extraction in relation classification. The literature (Dey and Salem, 2017) introduces Gated Recurrent Unit (GRU), which has only two gates—reset and update gates. The reset gate controls the retention of historical information, while the update gate controls the selective update of current information, balancing information retention and forgetting. GRU has fewer parameters, making it more efficient computationally.

In Chinese entity relation extraction research, due to structural differences between Chinese and English, the literature (Zhang and Hu, 2018) addresses Chinese structural characteristics and designs a bidirectional GRU-based dual attention mechanism for Chinese relation extraction, aiming to alleviate noise issues in distant supervision data. The literature (Li et al., 2019) takes into account the vocabulary, syntax, semantics, and positional features of Chinese text, using self-attention bidirectional LSTMs for relation classification.

In 2018, Google released the BERT pre-trained language model based on the Transformer (Devlin et al., 2019), capable of generating high-quality word vectors. The literature (Du et al., 2021) constructs dynamic word vectors using BERT, extracting context features with bidirectional GRU and CNN, achieving good results. The literature (YAO et al., 2021) considers text, entity, and entity context information, embedding word vectors using BERT, and extracting information at different levels with bidirectional LSTMs and fully connected networks for integrated relation class prediction. The literature (Zhao et al., 2022) addresses the issue of conventional relation extraction ignoring other possible relationships in a sentence, constructing an end-to-end relation extraction model that integrates BERT and label-dependent knowledge. This model uses graph neural networks to learn dependencies between labels, improving the performance of relation extraction tasks. The literature (ZHAO et al., 2023) designs a model based on BERT and a hybrid neural network for medical entity relation extraction. It uses bidirectional LSTMs to capture long-contextual dependencies and CNNs to capture local text features, then combines the two features for relation classification.

In summary, Deep learning methods like CNN, RNN and BERT, along with methods combining pre-trained models, show promise in extracting text features and handling long-distance dependencies. However, certain challenges still require further research.

Chinese and English corpora exhibit significant differences in person relation extraction. Chinese, with its complex linguistic structure and limited morphological variations, poses challenges such as synonyms and polysemous words, requiring careful consideration of contextual clues for accurate extraction. Traditional single neural network models face difficulties with Chinese, highlighting the need for more intricate models. Traditional word vector methods yield static representations, overlooking semantic relationships within contexts. In contrast, BERT generates comprehensive vectors, capturing contextual semantics and enhancing performance in relation extraction tasks by better capturing semantic information.

In addressing the aforementioned issues, this paper makes the following main contributions:

(1) Proposes a model for Chinese person relation extraction that integrates original textual information and adjacent word-level information in Chinese corpora, organically combining with the Bert-BiGRU-CNN hybrid neural network model.

(2) Constructs a Chinese person relation extraction corpus by combining the CNDBpedia knowledge base and encyclopedia website, conducting experiments on the constructed corpus. The proposed model outperforms relatively superior methods, achieving improvements of 6.61% in recall and 5.31% in F1 score metrics.

2. Hybrid Neural Network Model

Aiming to address the intricate grammatical structure of the Chinese corpus and the challenge of extracting comprehensive text features using a single neural network, we propose the adoption of a hybrid neural network model, MLI-HNN (MultiLevel Infused Hybrid Neural Networks), which integrates multi-level information. The model framework diagram is presented in Figure 1 and model architecture is presented in Figure 2. The model primarily comprises three layers: the embedding layer, the feature extraction layer, and the fusion layer.

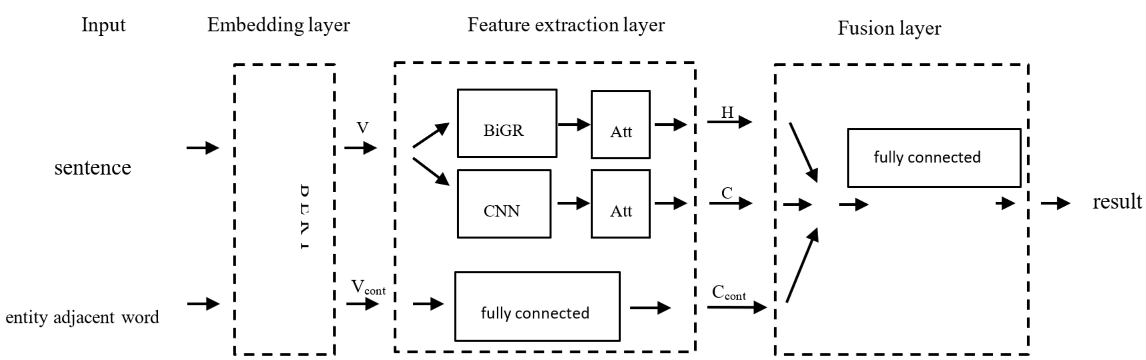

Figure 1: Model framework diagram.

2.1. Embedding layer

In MLI-HNN, the pre-training model BERT serves as an embedding layer to map each word in the text to a lowdimensional vector.

When employing BERT for constructing word vectors in a text sequence $X = x_1, x_2, x_3, ..., x_n$. it is essential to add "[CLS]" at the beginning and "[SEP]" at the end. The "[CLS]" signifies the beginning, and the "[SEP]" indicates the end of the text. This processed sequence is then input into the BERT model to obtain the word vector V.

2.2. Multi-level information feature extraction layer

In the feature extraction layer, MLI-HNN employs a Hybrid Neural Network to extract informative features from the text layer. The Hybrid Neural Network includes bidirectional GRU for capturing textual context information and CNN for generating local feature vectors. Additionally, two fully connected networks are used to extract features at the entity adjacent word level.

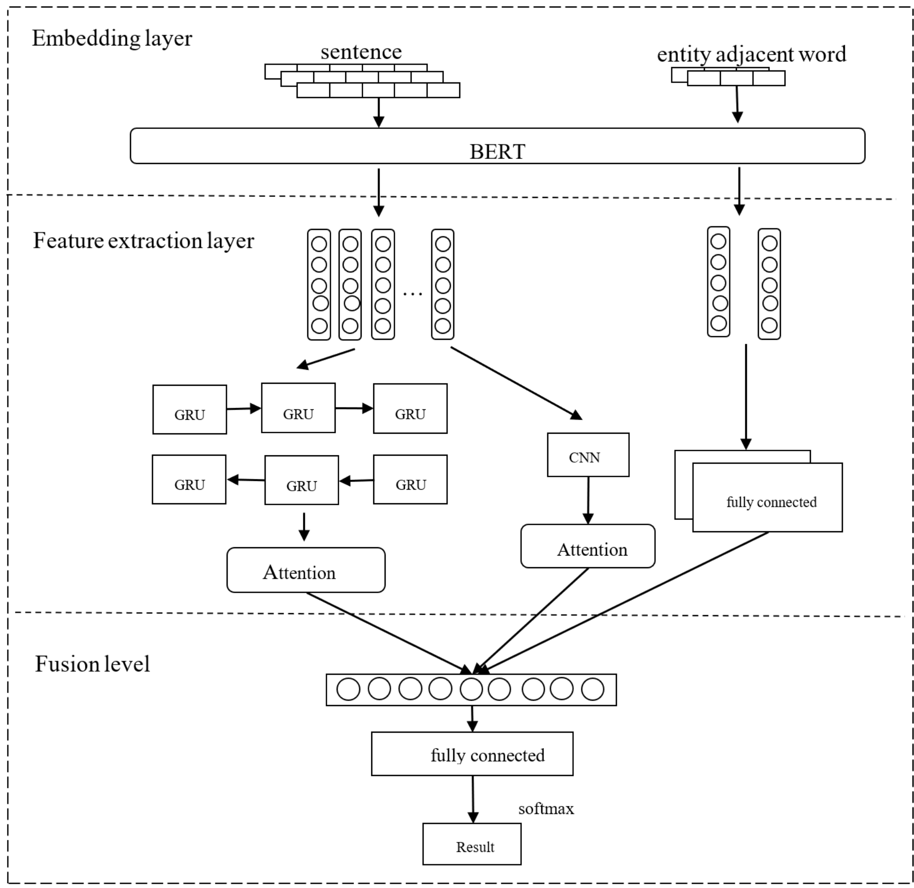

Figure 2: Model architecture.

At the sentence level, the word vectors $V = E_1, E_2, E_3, ..., E_n$ are fed into BiGRU and CNN for feature extraction. MLI-HNN utilizes bidirectional GRU to establish textual context information. This architecture can concurrently incorporate both forward and backward information of the input sequence and are computed as shown in Equations 1, 2, 3, enhancing the model's ability to capture contextual correlations within the sequence. Subsequently, an attention layer is applied to optimize ht, resulting in the feature representation H after attention weighting.

$$\overrightarrow{h}_t = GPU(x_t, \overrightarrow{h}_{t-1}) \tag{1}$$

$$\overleftarrow{h}_t = GPU(x_t, \overleftarrow{h}_{t-1}) \tag{2}$$

$$h_t = W_t \overrightarrow{h}_t + V_t \overrightarrow{h}_t + b_t \tag{3}$$

Considering that the features of different entities, different characters in the text are very important for the relationship prediction of the model, MLI-HNN uses CNN to extract the local features of each entity of the text. The convolution of convolutional feature p is calculated as shown in Equation 4.

$$p_c = f(W_{c \cdot e2:i+l-1} + b_c) \tag{4}$$

Pooling operation is required for the obtained convolutional feature p. The maximum pooling is selected to pool p as shown in Equation 5.

$$c = \max(p_c) \tag{5}$$

The pooled feature vector c is obtained, and the attention feature weighting is performed on c to obtain the weighted feature representation C.

At Entity Proximity Word Hierarchy, the proximity words of two entities are fed into two separate fully connected networks. These two networks perform feature extraction for each entity's proximity word segments separately, as a way to capture the semantic associations between entities and their neighboring words. The proximity word segments of two entities are denoted by V_{cont1}, V_{cont2} respectively, and the relationship features between entities and entity proximity words are extracted, which are denoted by C_{cont1}, C_{cont2} respectively, and are computed as shown in Equations 6 and 7.

$$C_{cont1} = f(W \cdot V_{cont1} + d) \tag{6}$$

$$C_{cont2} = f(W \cdot V_{cont2} + d) \tag{7}$$

2.3. Fusion level

At the multi-level information extraction stage, input text context information H is obtained through BiGRU. The local information of the text is constructed using CNN, denoted as C.

Information about entity adjacent words, represented as C_{cont1}, C_{cont2} is acquired through fully connected networks. The concatenation fusion method is applied to combine different features in sequence, as shown in Equation 8.

$$O = [H \oplus C \oplus C_{cont1} \oplus C_{cont2}] \tag{8}$$

Subsequently, the feature vector O is fed into a fully-connected layer that linearly transforms the feature vector through the weight matrix W as well as the bias b to provide inputs to the Softmax function, which ultimately converts the result of the linear transformation into a probability distribution for predicting the class of relationships between pairs of entities, as shown in Equation 9.

$$p\,(y) = soft\max(z) = \frac{\exp(z_i)}{\sum_{j=1}^{N_r} \exp(z_j)} \tag{9}$$

During the training process, the parameters of the model are optimized by minimizing the loss function, which is usually defined as cross-entropy loss, as shown in Equation 10.

$$J(\theta) = -\sum\nolimits_{i=1}^{N_s} \log(y_i|x_i,) \tag{10}$$

In order to efficiently find the optimal solution of the parameters, the MLI-HNN model adopts the Adam optimizer, which dynamically adjusts the learning rate in each iteration to accelerate the convergence and improve the model performance. Through multiple iterations of training, the Adam optimizer helps the model achieve parameter optimization in the Chinese character relationship extraction task.

3. Experimental Study

3.1. Dataset

Currently, due to the lack of high-quality datasets for Chinese character relationship extraction, this poses a challenge for research. In order to solve this problem. In this paper, we extracted Chinese character relationship triad information by integrating online knowledge base resources such as CN DBpedia. Web crawler technology is further utilized to retrieve entity pair co-occurring sentences related to the triad from authoritative encyclopedia websites such as Baidu Encyclopedia and Wikipedia to ensure the breadth and representativeness of the data. In the data preprocessing stage, data cleaning and screening were performed to eliminate noise and irrelevant information in order to improve data quality. Manual review was also performed to ensure the accuracy of entity relationship labeling in the dataset. In the process of constructing the dataset, data cleaning and labeling consistency testing measures were used to improve the reliability and usefulness of the dataset.

The final character relationship dataset contains 22,204 instances, and during the training process, the whole dataset is randomly divided into three copies most training set, testing set and validation set respectively, with the ratio of 7:2:1.

3.2. Experimental Results and Analysis

3.2.1. Ablation Experiment

To explore the influence of different modules within the MLI-HNN model on the performance of relationship extraction, ablation experiments were performed on distinct modules. The results of these ablation experiments are summarized in Table 1.

Table 1: Results of ablation experiments.

Model	Precision (%)	Recall (%)	F1-score (%)
MLI-HNN	87.03	89.2	88.1
-BiGRU	83.68	84.84	84.26
-CNN	85.13	85.47	85.30

Observing the table above, it is evident that eliminating the BiGRU layer leads to a 1.84% reduction in the model's F1 score, accompanied by a decrease in recall. This highlights the substantial impact of using BiGRU to extract long dependency relationships in text on relationship extraction. The absence of the CNN layer results in a 2.8% decrease in the model's F1 score, accompanied by reductions in precision and recall. This underscores the crucial role of CNN in capturing local features in the text. Removing the entity proximity features causes a 3.35% decrease in the model's F1 score.

3.2.2. Contrast Experiment

To establish a basis for comparison, the following models were selected:

BiGRU+WordAtt+SenATT: Utilizes BiGRU for sentence encoding, followed by attention mechanisms at the character and sentence levels to extract weighted vectors for relationship classification.

BiLSTM+SAtt: Employs BiLSTM to capture contextual information in the text, coupled with self-attention mechanisms for entity relationship extraction.

BERT: Constructs word vectors using pre-trained BERT models and performs entity relationship extraction through a fully connected network.

BERT+BiGRU-CNN: Constructs word vectors using pretrained BERT models, utilizes BiGRU to capture global text features, and incorporates CNN to capture local textual features for entity relationship extraction.

To ensure experimental reliability, all the above models were subjected to the same inputs. The performance comparison of different models is presented in Table 2.

Table 2: Performance comparison of different models.

Model	Precision (%)	Recall (%)	F1-score (%)
BiGRU+WordAtt+SenATT	81.67	50.08	62.09
BiLSTM+SAtt	75.23	74.46	74.84
BERT	79.48	73.14	76.18
BERT+BiGRU- CNN	87.21	86.57	86.89
MLI-HNN	91.03	92.20	92.10

MLI-HNN demonstrates superior performance in recall and F1 score compared to other models in Table 1. Models like BiGRU+WordAtt+SenATT and BiLSTM+SAtt utilize different attention mechanisms but struggle due to static word vectors from Word2Vec, limiting their effectiveness in capturing hierarchical information.

The BERT-based model achieves 79.48% accuracy, 73.14% recall, and 76.18% F1 score. BERT's dynamic word vectors enhance performance over static ones. Adding BiGRU and CNN models on top of BERT further boosts accuracy, recall, and F1 score, effectively capturing global and local features in longer texts.

Constructing semantically rich word vectors with BERT significantly enhances relationship extraction performance. Combining more model structures on top of BERT can further improve performance.

The MLI-HNN model achieves the highest F1 score of 91.03% by incorporating hierarchical information, using BiGRU for global features, CNN for local features, and combining neighboring word features for relationship extraction. This indicates its capability in extracting comprehensive text features and addressing challenges posed by complex grammatical structures in Chinese text.

4. Conclusion

The article presents the Multilevel Information Hybrid Neural Network (MLI-HNN) model to tackle the challenge posed by the complex grammatical structure of Chinese text, which makes capturing efficient semantic features difficult. MLI-HNN uses BERT pre-training to generate dynamic word vectors with richer semantics and employs a hybrid neural network to extract both contextual and local feature information. Additionally, a fully-connected network extracts relational feature information between entities and neighboring words. Experimental results show that MLI-HNN outperforms eight benchmark models in comprehensive text feature extraction. However, the method falls short in capturing finer-grained character relationships, suggesting future research explore finer annotation systems and diverse modeling strategies for extracting nuanced character relationships.

References

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding, 2019.

Rahul Dey and Fathi M. Salem. Gate-variants of gated recurrent unit (gru) neural networks. In *2017 IEEE 60th International Midwest Symposium on Circuits and Systems (MWSCAS)*, pages 1597–1600, 2017. doi: 10.1109/MWSCAS.2017.8053243.

HX Du, YL Shi, YC Cai, and LH Wang. Research on person relation extraction based on bert and convolutional neural network. *Journal of Northeast Normal University (Natural Science Edition)*, 53(03):49–55, 2021. ISSN 1000-1832. doi: 10.16163/j.cnki.22-1123/n.2021.03.010.

Sepp Hochreiter and Jürgen Schmidhuber. Long Short-Term Memory. *Neural Computation*, 9(8): 1735–1780, 11 1997. ISSN 0899-7667. doi: 10.1162/neco.1997.9.8.1735.

WJ Li, T Li, and F Qi. Chinese entity relation extraction based on multifeatures self-attention bi-lstm. *Journal of Chinese Information Processing*, 33(10):47–56,72, 1 2019.

Yankai Lin, Shiqi Shen, Zhiyuan Liu, Huanbo Luan, and Maosong Sun. Neural relation extraction with selective attention over instances. In Katrin Erk and Noah A. Smith, editors, *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 2124–2133, Berlin, Germany, August 2016. Association for Computational Linguistics. doi: 10.18653/v1/P16-1200.

Linlin Wang, Zhu Cao, Gerard de Melo, and Zhiyuan Liu. Relation classification via multi-level attention CNNs. In Katrin Erk and Noah A. Smith, editors, *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1298–1307, Berlin, Germany, August 2016. Association for Computational Linguistics. doi: 10.18653/v1/P16-1123.

Bowen YAO, Biqing ZENG, Jian CAI, and Meirong DING. Chinese character relation extraction model based on pre-training and multi-level information. *Journal of Computer Applications*, 41 (12):3637, 2021. doi: 10.

Dongxu Zhang and Dong Wang. Relation classification via recurrent neural network, 2015.

LX Zhang and WX Hu. Character relation extraction in chinese text based on bidirectional gru neural network and dual-attention mechanism. *u Neural Network And Dual-Attention Mechanism [J]. Computer Applications and Software*, 35(11):130–135,189, 1 2018.

Shu Zhang, Dequan Zheng, Xinchen Hu, and Ming Yang. Bidirectional long short-term memory networks for relation classification. In Hai Zhao, editor, *Proceedings of the 29th Pacific Asia Conference on Language, Information and Computation*, pages 73–78, Shanghai, China, October 2015.

C Zhao, SX Xie, DJ Zeng, F Zheng, T Cheng, and LH Peng. Combination of pre-trained language model and label dependency for relation extraction. *Journal of Chinese Information Processing*, 36(1):75–82, 1 2022.

Dandan ZHAO, Junpeng ZHANG, Jiana MENG, Zhihao ZHANG, and Wen SU. Medical entity relation extraction based on pre-trained model and hybrid neural network. *Acta Scientiarum Naturalium Universitatis Pekinensis*, 59(01):65–75, 2023. ISSN 0479-8023. doi: 10.13209/j.0479-8023.2022.065.

Peng Zhou, Wei Shi, Jun Tian, Zhenyu Qi, Bingchen Li, Hongwei Hao, and Bo Xu. Attention-based bidirectional long short-term memory networks for relation classification. In Katrin Erk and Noah A. Smith, editors, *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 207–212, Berlin, Germany, August 2016. Association for Computational Linguistics. doi: 10.18653/v1/P16-2034.