Chinese Relation Extraction with External Knowledge-Enhanced Semantic Understanding

Shulin Lv[1], Xiaoyao Ding[2]*
Information Technology and Data Management Center
Henan open University, Zhengzhou, China[1]
Henan open University, Zhengzhou, China[2]
Tanac Automation Co.,Ltd., Jiaxing, China[2]

***Abstract*—Relation extraction is the foundation of constructing knowledge graphs, and Chinese relation extraction is a particularly challenging aspect of this task. Most existing methods for Chinese relation extraction rely either on character-based or word-based features. However, the former struggles to capture contextual information between characters, while the latter is constrained by the quality of word segmentation, resulting in relatively low performance. To address this issue, a Chinese relation extraction model enhanced with external knowledge for semantic understanding is proposed. This model leverages external knowledge to improve semantic understanding in the text, thereby enhancing the performance of relation prediction between entity pairs. The approach consists of three main steps: first, the ERNIE pre-trained language model is used to convert textual information into dynamic word embeddings; second, an attention mechanism is employed to enrich the semantic representation of sentences containing entities, while external knowledge is used to mitigate the ambiguity of Chinese entity words as much as possible; and finally, the semantic representation enhanced with external knowledge is used as input for classification to make predictions. Experimental results demonstrate that the proposed model outperforms existing methods in Chinese relation extraction and offers better interpretability.**

***Keywords*—Chinese relation extraction; knowledge graph; external knowledge; semantic understanding; attention mechanism**

I. Introduction

Relation extraction is a critical subtask of information extraction, aiming to identify relationships between entity pairs from unstructured text based on semantic understanding. It plays a significant role in the construction of knowledge graphs. As an essential branch of relation extraction, Chinese relation extraction is crucial for downstream tasks such as Chinese semantic understanding and Chinese knowledge base construction. However, the complexity of Chinese semantics and the diversity of meanings in Chinese words have limited research in this area. Instead, most scholars have focused on relation extraction in English texts, which benefit from more abundant datasets. In practice, however, Chinese text is more commonly encountered. With the growing volume of Chinese text data, Chinese relation extraction technology can better meet real-world needs and serves as a key module in building Chinese knowledge bases[1]. Therefore, it is particularly important to develop an efficient and robust Chinese relation extraction model.

*Corresponding authors.

Currently, most researchers focus on relation extraction in English texts [2] [3], but the large volume of Chinese text demonstrates the necessity of advancing Chinese relation extraction. Its progress will directly impact the level of Chinese knowledge graph construction and indirectly promote the development of Chinese corpora. Thus, the construction of effective Chinese relation extraction models is a significant task. This paper conducts research on Chinese relation extraction models to address these challenges.

Compared to English relation extraction, Chinese text presents unique and significant challenges. First, Chinese text is semantically rich but structurally less rigid than English text. Second, Chinese relies heavily on function words to connect sentences or clauses, while English typically conveys sentence semantics and structures through word order. These challenges underscore the importance of understanding sentence semantics in Chinese relation extraction. For example, as shown in Fig. 1, consider the sentence “牛顿研究所有苹果” (“Newton Research Institute has apples”). The relationship between the entities “牛顿” (“Newton”) and “苹果” (“apple”) depends on the quality of word segmentation. For instance, the segmentation “牛顿/研究/所有/苹果” (“Newton/studies/all/apple”) suggests the relationship “study,” while the segmentation “牛顿/研究所/有/苹果” (“Newton/research institute/has/apple”) suggests the relationship “ownership.” Both segmentation results are valid when the sentence is isolated. Furthermore, the word “苹果” (“apple”) has two possible meanings: “fruit apple” and “Apple Inc.” Both interpretations are valid in the context of the sentence, but this ambiguity is a common challenge in Chinese text. These challenges impose higher requirements on Chinese relation extraction models. First, word segmentation must incorporate dynamic word embeddings, meaning segmentation must adapt to the context of the sentence. Second, attention mechanisms should be employed to capture not only the semantic meaning of entities but also the information from the sentence and the entire text. This ensures accurate word meanings and effectively resolves ambiguity.

Our paper proposes a Chinese relation extraction model enhanced with external knowledge for improved semantic understanding. First, the Chinese sentence containing the target entities is fed into the ERNIE [4] pre-trained model to obtain dynamic word embeddings. Then, an attention mechanism [5] is used to further enrich the semantics of the sentence where the entities are located, generating vector representations that incorporate sentence information. To further reduce ambiguity in Chinese words and enhance semantic understanding, the

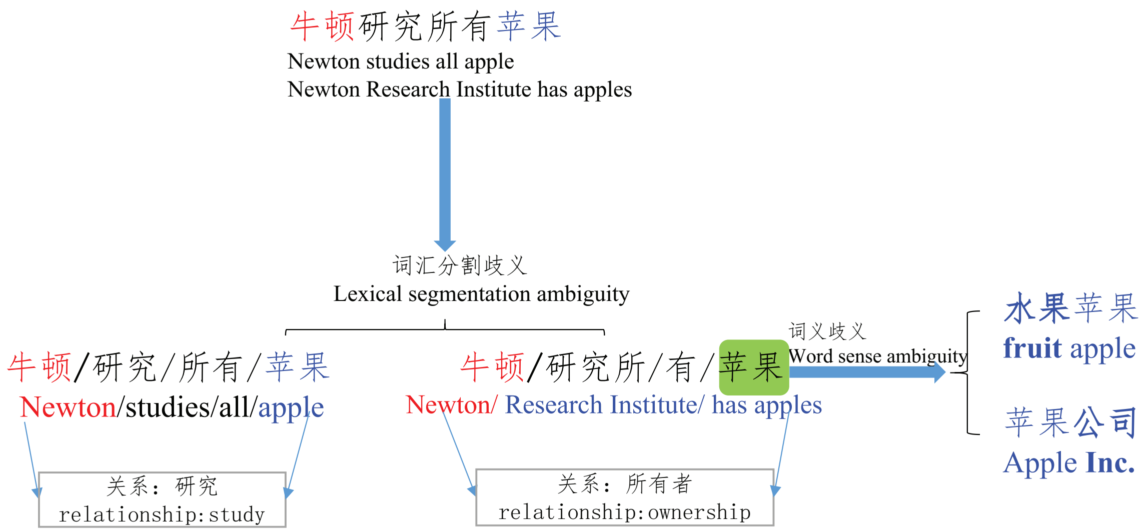

Fig. 1. Chinese relation extraction example.

model incorporates external knowledge, allowing the entities to refine their representations under the guidance of this external knowledge. These stages specifically address the challenges of insufficient entity semantics and entity ambiguity in Chinese relation extraction, enabling the model to pass richer semantic information to the classifier and achieve better performance in Chinese relation extraction. The proposed model was evaluated on a Chinese relation extraction dataset. Experimental results demonstrate that the model outperforms existing Chinese relation extraction methods.

II. Related Work

In recent years, neural networks and deep learning technologies have been widely and deeply applied, leading to rapid advancements in Chinese relation extraction within the field of natural language processing. Consequently, the number of studies and publications on Chinese relation extraction has been steadily increasing. The construction methods for relation extraction models can generally be categorized into two types: models based on traditional neural networks and models based on pre-trained language models.

A. Models Based on Traditional Neural Networks

In traditional neural network-based models, Convolutional Neural Networks(CNN) and Recurrent Neural Networks(RNN) are primarily applied. Liu et al. [6] were pioneers in proposing the use of CNNs to learn semantic features from text for relation extraction. Subsequent researchers enriched and extended CNNs, proposing models such as CNNs with max-pooling [7] and CNNs enhanced with attention mechanisms [8]. Although CNN-based models possess unique advantages in parallel computing, they exhibit significant shortcomings in semantic understanding and contextual modeling for Chinese text. Following this, researchers shifted their focus to RNNs. Zhang et al. [9] were among the first to apply RNNs to relation extraction models, achieving improved extraction performance. As a variant of RNNs, Long Short-Term Memory(LSTM) networks have also been widely used in Chinese relation extraction. Zhang and Yang [10] proposed the Lattice+LSTM model, which incorporates lexical information into the LSTM framework to better address the integration of character and word-level features. However, the model's performance in word segmentation remained constrained by the quality of word segmentation. To address this issue, Li et al. [11] proposed a multi-granularity lattice framework that leveraged potential semantic information from both characters and character sequences as input, thereby mitigating some ambiguity issues in entities. Similarly, Gao et al. [30] introduced the MGLT model, which integrates external lexical information and self-matching of lexical meanings to combine word-level features with their associated semantics, alleviating ambiguities in text.

Despite these advancements, traditional neural networks face inherent disadvantages when handling long-distance dependencies between entity pairs. As a result, an increasing number of researchers have turned to pre-trained language models to address the challenges of long-distance dependencies in Chinese relation extraction. This evolution reflects the shift from traditional approaches to more sophisticated techniques that leverage the power of pre-trained models for enhanced performance and contextual understanding in Chinese relation extraction.

B. Models Based on Pre-trained Language Models

In recent years, pre-trained language models have achieved remarkable success in the field of natural language processing, delivering superior performance in Chinese relation extraction tasks. To address the issues of character information loss and the inability to share lexical information in Li et al. [11] LSTM-based model, Kong et al. [12] proposed a hybrid approach combining LSTM and BERT [13] at the encoding layer. This method allows the character representations to include all matched lexical information, thereby mitigating the problem of information loss.

Eberts and Ulges [14] proposed a relation extraction model based on the pre-trained language model BERT, which incorporates the concept of spans. However, calculating the spans between every pair of characters results in high computational complexity. Zhong and Chen [15] introduced a model that uses two different encoders BERT and ALBERT [16] to independently learn features for entities and relationships. While using different encoders facilitates the representation of distinct features, it disrupts information sharing between entity representation and relation extraction. Zhou and Chen [17] proposed techniques to improve entity representation for enhanced extraction performance. Their model utilized BERT and RoBERTa [18] as encoders, with RoBERTa offering comprehensive optimizations over BERT, such as dynamic masking and sentence-level input. Cui et al. [19] introduced MacBERT, a further improvement over BERT, which demonstrated better results than RoBERTa in Chinese relation extraction. However, MacBERT still lacks external knowledge, making it less effective at resolving word sense disambiguation.Yang et al. [20] proposed a hybrid expert framework with BERT as the encoder. This framework dynamically learns multi-perspective semantic features by combining different granularities and views with the pre-trained model, which benefits Chinese relation extraction. Zhao et al. [21] introduced an ambiguity feedback mechanism to address word ambiguity, combining CNN and RoBERTa in the encoding module to effectively represent multi-granular information features. However, the performance of word-based representations was found to be inferior to character-based representations. Although models that integrate character and word-level features address the disadvantages of each approach, their use of contextual information remains limited. This results in unclear semantic representations for entities, leading to ambiguity. In 2019, Baidu introduced ERNIE [4], an improvement on BERT tailored for Chinese text. ERNIE incorporates masked training on continuous entity words and phrases to learn better semantic knowledge, thereby improving performance. However, it still falls short in addressing word sense disambiguation effectively.

Methods based on pre-trained language models have achieved more competitive performance, but two problems remain to be addressed. The first issue is that, although later pre-trained language models can convert input sequences into dynamic vector representations, thereby alleviating the problem of polysemy to some extent, they still fail to fully capture the in-depth understanding of word meanings in sentences, which is the foundation of Chinese relation extraction. The second issue is that the semantic information contained in target entities within sentences is still relatively insufficient, and in some cases, the representation of entities remains ambiguous, ultimately affecting relation extraction performance. Our model provides specific solutions to these two issues. For the first issue, the current ERNIE pre-trained model can effectively address word vector representation problems; however, relying solely on the ERNIE model is not enough. Attention mechanisms should be further utilized for downstream tasks to address the thin representation of word meanings and semantics. By emphasizing important words and the interdependencies between words in dynamic vectors through weighted attention, the semantic understanding of the model can be enhanced. For the second issue, to better avoid ambiguity in target entities, the assistance of external knowledge is required. External knowledge can provide supplementary context when entities are ambiguous, enriching the semantic representation of sentences. By addressing these two critical issues, the model ultimately enhances the semantic representation of entities and improves the performance of Chinese relation extraction.

III. MODEL

The proposed model framework, as shown in Fig. 2 is mainly divided into three levels: the encoding layer, the semantic understanding layer, and the classification layer. In the encoding layer, the sentence containing the entity pair whose relationship needs to be determined is first subjected to entity marking, and then input into the ERNIE pre-trained model to obtain dynamic word embeddings as output. In this way, the word embeddings gain entity awareness and contextual capturing ability during the encoding stage. In the semantic understanding layer, the self-attention mechanism is used to calculate the influence of other words in the sentence on the target entity pair, that is, the interaction weights between other words and the target entity pair. This allows the semantics of the sentence to be absorbed by the target entity pair. To avoid semantic ambiguity of the target entities, the HowNet [22] and ConceptNet[23] knowledge bases are used as external knowledge to supplement the representation of the target entities, further enhancing their semantic understanding ability. In the classification layer, the sentence semantic representation is fed into the classifier to compute the relationship type of the target entity pair.

A. Encoding Layer

To enable the pre-trained model to accurately identify the entity pair whose relationship needs to be determined, the entity pair in the sentence must first be marked. For example, in the sentence "凯特今天上午在超市购买了苹果" ("Kate bought apples at the supermarket this morning"), if we want to determine the relationship between the entities "凯特" ("Kate") and "苹果" ("apples"), the target entity pair "(凯特, 苹果)" needs to be marked before transforming the sentence into word vectors. The marked result would be: "ES_1凯特ED_1今天上午在超市购买了ES_2苹果ED_2." Here, "ES_1" and "ED_1" are the markers placed to the left and right of the first target entity "凯特", "ES_2" and "ED_2" are the markers placed to the left and right of the second target entity "苹果".

Once the sentence containing the target entity pair has been marked, it can proceed to the next step of vectorization. Assuming the marked sentence consists of *n* characters, where w_i represents the *i-th* character in the sentence, the sentence is then input into the ERNIE pre-trained model to obtain the following vectorized representation:

$$H = [h_1, h_2, h_3...h_n] = ERNIE([w_1, w_2, w_3, ...w_n]) \quad (1)$$

B. Semantic Understanding Layer

Semantic understanding involves integrating multi-granular semantic perception related to entities into the sentence representation. Multi-granularity specifically refers to modeling at the levels of words, sentences, and concepts.

The dynamic word vectors output by the ERNIE pre-trained model indeed contain information about the sentence where the entity pair resides to some extent. However, this

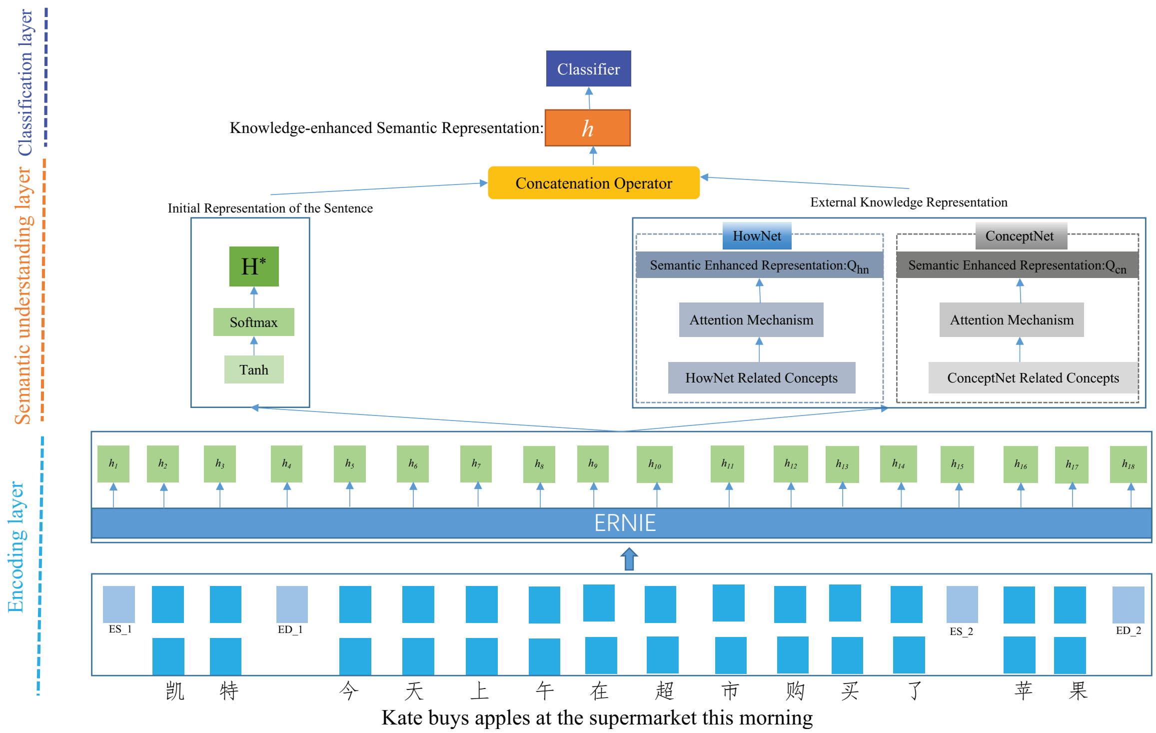

Fig. 2. Framework of the model.

semantic information is insufficient for more accurate prediction of the relationship between the entity pair. The sentence representation of the entities should encompass richer semantic information. Therefore, in semantic understanding, attention mechanisms and external knowledge are jointly used to enrich the sentence semantics.

Each character in the sentence contributes differently to the target entities. In other words, some characters are more helpful in determining the relationship between the entity pair. For example, in the sentence “凯特今天上午在超市购买了苹果” (“Kate bought apples at the supermarket this morning”), the word “超市” (“supermarket”) has a different impact on determining the relationship between the entity pair “(凯特, 苹果)” compared to “今天上午” (“this morning”). Characters that have a greater influence on the entity pair should be assigned higher weights, while those with less influence should be assigned lower weights. Using the attention mechanism, the initial representation of the sentence containing the entities can be calculated as follows:

$$\mathrm{S} = \tanh(\mathrm{H}) \tag{2}$$

$$\alpha = \mathrm{softmax}(b^{\mathrm{T}}\mathrm{S}) \tag{3}$$

$$\mathrm{H}^{*} = \mathrm{H}\alpha^{\mathrm{T}} \tag{4}$$

where $b \in R^{d^h}$ is a trainable parameter, and $\alpha \in R^n$ is a weight parameter.

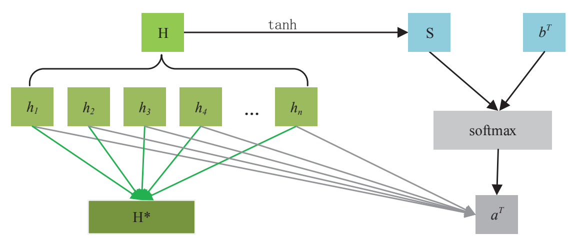

Fig. 3. The Process of computing the initial representation of the sentence in which the entity is located.

The specific calculation process of the initial representation of the sentence vector where the entity resides is shown in Fig. 3. During the model training process, the vectorized representation of the sentence is first subjected to tanh activation to alleviate the gradient vanishing problem during training. Then, the output S of this process is multiplied by the transpose of the trainable parameter b to calculate the weight a of each character's contribution to the entities in the sentence. Based on the obtained weight values, the vectorized representation of the sentence is updated. The model's prediction results based on the new sentence vector are compared with the actual labels in the data to calculate the loss function value. The error is used to compute parameter gradients via backpropagation, and the trainable parameters of the model are updated using the gradient descent method. In this way, through the attention mechanism, character-level features are integrated into sentence-level features.

Considering the particularity and complexity of Chinese word meanings, the same word can have different meanings in different contexts. As shown in the example in Fig. 1, the word “苹果” (“apple”) has different meanings in different contexts. For instance, it could mean “水果苹果” (“fruit apple”) or “水果公司”(“Apple Inc.”) depending on the context. However, previous models could not correctly distinguish these word meanings due to the lack of prior knowledge. In contrast, the model proposed in this paper introduces two external knowledge bases, HowNet and ConceptNet, aiming to further enhance the model’s ability to handle word sense disambiguation.

For the HowNet knowledge base, an official API interface is provided, which can be directly called for use. For example, using the API interface to search for meanings related to “苹果” (“apple”) returns external knowledge such as “[fruit|水果, tool|用具, PatternValue|样式值, able|能, bring|携带, SpeBrand|特定牌子, communicate|交流].” This knowledge serves as a rich and complementary source of information for “苹果. Assuming that an entity retrieves k related concepts through HowNet, the softmax function is applied to normalize their weight values, obtaining the attention weight c_k corresponding to each concept of the entity. Finally, the attention mechanism is used to enhance the semantic representation of the external knowledge from HowNet corresponding to the entity:

$$\beta_j = \frac{\exp(\mathrm{c}_k)}{\sum\limits_k \exp(\mathrm{c}_k)} \tag{5}$$

$$O_{\mathrm{hn}} = \sum_k \beta_j e_i \tag{6}$$

For the ConceptNet knowledge base, an official API interface is also provided. Referring to the usage of HowNet mentioned above, we can retrieve the keyword “苹果” (“apple”) through the interface and obtain knowledge relationships such as “水果、实物、植物、商品” (“fruit、physical object、plant、commodity”) etc., each with different weights. Similarly, k related entity relationship concepts are used, and their weights are normalized using the softmax function. The attention mechanism is then applied to obtain the semantic representation of the external knowledge from ConceptNet corresponding to the entity:

$$\gamma_t = \frac{\exp(\mathrm{q}_k)}{\sum\limits_k \exp(\mathrm{q}_k)} \tag{7}$$

$$O_{\mathrm{cn}} = \sum_k \gamma_t e_i \tag{8}$$

Finally, the initial sentence representation, the semantic representation of external knowledge from HowNet, and the semantic representation of external knowledge from ConceptNet are concatenated to obtain the external knowledge-enhanced semantic representation.

$$\mathrm{h} = [\mathrm{H}^*; O_{\mathrm{hn}}; O_{\mathrm{cn}}] \tag{9}$$

where $[;]$ is the concatenation operator.

C. Classification Layer

Through the effect of multi-level semantic awareness, the final representation of the sentence containing the entity pair is obtained. The predefined Chinese relation extraction task is treated as a binary classification task for each relation, and the sigmoid function is used to calculate the probability of relation r in the set of relations:

$$p_r = \mathrm{sigmoid}(W_r h + b_r) \tag{10}$$

where $p_{\mathrm{r}} \in R^{|\Re|}$, W and b are trainable parameters.

Finally, the binary cross-entropy is used to define the loss function, and during model training, the Adam optimizer [24] is employed to adjust the loss function. The loss function is defined as follows:

$$\mathcal{L}_{loss} = -\sum_{r\in\mathcal{R}} \left(y_r \log(p_r) + (1-y_r)\log(1-p_r)\right) \tag{11}$$

where $y_r \in \{0,1\}$ represents the true value of the relation label r.

IV. Experiment

A. Dataset

The model proposed in this paper is evaluated on the SanWen [25] and FinRE [26] datasets, which are commonly used in multiple studies.

The SanWen dataset is a manually annotated Chinese dataset containing 837 documents with a total of 21,240 sentences. Among them, 81.1% are randomly selected as the training set, 8.4% as the validation set, and the remaining 10.5% as the test set. There are nine types of relationships between entities, including location, proximity, part-whole, general-specific relationship, family, social, ownership, usage, and creation.

The FinRE dataset consists of 18,702 instances extracted from 2,647 Sina Finance news articles. Of these, 13,486 instances are used as the training set, 1,489 as the validation set, and the remaining 3,727 as the test set. It includes 44 types of relationships between entities, such as competition, cooperation, stock reduction, and other financial-specific relationships.

B. Evaluation

For model evaluation, the commonly used performance metric F1-score is adopted. The application of the F1-score is primarily aimed at balancing precision and recall, ensuring that both are taken into account as much as possible. The three performance calculation methods are as follows:

$$Precision = \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FP}} \tag{12}$$

$$Recall = \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FN}} \tag{13}$$

$$F1 = \frac{2 * Precision * Recall}{Precision + Recall} \quad (14)$$

where TP (True Positive) indicates instances where the model predicts a positive instance, and it is indeed a positive instance in reality. FP (False Positive) indicates instances where the model predicts a positive instance, but it is actually a negative instance. FN (False Negative) indicates instances where the model predicts a negative instance, but it is actually a positive instance.

C. Experimental Setup

Table I presents the parameter settings of the model at its best performance. The model is based on the ERNIE encoder, so the dimension of the output dynamic vector is 768, which is consistent with BERT. During the process of parameter tuning, the dropout rate was set to 0.5, the initial learning rate was set to 1.0, and the optimal learning rate was 5e-5. The optimal training model can be achieved after 100 epochs.

TABLE I. Hyper-Parameters Settings

Parameter	Value
Batch Size	10
Encoder Hidden Size	768
Optimizer	Adam
Dropout	0.5
Learning Rate	0.0005
Epoch	100

D. Compared Models

To better compare and study the proposed model, we evaluate its performance against other models on the same dataset. The compared models include both traditional neural network-based models and pre-trained language model-based approaches (1) Traditional Neural Network-Based Models: PCNN+Att [8], Lattice+LSTM [10], Lattice+MG [11], PCNN [27], MGRSA [28], MGLT [30]. (2) Pre-Trained Language Model-Based Models: ERNIE [4], PURE [15], IERE [17], PRM [21], OPT-FLAT [29].

V. Experiments and Results

A. SanWen Dataset Results

The results of the proposed model on the SanWen dataset are shown in Table II. It can be observed that our model outperforms existing models, with the F1-score surpassing the current best model by 0.41. This demonstrates the critical importance of processing the semantics of the sentences where the entities are located. Moreover, relying solely on the semantics of the sentence itself cannot achieve optimal performance; external knowledge must be incorporated as a supplement. When the model encounters ambiguity in word meanings, external knowledge can effectively resolve this issue. Additionally, the ERNIE pre-trained model significantly enhances the understanding of entity words, further contributing to improved performance.

TABLE II. Experiment Results on SanWen

Method	F1/%
PCNN+Att[8]	60.55
PCNN[27]	61.23
ERNIE[4]	63.25
Lattice+LSTM[10]	63.88
IERE[17]	63.99
PURE[15]	64.70
Lattice+MG[11]	65.61
MGRSA[28]	67.12
PRM[21]	67.72
OPT-FLAT[29]	68.35
MGLT[30]	69.50
Our model	**69.91**

TABLE III. Experiment Results on FinRE

Method	F1/%
PCNN[27]	45.51
PCNN+Att[8]	46.13
Lattice+LSTM[10]	47.41
ERNIE[4]	47.45
IERE[17]	49.09
Lattice+MG[11]	49.26
PURE[15]	46.61
OPT-FLAT[29]	50.60
MGRSA[28]	52.61
PRM[21]	52.97
MGLT[30]	53.22
Our model	**53.47**

B. FinRE Dataset Results

The results of the proposed model on the FinRE dataset are shown in Table III. It can be observed that our model also outperforms existing models, with the F1-score surpassing the current best model by 0.25. However, it is evident that the advantage of the proposed model on the FinRE dataset is lower compared to its performance on the SanWen dataset. After analysis, this difference can be attributed to the varying number of relationship types in the two datasets. The SanWen dataset contains nine types of relationships, whereas the FinRE dataset includes 44 types. This discrepancy increases the difficulty of relationship extraction, as distinguishing between similar relationships imposes higher demands on the model.Moreover, the two types of external knowledge incorporated in the proposed model abstract entities into concepts, which indirectly amplifies the influence of entities on their corresponding relationship types. This highlights a potential direction for future model improvements: ensuring greater precision when incorporating external knowledge to further enhance performance.

C. Ablation Study

From Tables II and III, it can be seen that the proposed model demonstrates strong competitiveness compared to other models. To further explore the role of the three components in the model, we conducted an ablation study on the SanWen dataset. Based on the analysis above, the key modules of the model include the ERNIE encoder, the sentence representation module, and the external knowledge representation module. To clearly observe the contribution of each module, we disabled one module at a time and analyzed the results. First, we replaced the ERNIE encoder with the Chinese version of BERT-base. The results showed that the model's performance dropped by 2.01, indicating that the ERNIE encoder helps alleviate the issue of insufficient semantic information at a fundamental level. Its enhanced pre-training strategies play an important role in understanding entity-related semantics. Second, when the sentence representation module was removed, the model's performance dropped significantly by 2.73. This result shows that the sentence representation module is critical for improving the model's performance. By leveraging the attention mechanism, this module enhances the characters in the sentence that are important to the entities, thereby enriching the semantic representation of entity words in a targeted manner. Finally, when the external knowledge representation module was removed, the model's performance showed a slight decrease. This indicates that external knowledge is helpful in resolving word ambiguities, but it also introduces a certain amount of noise. As a result, the improvement brought by this module is not as significant as the other components. From the ablation experiments, it can be concluded that the ERNIE encoder and the sentence representation module play a core role in enhancing the model's performance, the external knowledge representation module helps the model handle ambiguity issues. However, due to the complexity and diversity of external knowledge sources, their quality cannot be fully guaranteed, and they may contain information that is irrelevant or even contradictory to the current task. Such low-quality or irrelevant knowledge may interfere with the model's learning process, leading to noise accumulation and ultimately affecting the extraction performance. This issue becomes particularly prominent when dealing with high-noise knowledge sources or when the knowledge integration method lacks precision. Therefore, the presence of noise in external knowledge is also a limitation of this study. From Table IV, it can also be seen that the three modules in the model all contribute to improving the overall performance of the model to varying degrees. The organic combination of these three modules ultimately leads to a significant performance improvement compared to previous models.

TABLE IV. ABLATION EXPERIMENT

Parameter	F1/%
Our Model	69.91
- ERNIE encoder	67.90
- Sentence representation module	67.18
- External knowledge representation module	68.24

VI. CONCLUSION

This paper proposes a Chinese relation extraction model enhanced by external knowledge to improve semantic understanding. Experiments demonstrate that the proposed model achieves better performance in handling Chinese relation extraction tasks. In future research, we will further explore more precise knowledge filtering and integration strategies to maximize the benefits of external knowledge while minimizing the introduction of noise, ultimately improving the model's stability and generalization ability.

ACKNOWLEDGMENT

This work is supported by the Key Scientific and Technological Project of Henan Province (242102320163, 252102211058, 252102210152, 252102210122), the Key Scientific Research Project of Higher Education Institutions in Henan Province (24B520046, 25A520058), and the Research Project of Henan Federation of Social Sciences (SKL-2024-1670, SKL-2024-997).

REFERENCES

[1] J. Zhang, K. Hao, X.Tang, et al. A multi-feature fusion model for Chinese relation extraction with entity sense[J]. Knowledge-Based Systems, 2020, 206: 106348.

[2] H. Tang, Y. Cao, Z. Zhang, et al. HIN: Hierarchical inference network for document-level relation extraction[C]//Proceedings of the 24th Pacific-Asia Conference. Singapore: PAKDD, 2020: 197-209.

[3] Q. Tan, R. He, L. Bing, et al. Document-Level Relation Extraction with Adaptive Focal Loss and Knowledge Distillation[C]//Findings of the Association for Computational Linguistics. Dublin: ACL, 2022: 1672-1681.

[4] Y. Sun, S. Wang, Y. Li, et al. Ernie 2.0: A continual pre-training framework for language understanding[C]//Proceedings of the AAAI Conference on Artificial Intelligence. New York: AAAI, 2020, 34(05): 8968-8975.

[5] A. Vaswani, N. Shazee, N. Parmar, et al. Attention is all you need[C]//31st Annual Conference on Neural Information Processing Systems. Long Beach: NIPS, 2017: 5998-6008.

[6] C. Y. Liu, W. B. Sun, W. H. Chao, et al. Convolution neural network for relation extraction[C]//Advanced Data Mining and Applications: 9th International Conference. Hangzhou: ADMA, 2013: 231-242.

[7] D. Zeng, K. Liu, S. Lai, et al. Relation classification via convolutional deep neural network[C]//Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics. Dublin: COLING, 2014: 2335-2344.

[8] Y. Shen, X. J. Huang. Attention-based convolutional neural network for semantic relation extraction[C]//Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics. Osaka: COLING, 2016: 2526-2536.

[9] D. Zhang, D. Wang. Relation classification via recurrent neural network[J]. arXiv preprint arXiv:1508.01006, 2015.

[10] Y. Zhang, J. Yang. Chinese NER using lattice LSTM[C]// Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics. Melbourne: ACL, 2018:1554-1564.

[11] Z. Li, N. Ding, Z. Liu, et al. Chinese relation extraction with multi-grained information and external linguistic knowledge[C]//Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. Firenze: ACL, 2019: 4377-4386.

[12] B. Kong, S. Liu, F. Wei, et al. Chinese Relation Extraction Using Extend Softword[J]. IEEE Access, 2021, 9: 110299-110308.

[13] J. Devlin, M. W. Chang, K. Lee, et al. BERT: Pre-training of deep bidirectional transformers for language understanding[C]//Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics. Minneapolis: NAACL, 2019: 4171–4186.

[14] M. Eberts, A. Ulges. Span-Based Joint Entity and Relation Extraction with Transformer Pre-Training[C]// European Conference on Artificial Intelligence. Santiago: ECAI, 2020.

[15] Z. Zhong, D. Chen. A Frustratingly Easy Approach for Entity and Relation Extraction[C]//Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics. Mexico City: NAACL, 2021: 50-61.

[16] Z. Lan, M. Chen, S. Goodman, et al. ALBERT: A lite bert for self-supervised learning of language representations[C]//In International Conference on Learning Representations. Addis Ababa: ICLR, 2020.

[17] W. Zhou, M. Chen. An improved baseline for sentence-level relation extraction[J]. arXiv preprint arXiv:2102.01373, 2021.

[18] Y. Liu, M. Ott, N. Goyal, et al. Roberta: A robustly optimized bert pretraining approach[J]. arXiv preprint arXiv:1907.11692, 2019.

[19] Y. Cui, W. Che, T. Liu, et al. Revisiting Pre-Trained Models for Chinese Natural Language Processing[C]//Findings of the Association for Computational Linguistics. Punta Cana: EMNLP, 2020: 657-66.

[20] J. Yang, B. Ji, S. Li, et al. Dynamic Multi-View Fusion Mechanism For Chinese Relation Extraction[C]//Advances in Knowledge Discovery and Data Mining. PAKDD, 2023: 405–417.

[21] Q. Zhao, T. Gao, N. Guo. A novel chinese relation extraction method using polysemy rethinking mechanism[J]. Applied Intelligence, 2023, 53(7): 7665-7676.

[22] Z. Dong, Q. Dong. HowNet-a hybrid language and knowledge resource[C]//International conference on natural language processing and knowledge engineering. San Francisco: IEEE, 2003: 820-824.

[23] R. Speer, J. Chin, C. Havasi. Conceptnet 5.5: An open multilingual graph of general knowledge.[C]. Proceedings of the 31st AAAI Conference on Artifcial Intelligence, 2017: 4444–4451.

[24] D. P. Kingma, J. Ba. Adam: A method for stochastic optimization[J]. arXiv preprint arXiv:1412.6980, 2014.

[25] J. Xu, J. Wen, X. Sun, et al. A discourse-level named entity recognition and relation extraction dataset for chinese literature text[J]. arXiv preprint arXiv:1711.07010, 2017.

[26] Z. R. Li, N. Ding, Z. Y. Liu, et al. Chinese relation extraction with multi-grained information and external linguistic knowledge[C]//Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. ACL, 2019: 4377-4386.

[27] D. Zeng, K. Liu, Y. Chen, et al. Distant supervision for relation extraction via piecewise convolutional neural networks[C]//Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing. Lisbon: EMNLP, 2015: 1753-1762.

[28] Z. Zhong. Chinese Entity Relation Extraction Based on Multi-level Gated Recurrent Mechanism and Self-attention[C]//2021 2nd International Conference on Artificial Intelligence and Information Systems. Chongqing: ICAIIS, 2021: 1-7.

[29] X. Zeng X, J. Zhong, C. Wang, et al. Chinese relation extraction with flat-lattice encoding and pretrain-transfer strategy[C]//Knowledge Science, Engineering and Management: 14th International Conference. Tokyo: KSEM, 2021: 30-40.

[30] Y. B. Gao, G. H. Gong, and L. Ni. Chinese relation extraction in military field based on multi-grained lattice transformer and imbalanced data classification[J]International Journal of Modeling, Simulation, and Scientific Computing, 2024, 15(05):2350052.