NCRE: A Benchmark for Document-level Nominal Compound Relation Extraction

Jincheng Cao, Bobo Li*, Jiang Liu, Donghong Ji*

Key Laboratory of Aerospace Information Security and Trusted Computing, Ministry of Education, School of Cyber Science and Engineering, Wuhan University, Wuhan, China
{caojincheng,boboli,liujiang,dhji}@whu.edu.cn

Abstract

Entity and relation extraction is a conventional task in the field of information extraction. Existing work primarily focuses on detecting specific relations between entities, often constrained to particular fields and lacking general applicability. In response, we propose a novel task: nominal compound relation extraction (NCRE), which concentrates on abstract and broadly applicable relation extraction between noun phrases. This task diverges significantly from traditional entity and relation extraction in two key respects. Firstly, our task involves general nominal compounds rather than named entities, which are longer and encompass a broader scope, presenting significant challenges for extraction. Secondly, relation extraction in NCRE demands an in-depth understanding of context to detect abstract relations. We manually annotate a high-quality Chinese dataset for the NCRE task and develop a model incorporating the rotary position-enhanced word pair (RoWP) detection schema. Experimental results demonstrate the efficiency of our RoWP model over previous baselines, while the suboptimal F1 scores indicate that NCRE remains a challenging task. Our code and data are available at https://github.com/yeecjc/NCRE.

1 Introduction

Noun phrases (Breheny, 2008) are fundamental units in information extraction and knowledge representation, encapsulating rich semantic information. The relationships among noun phrases not only influence semantic understanding but are also core to effective text content analysis. For instance, in medical and legal documents, precise comprehension of noun phrases and their interrelationships enables the rapid extraction of essential information (Merabti et al., 2014) , thus enhancing the efficiency of information retrieval. This underscores the importance of complex relationships among noun phrases.

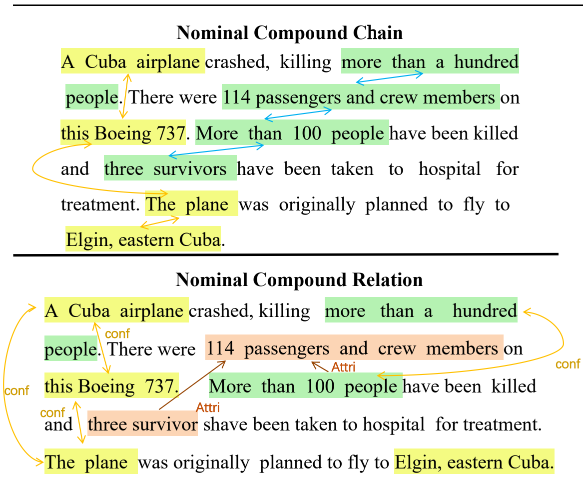

Figure 1: The difference between NCCE and NCRE tasks. Among them, *conf* refers to the coreference relationship, and *Attri* refers to the AttributesName relationship.

Significant advancements have been made in noun phrase extraction techniques, such as structural prediction (Matusevych and Culbertson, 2022) and graph neural network-based extraction (Gui et al., 2019). However, the extraction for relationships among noun phrases remain underexplored. Nominal compound chain extraction (Li et al., 2020) has been proposed to detect the chain relations of nominal compounds; however, it primarily focuses on semantically related relations and fails to distinguish among types of relations or detect complex relationships. Moreover, although named entity relation extraction has defined multiple types of relations, such as those between humans and organizations or places (Yao et al., 2019), these categories are limited. For noun phrases, it is impractical to predetermine all types of named entities and potential relations. Furthermore, the performance of open relation extraction (Wang et al.,

*These authors are co-corresponding authors.

Proceedings of the 31st International Conference on Computational Linguistics, pages 10531–10540
January 19–24, 2025. ©2025 Association for Computational Linguistics

2021) is still suboptimal (Dong et al., 2023), indicating that current methods and datasets struggle to comprehensively address noun phrases and their relations from a general and abstract perspective.

In light of these challenges, it is necessary to develop a benchmark to support the evaluation of relationships between noun phrases. In this paper, we categorize noun phrase relations (Huang et al., 2019) and select several key types, such as *Value* relationships and *element* (group-part) relationships, to denote the various relationships between nouns. These relationships do not depend on specific entity types and extend beyond coreference, making them suitable for describing inter-noun relationships. For example, as illustrated in Figure 1, *AttributesNames* relationships can exist between different entity pairs. Thus, this framework is highly suitable for describing the relationships between noun phrases.

We have developed a Chinese dataset named Nominal Compound Relation Extraction (NCRE). The corpus is derived from the social news sections of public news websites, such as Xinhua.net. We selected 1,025 documents for annotation, initially utilizing the Stanza Parser (Qi et al., 2020) to determine the boundaries of the longest nominal compounds, while splitting compounds to avoid continuous 'of' sequences. By adopting the longest noun phrase method, the issues of entity nesting and ambiguity are effectively avoided. Subsequently, we annotated the relationships between nominal compounds using crowdsourcing. Rigorous cross-validation ensured the dataset's high quality, providing a reliable benchmark for the comprehensive evaluation of extraction methods in the NCRE task.

Two primary challenges arise in conducting the NCRE task. Firstly, in the NCRE task, each noun phrase consists of the longest nominal compounds, which presents a challenge in determining precise boundaries for these compounds. Secondly, models should not only consider the direct relationships between compounds but also abstract and categorize different types of relationships, further complicating the extraction task.

In response, we devised a model named Rotary Position Enhanced Token Pair Relation (RoWP) to address this task. RoWP builds on the pairwise relationships between tokens, employing Multi-granularity convolution (Li et al., 2022) to capture pair relations at varying distances. Additionally, we adopted Rotary Position Embedding (RoPE) (Su et al., 2021) to ascertain the relative distances between tokens, combined with biaffine (Li et al., 2021b) and MLP to precisely recognize and classify the relationships between tokens in the document.

We conducted a series of experiments on the NCRE dataset. The results demonstrate that the RoWP model outperforms previous work on multiple evaluation metrics. The analysis further demonstrates ts excellent performance in detecting long compounds and extracting relationships across sentences.

Finally, the contributions of our work can be summarized as follows:

- We first introduce the document-level nominal compound relation extraction (NCRE) task, which is aimed at detecting abstract relationships between noun phrases. The relation in NCRE is broader and more general in scope, clearly distinguishing it from traditional entity relation tasks.

- We provide a high-quality, large-scale Chinese NCRE dataset, establishing a robust benchmark for understanding nominal compounds.

- Our proposed RoWP model shows significant superiority in the NCRE task, optimizing detection of long compounds, long distances, and cross-sentence extraction.

2 Related Work

2.1 Relation Triple Extraction Methods

Relation Triple Extraction methods can be summarized into four categories, based on different extraction ideas and technical means. Firstly, tagging-based methods identify entities and relations through multiple sequence annotation modules, such as the novel annotation scheme proposed by Zheng et al. (2017) to transform the RTE task into a sequence labeling problem. Furthermore, Zheng et al. (2021) proposed a novel extractor to reduce redundancy in relation prediction by introducing latent relations and global correspondence mechanisms. The Bi-RTE framework proposed by Ren et al. (2022) adopts a bidirectional extraction strategy to identify triples based on entity pairs extracted from complementary directions. Secondly, table-filling methods identify head and tail entities by classifying the relationship between token pairs. TPLinker (Wang et al., 2020) is a typical example, which adopts an innovative alignment boundary labeling scheme to identify entity pairs under each relation type. Thirdly, text generation methods, such as the work Zeng et al. (2018) and Ye et al. (2021), adopt a sequence-to-sequence (seq2seq) structure to generate triples. Finally, graph linking methods, such as the research of Shang et al. (2022b), regard the triple extraction problem as a graph linking the problem of enumerating candidate entities on a bipartite graph. Despite this, there are relatively few studies on document-level triple extraction.

2.2 Long Nominal Compound Analysis

Mascarell (2017) utilized word embedding technology to compute semantic similarity and enhanced results by integrating contextual information. Qian (2019) extensively investigated the multi-level distribution characteristics of Chinese noun phrases, identifying three major challenges: data sparsity, structural ambiguity, and boundary uncertainty. To tackle these challenges, they proposed a recognition method based on a conditional random field model and a noun block lifting rule. However, existing joint extraction methods often overlook latent semantic information, such as topic information (Fei et al., 2020). There are few studies on noun compounds, but noun compounds contain rich information and have great research value.

3 Dataset

3.1 Data construction

To facilitate our task, we manually annotated a high-quality Chinese dataset based on the brat platform (Stenetorp et al., 2011) [1]. Experienced computer and linguistics experts built an annotation manual. Undergraduate and master students of computer science were hired to annotate. Master's students underwent training and testing to ensure annotation quality, with only those meeting the assessment criteria proceeding to formal annotation.

In this process, we also improved and revised the annotation manual. In the end, 10 students were involved in the annotation. Two students annotated noun compounds and relations for each document, and each student was responsible for 100 documents on average. For inconsistencies in the annotation, linguistic experts determined the final annotation results. Then the language experts proofread in Chinese to ensure high consistency of the labels and data quality.

[1] http://brat.nlplab.org

	Train	Dev	Test	Total
Document	821	102	102	1025
Compound	30143	5026	5036	40205
Max.compound length	142	136	172	150
Avg.compound length	4.17	4.13	4.53	4.28
Relation	9796	1655	1059	12510
Intra-sentence-relation	6106	894	675	7675
Cross-sentence-relation	3985	766	384	4835
RE.Coreference	3954	747	384	5085
RE.AttributesName	1865	191	215	2271
RE.Agent-Patient	928	139	87	1154
RE.Value	978	146	111	1235
RE.Possess	775	183	84	1042
RE.Element	705	123	70	898
RE.Generic	360	114	88	562
RE.AttributeMul	232	12	21	265

Table 1: The statistics of the dataset. We counted the number of entities and relations, the average entity length, the maximum entity length, and the number of each relation.

3.2 Dataset Analysis and Insights

The final dataset comprises 1,025 documents and 40,205 nominal compounds, encompassing a total of 12,510 relationshipsk. We randomly divided the data into a training set, a development set, and a test set, with 821, 102, and 102 documents respectively. Table 1 shows the statistics of the dataset.

In the table, we first counted the number of documents in the dataset. At the entity level, we counted the total number of noun compounds and recorded the maximum and average length of noun compounds in the dataset, which helps us understand the complexity of the entities. At the relationship level, we counted the number of all relationship types defined in the dataset. Furthermore, we distinguished between cross-sentence relationships and intra-sentence relationships to evaluate the performance of the model when dealing with different contextual relationships.

In Table 2, we compare our dataset with existing entity relationship extraction datasets. For example, the document-level relationship CDR dataset is oriented to a specific biological field, while our dataset mainly comes from news, including politics, military, technology, and finance, covering many fields and providing a wider range of application scenarios. In addition, the definition of relations in our dataset is an abstract semantic relationship, not defined for a specific type of entity. This design makes our dataset universal, useful, and comprehensive.

	DuIE	CDR	DocRED	NCRE
Entity	✓	✓	✓	✓
Relation	✓	✓	✓	✓
Document level	✗	✓	✓	✓
Multi-field	✓	✗	✓	✓
General Relation	✗	✗	✗	✓

Table 2: A comparison between our NCRE dataset and existing popular entity relation extraction datasets, including: DuIE (Li et al., 2019), CDR (Li et al., 2016), and DocRED (Yao et al., 2019).

4 Methodology

The architecture of our framework is shown in Figure 2 and consists of three main parts. First, we utilize widely adopted pre-trained language models, specifically BERT (Devlin et al., 2019) and bidirectional LSTM (Lample et al., 2016), as encoders to generate contextual word representations from the input sentences. Next, we employ convolutional layers to construct and refine the representation of the word grid, facilitating subsequent word-word relationship classification. Finally, we integrate a co-predictor layer (Li et al., 2021b), which includes a biaffine classifier and a multilayer perceptron, with the prediction layer of our proposed RoWP to jointly infer the relations between all word pairs.

4.1 Task Definition

Given a document $\mathrm{S} = \{w_1, w_2, \ldots, w_N\}$, where N is the length of the document. The NCRE task aims to extract the set of all potential relation triplets.

4.2 Encoder layer

We use BERT (Devlin et al., 2019) as the text encoder of our model. Given a document S = { w_1, w_2, …, w_N }, and input them into a pre-trained model BERT. To further enhance context modeling, we adopted bidirectional LSTM (Lample et al., 2016) based on previous work (Li et al., 2021a). After BERT encoding, the document S can be represented as:

$$H = \{h_1, h_2, \ldots, h_N\}, \tag{1}$$

where $h_i \in \mathbb{R}^{d_h}$ is the representation of the i-th word and d_h represents the dimension of a word representation. Since all the documents in our dataset are no more than 512 in length, it is possible to encode the documents directly in BERT.

4.3 Convolution Layer

We utilize convolutional neural networks (CNNs) to refine representations. First, we apply a conditional normalization layer (CLN) (Liu et al., 2021) to generate word representations on a grid. This grid can be viewed as a three-dimensional matrix $V \in \mathbb{R}^{N \times N \times d_h}$, where each element V_{ij} represents a word pair (x_i, x_j):

$$V_{ij} = CLN(h_i, h_j) = \gamma_{ij} \odot (\frac{h_j - \mu}{\sigma}) + \lambda_{ij}, \tag{2}$$

where h_i is the condition to generate the gain parameter $\gamma_{ij} = W_\alpha h_i + b_\alpha$ and bias $\lambda_{ij} = W_\beta h_i + b_\beta$ of layer normalization. μ and σ are the mean and standard deviation across the elements of h_j.

To enhance the grid representation, we incorporate the relative word position information $E^d \in \mathbb{R}^{N \times N \times d_{E_d}}$ for each word pair, and the grid position information $E^t \in \mathbb{R}^{N \times N \times d_{E_t}}$, which differentiates the upper and lower triangular areas of the grid. These positional embeddings are then integrated with the word pair information $V \in \mathbb{R}^{N \times N \times d_h}$, This fused representation is processed through a multi-layer perceptron to produce the position-aware contextual representation $C \in \mathbb{R}^{N \times N \times d_c}$. The overall process can be formulated as:

$$C = MLP_1([V; E^d; E^t])\,, \tag{3}$$

where the tensor $V \in \mathbb{R}^{N \times N \times d_h}$ represents word information, a tensor $E^d \in \mathbb{R}^{N \times N \times d_{E_d}}$ represents the relative position information between each pair of words, and a tensor $E^t \in \mathbb{R}^{N \times N \times d_{E_t}}$ represents the region information for distinguishing lower and upper triangle regions in the grid.

Next, we employ multiple 2D dilated convolutions (DConv) with varying dilation rates to capture the interactions among words at different distances. This process is formulated as follows:

$$Q = GeLU(DConv(C))\,, \tag{4}$$

where $Q \in \mathbb{R}^{N \times N \times d_q}$ is the output and $GELU$ is a activation function.

4.4 Co-Predictor Layer

After the convolutional layers, we obtain the word grid representation Q, which is fed into the MLP to predict the relationship between each pair of words. In addition, we enhance the relation classification by combining the MLP predictor with a

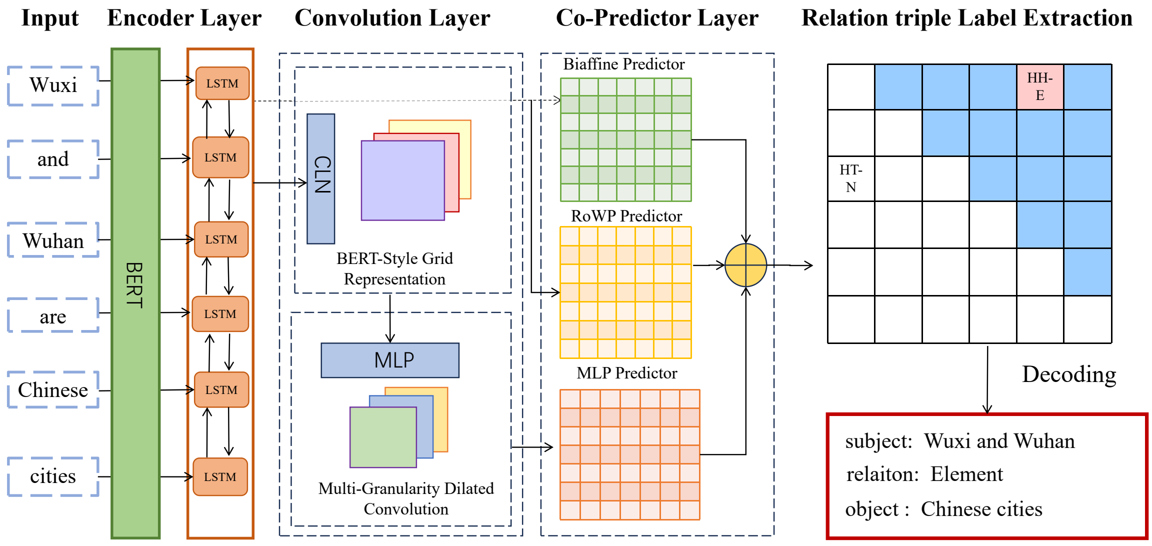

Figure 2: Overall model architecture. CLN and MLP represent conditional layer normalization and multi-layer perceptron. ⊕ represents element-wise addition.

biaffine predictor. At the same time, we added a RoWP prediction layer. Therefore, we use these three predictors to simultaneously calculate two independent relation distributions (x_i, x_j) for word pairs and combine them as the final prediction.

Biaffine Predictor The input to the biaffine predictor is the output $H = \{h_1, h_2, \ldots, h_N\}$ of the encoder layer. We employ two MLPs to compute the word representations s_i and o_j for the subject (x_i) and object (x_j) respectively. Subsequently, a biaffine classifier (Dozat and Manning, 2017) calculates the relationship score between a pair of subject and object words (x_i, x_j).

$$s_i = MLP_2(h_i), \tag{5}$$

$$o_j = MLP_3(h_j), \tag{6}$$

$$y'_{ij} = {s_i}^T U o_j + W[s_i; o_j] + b, \tag{7}$$

where U, W and b are trainable parameters, s_i and o_j denote the subject and object representations of the i-th and j-th word respectively.

MLP Predictor For MLP, its input is the word pair grid representation Q output by the convolutional layer, so the relationship score of each word pair (x_i, x_j) using Qij is calculated as:

$$y''_{ij} = MLP(Q_{ij}), \tag{8}$$

RoWP Predictor For the RoWP prediction layer, its input is the word embedding from the encoding layer. After passing through the RoWP prediction layer, the attention score of each word pair (x_i, x_j) is computed as follows:

$$y'''_{ij} = \frac{(R_m(W_q h_i))^T (R_n(W_k h_j))}{\sqrt{d_k}}, \tag{9}$$

where W_q and W_k are trainable parameters, R_m and R_n is the rotary matrix, used to encode the position information into the key h_i and h_j.
The final relation probabilities y_{ij} for the word pair $(x_i,\ x_j)$ are calculated by combining the scores from the biaffine , MLP predictors and RoWP Predictor:

$$y_{ij} = Softmax(y'_{ij} + y''_{ij} + y'''_{ij}). \tag{10}$$

4.5 Decoding

Entity Decoding In this work, we adopt a straightforward head-to-tail approach to decode the result matrix. Specifically, the horizontal and vertical coordinates corresponding to each position in the lower triangle of the matrix represent the start and end positions of a fragment, respectively. This choice is informed by the characteristics of our Chinese dataset, which contains almost no discontinuous entities.The head-to-tail decoding method not only enhances the interpretability of the model but also aligns better with the dataset's characteristics.

Joint Decoding Since the relationship is directional, we traverse the upper triangle of the matrix to decode the relationship. We use the head-to-head entity relationship filling method, where the horizontal coordinate of each position represents the starting coordinate of the head entity, the vertical coordinate represents the starting coordinate of the tail entity, and the corresponding table value is filled with the relationship ID. During the decoding process, after the model obtains the starting coordinates of the head and tail entities, we traverse the entity head and tail labels in the lower triangular matrix, and decode the corresponding entity head coordinates through this label to obtain the entity tail coordinates, thereby extracting the corresponding entity, and then extracting the entity relationship triple.

Learning Given a document $S = \{w_1, w_2, \ldots, w_N\}$, our training objective is to minimize the negative log-likelihood losses with regard to the corresponding gold labels, formalized as:

$$\mathcal{L} = -\frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{r=1}^{|R|} \hat{y}_{ij}^{r} \log y_{ij}^{r}, \tag{11}$$

where N is the number of words in the document, $\hat{y}_{ij}^{r}$ is the binary vector that denotes the gold relation labels for the word pair (x_i, x_j), and y_{ij} are the predicted probability vector. r indicates the r-th relation of the pre-defined relation set $\mathcal{R}$.

5 Experiments

5.1 Experimental Settings

Parameter settings We use the pretrained weights from BERT-base-Chinese version to initialize the BERT encoder, which consists of 12 layers and a hidden state dimension of 768. For optimization, we utilize the Adam optimizer with an initial learning rate of 1×10^{-3}. We also implement a warm-up rate of 0.1 and an L2 weight decay of 0.1. The maximum document length is set to 512 tokens. To prevent overfitting, we apply a dropout rate of 0.5 after the embedding and convolutional layers, and a dropout rate of 0.33 after the output layer.

Evaluation indicators In order to make a fair comparison with the prior works, we report the standard F1 precision and recall score as our three evaluation metrics. At the same time, we require that the extracted relation triples are considered correct only when both entity and relation are correct.

Baseline We used three strong baseline models for joint extraction of entity relations.

- The CasRel proposed by Wei et al. (2020) models relations as functions that map subjects to objects in a sentence, which naturally handles the overlapping problem.

- GlobalPointer (Su et al., 2022) utilizes position encoding and attention mechanisms to compute scores between each pair of positions, subsequently populating these scores into corresponding cells within a table. With the aid of the global pointer mechanism, the model can concurrently discern entities and extract their relationships, thus effectively extracting intricate relation triplets from textual data.

- OneRel (Shang et al., 2022a) addresses the limitations of traditional pipeline and multi-step approaches by adopting a single module that performs entity and relation extraction in one unified step, treating the task as a fine-grained triple classification problem. The model incorporates a scoring-based classifier and a relation-specific horns tagging strategy to identify and decode entities and relations efficiently, aiming to reduce cascading errors and redundant information.

5.2 Main Results

We compared our proposed model with three baseline models, and the experimental results on our datasets are presented in Table 3. As shown, our model outperforms all baselines and achieves state-of-the-art performance in F1 scores across all datasets. The experimental results further reveal the following observations:

Comparison with CasRel, Our model demonstrates significant performance improvements. This indicates that the one-stage triplet extraction approach adopted in our model effectively mitigates error propagation and exposure bias that are common in multi-module cascade methods. Moreover, the triplet encoding and decoding strategy in our model can simultaneously decode triplets belonging to the same relationship, even when their first and last entities are nested within each other, a challenge for tagging-based methods. Comparison with

model	Intra-sentence Precision(%)	Recall(%)	F1(%)	Cross-sentence Precision(%)	Recall(%)	F1(%)	Document level Precision(%)	Recall(%)	F1(%)
CasRel(Wei et al., 2020)	17.68	32.71	22.95	13.67	54.30	21.84	-	-	-
GlobalPointer(Su et al., 2022)	17.29	33.17	22.73	24.23	45.19	31.54	4.65	12.16	6.73
OneRel(Shang et al., 2022a)	25.20	13.30	17.40	33.33	14.34	20.05	21.27	8.95	12.60
RoWP (Ours)	25.39	26.67	26.01	32.51	37.76	34.94	27.68	35.13	30.96
w/o RoWP	21.09	27.59	23.90	26.29	43.75	32.84	27.53	32.67	29.88

Table 3: Results on joint extraction of document-level relation triples. The baseline models are all joint extraction models. Relation types are divided into intra-sentence relations, cross-sentence relations, and relations of the entire document.

GlobalPointer and OneRel, Our model maintains strong performance. This verifies the effectiveness of our labeling system and the added benefit of RoPE prediction scores. The results suggest that our model not only leverages global context effectively but also benefits from a robust prediction mechanism.

In addition, we also conducted corresponding ablation experiments to analyze the effect of adding RoWP model in different intervals. The experimental results show that the model integrated with RoWP has different degrees of improvement in different intervals. In the sentence interval, the F1 score improves by 2.1%, indicating that the RoWP model enables the model to recognize that the closer the entities in the text are, the more likely they are to be related, which is in line with the intuition of natural language coding that the further away the words in the text are, the less relevant they are. In the cross-sentence interval and document interval, the F1 score increased by 2.1% and 1.1% respectively, and the prediction accuracy also improved, which shows that the RoWP model can effectively enhance the model's sensitivity to position information in the text, thereby improving the accuracy of relationship extraction.

5.3 In-depth Analysis

Q1: How do the compound lengths influence the extraction result?

A1: In this study, we faced a key challenge: accurately extracting longer noun compounds. This task is more complex and challenging than short words in traditional entity recognition. To gain a deeper understanding of the difficulty of extracting compounds of different lengths, we analyzed the performance of the baseline models under different configurations. As shown in Figure 3, we observed that various baseline models showed relatively good recognition results for noun compounds with a length of less than 5 words. We analyzed that this may be because short compounds are not much

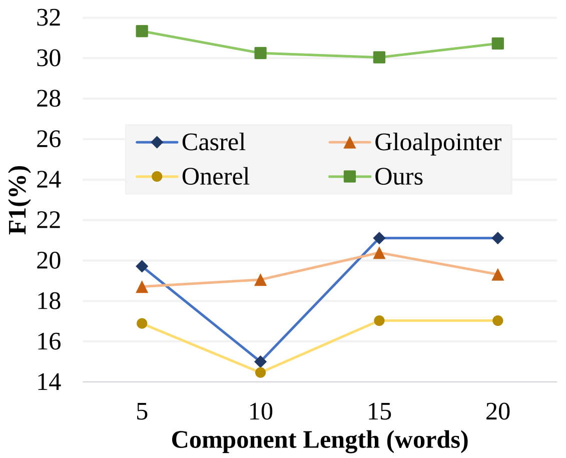

Figure 3: F1 scores of noun compounds of different lengths.

different from ordinary noun entities in structure, allowing the model to effectively extract these short compounds by leveraging its recognition ability for ordinary entities.

However, when the length of the compound increased to 5 to 10 words, we noticed a significant increase in the difficulty of extraction. Compounds in this length range face greater challenges in recognition, probably because as the length of the compound increases, its internal structure becomes more complex, causing the model to face difficulties in determining its boundaries and integrity.

When the length of the compound exceeds 15 words, the extraction difficulty decreases and the results become more stable. We speculate that this may be because long compounds tend to have more obvious characteristics or patterns, allowing the model to recognize them through context clues even when the boundaries are not clear. In addition, long compound words may be more likely to follow specific grammatical rules or appear in specific language environments, and these regularities provide an additional recognition basis for the model.

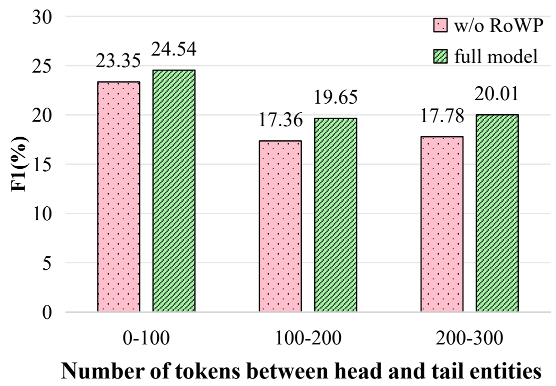

Figure 4: F1 scores under different token distances between head and tail entities.

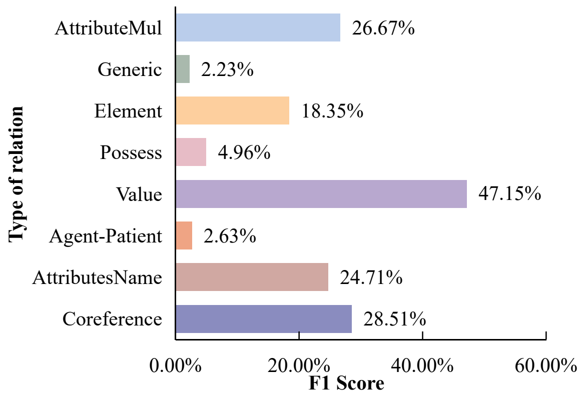

Figure 5: F1 scores predicted for each relationship.

	Pre(%)	Recall(%)	F1(%)
Ours	27.68	35.13	30.96
-Biaffine	27.30(-0.38)	33.15(-1.98)	29.94 (-0.96)
-MLP	32.02(+4.34)	28.39(-6.74)	30.10 (-0.86)
-All DConv	26.29(-1.39)	24.18(-10.59)	25.19 (-5.77)
-DConv(l=1)	31.44(+3.76)	28.39(-6.74)	29.84 (-1.12)
-DConv(l=2)	30.81(+3.13)	29.12(-6.01)	29.94 (-1.02)
-DConv(l=3)	32.75(+5.07)	27.29(-7.84)	29.77 (-1.19)

Table 4: Model ablation studies. DConv(l=1) denotes the convolution with the dilation rate 1.

Q2: How does the RoWP model perform differently at different token distances?
A2: After an in-depth analysis of the performance of the RoWP model at different head-entity-tail-entity distance intervals, the experimental results are shown in Figure 4.

We observe that the F1 score of the model is improved on multiple token intervals. Specifically: when the distance between the head entity and the tail entity is between 0 and 100 tokens, the F1 score of the RoWP model achieves an increase of 1.19%, which indicates that the model is able to more accurately predict the relationship between pairs of closely connected entities when dealing with them. When the distance is further extended to the interval of 100 to 200 tokens, the F1 score of the RoWP model improves by 2.29%, which shows the model's effectiveness in dealing with slightly more complex text structures. In the more distant intervals where the distance between the head entity and the tail entity reaches 200 to 300 tokens, the RoWP model still maintains a 2.23% increase in its F1 score. These findings indicate that the RoWP model is able to effectively capture inter-entity correlations in different distance intervals and maintains high accuracy even in long-distance relationship prediction, further demonstrating the effectiveness of our RoWP model.

Q3: How does the RoWP model perform differently at different token distances?
A3: We verified the prediction effect of each relation separately, using the F1 score as the evaluation metric, and the results are shown in Figure 5. We found that the recognition success rate is higher for Value and Coreference, and lower for Agent-Patient and Generic. Taking Coreference and Agent-Patient as an example, one possible reason is that Coreference recognizes the same entity with more relative overlap and a higher recognition success rate. Whereas Agent-Patient related relationships involve the understanding of the relationship between nouns and verbs, which requires the model to have a stronger contextual understanding, and therefore, the recognition is less effective.

5.4 Ablation study

In addition, we performed the corresponding ablation experiments as shown in Table 4. In the joint prediction layer, we additionally compared the biaffine and MLP prediction layers and found a slight decline in performance. This validates the effectiveness of our proposed RoWP prediction layer. By removing all convolutions, there is a significant decrease in performance, validating the usefulness of multi-granularity dilated convolution. In addition, removing convolutions with different dilation rates also resulted in performance degradation, especially for convolutions with dilation rate 3.

6 Conclusion

In this work, we propose the document-level Noun Compound Relation Extraction (NCRE) task, which provides a new perspective for deeply understanding text content. First, we define the NCRE task, focusing on the complex relations between noun compound phrases. Second, we develop a large-scale, high-quality Chinese NCRE dataset to provide a reliable benchmark for evaluating noun phrase relation extraction methods. We design the "Rotated Position Enhanced Token Pair Relation" (RoWP) model, which effectively improves the accuracy of noun phrase relation extraction by combining multi-granularity convolution, rotated position embedding, biaffine, and MLP techniques. Through extensive experiments on the NCRE dataset, we demonstrate that the RoWP model outperforms existing methods on multiple evaluation metrics.

Limitations

In this paper, our main contributions are the introduction of a new dataset centered on noun compounds and develop a document-level joint extraction model that leverages table completion. Due to the complexity of the dataset and the specificity of the task, we only compare our model with a limited number of baseline models. In addition, our model has only been experimented on our dataset, and we will expand to other similar datasets in the future.

Acknowledgments

This work is supported by the National Natural Science Foundation of China (No. 62176187), and the Social Science Fund of the Ministry of Education (Grant No. 20YJA740062).

References

Richard Breheny. 2008. A new look at the semantics and pragmatics of numerically quantified noun phrases. *J. Semant.*, 25(2):93–139.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: pre-training of deep bidirectional transformers for language understanding. In *Proceedings of NAACL-HLT*, pages 4171–4186.

Kuicai Dong, Aixin Sun, Jung-Jae Kim, and Xiaoli Li. 2023. Shall we trust all relational tuples by open information extraction? A study on speculation detection. *CoRR*, abs/2305.04181.

Timothy Dozat and Christopher D. Manning. 2017. Deep biaffine attention for neural dependency parsing. In *ICLR, Conference Track Proceedings*.

Hao Fei, Meishan Zhang, and Donghong Ji. 2020. Cross-lingual semantic role labeling with high-quality translated training corpus. In *Proceedings of ACL*, pages 7014–7026.

Tao Gui, Yicheng Zou, Qi Zhang, Minlong Peng, Jinlan Fu, Zhongyu Wei, and Xuanjing Huang. 2019. A lexicon-based graph neural network for chinese NER. In *Proceedings of EMNLP*, pages 1040–1050.

Te-En Huang, Tao-Hsing Chang, and Jon-Fan Hu. 2019. A tool to analyze verb phrase and noun phrase relationship in sentences. In *Proceedings of CogSci*, page 3479.

Guillaume Lample, Miguel Ballesteros, Sandeep Subramanian, Kazuya Kawakami, and Chris Dyer. 2016. Neural architectures for named entity recognition. In *NAACL HLT*, pages 260–270.

Bobo Li, Hao Fei, Yafeng Ren, and Donghong Ji. 2020. Nominal compound chain extraction: A new task for semantic-enriched lexical chain. In *NLPCC Proceedings, Part I*, pages 119–131.

Fei Li, Zhichao Lin, Meishan Zhang, and Donghong Ji. 2021a. A span-based model for joint overlapped and discontinuous named entity recognition. In *Proceedings of ACL/IJCNLP, (Volume 1: Long Papers),*, pages 4814–4828.

Jiao Li, Yueping Sun, Robin J. Johnson, Daniela Sciaky, Chih-Hsuan Wei, Robert Leaman, Allan Peter Davis, Carolyn J. Mattingly, Thomas C. Wiegers, and Zhiyong Lu. 2016. Biocreative V CDR task corpus: a resource for chemical disease relation extraction. *Database J. Biol. Databases Curation*, 2016.

Jingye Li, Hao Fei, Jiang Liu, Shengqiong Wu, Meishan Zhang, Chong Teng, Donghong Ji, and Fei Li. 2022. Unified named entity recognition as word-word relation classification. In *AAAI*, pages 10965–10973.

Jingye Li, Kang Xu, Fei Li, Hao Fei, Yafeng Ren, and Donghong Ji. 2021b. MRN: A locally and globally mention-based reasoning network for document-level relation extraction. In *Findings of ACL/IJCNLP*, pages 1359–1370.

Shuangjie Li, Wei He, Yabing Shi, Wenbin Jiang, Haijin Liang, Ye Jiang, Yang Zhang, Yajuan Lyu, and Yong Zhu. 2019. Duie: A large-scale chinese dataset for information extraction. In *NLPCC*, pages 791–800.

Ruibo Liu, Jason Wei, Chenyan Jia, and Soroush Vosoughi. 2021. Modulating language models with emotions. In *Findings of ACL/IJCNLP*, pages 4332–4339.

Laura Mascarell. 2017. Lexical chains meet word embeddings in document-level statistical machine translation. In *Proceedings of DiscoMT@EMNLP*, pages 99–109.

Yevgen Matusevych and Jennifer Culbertson. 2022. Trees neural those: Rnns can learn the hierarchical structure of noun phrases. In *Proceedings of the 44th Annual Meeting of the Cognitive Science Society, CogSci*.

Adila Merabti, Lina Fatima Soualmia, and Stéfan Jacques Darmoni. 2014. Detecting noun phrases in biomedical terminologies: The first step in managing the evolution of knowledge. In *HIS, Proceedings*, pages 109–120.

Peng Qi, Yuhao Zhang, Yuhui Zhang, Jason Bolton, and Christopher D. Manning. 2020. Stanza: A python natural language processing toolkit for many human languages. In *Proceedings of ACL*, pages 101–108.

Xiaofei Qian. 2019. Research on the recognition of the longest inner noun phrase in chinese. *Journal of Zhejiang International Studies University*, pages 59–67.

Feiliang Ren, Longhui Zhang, Xiaofeng Zhao, Shujuan Yin, and Shilei Liu and. 2022. A simple but effective bidirectional framework for relational triple extraction. In *WSDM '22: The Fifteenth ACM International Conference on Web Search and Data Mining*, pages 824–832.

Yuming Shang, Heyan Huang, and Xianling Mao. 2022a. Onerel: Joint entity and relation extraction with one module in one step. In *AAAI*, pages 11285–11293.

Yuming Shang, Heyan Huang, Xin Sun, Wei Wei, and Xian-Ling Mao. 2022b. Relational triple extraction: One step is enough. In *Proceedings of IJCAI*, pages 4360–4366.

Pontus Stenetorp, Goran Topić, Sampo Pyysalo, Tomoko Ohta, Jin-Dong Kim, and Jun'ichi Tsujii. 2011. Bionlp shared task 2011: Supporting resources. In *Proceedings of BioNLP Shared Task 2011 Workshop*, pages 112–120.

Jianlin Su, Yu Lu, Shengfeng Pan, Bo Wen, and Yunfeng Liu. 2021. Roformer: Enhanced transformer with rotary position embedding. *CoRR*, abs/2104.09864.

Jianlin Su, Ahmed Murtadha, Shengfeng Pan, Jing Hou, Jun Sun, Wanwei Huang, Bo Wen, and Yunfeng Liu. 2022. Global pointer: Novel efficient span-based approach for named entity recognition. *CoRR*, abs/2208.03054.

Chengyu Wang, Xiaofeng He, and Aoying Zhou. 2021. Open relation extraction for chinese noun phrases. *IEEE Trans. Knowl. Data Eng.*, 33(6):2693–2708.

Yucheng Wang, Bowen Yu, Yueyang Zhang, Tingwen Liu, Hongsong Zhu, and Limin Sun. 2020. Tplinker: Single-stage joint extraction of entities and relations through token pair linking. In *Proceedings of COLING*, pages 1572–1582.

Zhepei Wei, Jianlin Su, Yue Wang, Yuan Tian, and Yi Chang. 2020. A novel cascade binary tagging framework for relational triple extraction. In *Proceedings of ACL*, pages 1476–1488.

Yuan Yao, Deming Ye, Peng Li, Xu Han, Yankai Lin, Zhenghao Liu, Zhiyuan Liu, Lixin Huang, Jie Zhou, and Maosong Sun. 2019. DocRED: A large-scale document-level relation extraction dataset. In *Proceedings of ACL*.

Hongbin Ye, Ningyu Zhang, Shumin Deng, Mosha Chen, Chuanqi Tan, Fei Huang, and Huajun Chen. 2021. Contrastive triple extraction with generative transformer. In *AAAI*, pages 14257–14265.

Xiangrong Zeng, Daojian Zeng, Shizhu He, Kang Liu, and Jun Zhao. 2018. Extracting relational facts by an end-to-end neural model with copy mechanism. In *Proceedings of ACL , Volume 1: Long Papers*, pages 506–514.

Hengyi Zheng, Rui Wen, Xi Chen, Yifan Yang, Yunyan Zhang, Ziheng Zhang, Ningyu Zhang, Bin Qin, Xu Ming, and Yefeng Zheng. 2021. PRGC: potential relation and global correspondence based joint relational triple extraction. In *Proceedings of ACL/IJCNLP, (Volume 1: Long Papers)*, pages 6225–6235.

Suncong Zheng, Feng Wang, Hongyun Bao, Yuexing Hao, Peng Zhou, and Bo Xu. 2017. Joint extraction of entities and relations based on a novel tagging scheme. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics, ACL, Volume 1: Long Papers*, pages 1227–1236.