

RoGloRE: A Global-Local Adaptive Joint Knowledge Extraction Framework for Chinese Cyber Threat Intelligence

Jipeng Tang¹[0009-0000-1871-7913], Hao Hu¹[0000-0003-4888-6368], Yingchang Jiang²[0009-0005-1632-6413], Yixiao Peng³[0009-0006-6829-991X], Nan Wang²[0009-0008-4997-1939], and Yongwei Wang³[0009-0007-5750-0902] ✉

¹ Information Engineering University, Zhengzhou, 450001, China

² Henan Key Laboratory of Information Security, Zhengzhou, 450001, China

³ State Key Laboratory of Mathematical Engineering and Advanced Computing Zhengzhou, 450001, China
wywyongweip@126.com

Abstract. Automated parsing of Cyber Threat Intelligence (CTI) is crucial for threat attribution and proactive defense. However, Chinese CTI texts are highly unstructured and semantically fragmented, posing dual challenges for existing models. In entity extraction, fragmented tokenization caused by high-entropy entities and complex nested structures leads to ambiguous entity boundaries. In relation extraction, critical attack clues are scattered across paragraphs, preventing traditional attention mechanisms from effectively capturing long-range dependencies and relative spatial structures. To address these limitations, we propose an adaptive global-local joint extraction framework designed for fragmented semantic aggregation in Chinese CTI. Within the entity recognition module, we introduce adaptive rotary position embeddings to correct low-level positional features. This mechanism, combined with a type-decoupled GlobalPointer, resolves recognition conflicts involving long-span entities and nested boundaries. In the relation extraction module, we design a dual-stage attention mechanism to dynamically integrate global cross-paragraph spatial clues with local entity neighborhood features. Additionally, an adaptive decoding strategy aware of class imbalance is implemented to enhance the robustness of the model against sparse long-tail relations. Experimental results on the CDTier dataset indicate that the proposed framework achieves a 9.45% improvement in the F1 score of entity extraction over the best existing baseline, alongside a precision of 93.3% and a recall of 95.5% for relation extraction. The proposed method overcomes the bottleneck of long-range semantic parsing in complex Chinese contexts, demonstrating superior generalization capabilities and practical utility.

Keywords: Cyber threat intelligence, Entity and relation extraction, Rotary position embedding, GlobalPointer.

1 Introduction

As cyber attacks evolve toward Advanced Persistent Threats (APTs), infiltrations of the supply chain, and zero-day exploits, attack behaviors become multi-stage, cross-platform, and highly concealed, posing severe challenges to critical information infrastructures[1]. Cyber Threat Intelligence (CTI) serves as the core medium for conveying Indicators of Compromise (IOCs), mechanisms of vulnerability exploitation, and strategies for defense. The automated parsing of CTI is essential for threat attribution and proactive defense. However, unlike conventional structured corpora, Chinese CTI reports are highly unstructured and semantically fragmented. Comprehensive clues of the kill chain often span long textual sequences, where key entities and core behaviors are scattered across multiple paragraphs, accompanied by numerous nested descriptions and technical terminology. Consequently, traditional models for extraction suffer from limited semantic understanding, an inability to capture long-range dependencies, and ambiguous boundaries of entities when processing Chinese CTI[2].

Recent CTI entity and relation extraction research has progressed significantly. Early rule-based and statistical methods struggle with novel attacks and long texts, driving the adoption of deep learning and pre-trained models. For example, Varghese et al.[3] utilized BERT to extract attacker and tool identifiers from dark web forums. To improve relation extraction, Yuan et al.[4] built a relation-aware attention network enhancing cross-sentence modeling. Additionally, CyberRel[5] integrates BERT and BiGRU for CTI triplet extraction. Regarding paradigms, joint methods alleviate pipeline error propagation via shared representations, whereas pipeline models remain popular in practical systems due to structural clarity and rule compatibility[6]. However, most studies target English. Applying either paradigm to Chinese CTI datasets like CDTier[7] exposes core bottlenecks, as Fig. 1 illustrates.

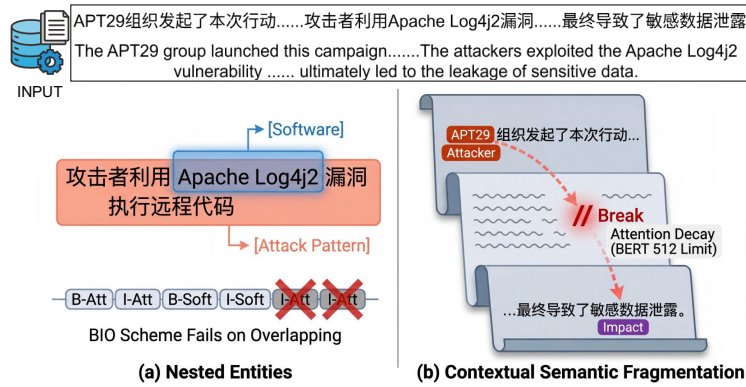


Fig. 1. Challenges of Existing Extraction Methods in Chinese Cyber Threat Intelligence Involving Nested Entities and Contextual Semantic Fragmentation.

First, complex boundaries and entity nesting frequently trigger recognition conflicts. Unlike English, Chinese lacks explicit word delimiters. Combining Chinese characters

with high-entropy attack features, such as hash values, causes boundary ambiguity. Furthermore, frequent long attributive structures in Chinese cause prevalent nesting, often embedding software entities within vulnerability descriptions. Traditional sequence labeling methods like the BIO scheme struggle to parse these overlapping boundaries, sharply reducing recall.

Second, long text truncation causes contextual semantic fragmentation. Chinese heavily emphasizes semantic coherence and frequently utilizes zero anaphora, necessitating deep reliance on long-range contexts. Moreover, attack clues within CTI reports are generally scattered non-linearly. Constrained by the fixed context windows and absolute positional embeddings of BERT-based models, text truncation easily loses implicit cross-paragraph dependencies, ultimately severing long-range entity associations[8].

To address these challenges, we propose RoGloRE (Rotary Global-Local Relation and Entity Extraction), an adaptive global-local relation extraction framework specifically designed to aggregate fragmented semantics in Chinese CTI. It overcomes complex context parsing bottlenecks by integrating dynamic position awareness and a dual attention mechanism. Our main contributions are as follows:

- We construct an entity extraction module based on Rotary Position Embedding (RoPE) and GlobalPointer. It utilizes multi-dimensional relative position modeling to resolve long entity recognition difficulties and nested boundary confusion.
- We design a dual-stage attention mechanism fusing global and local semantics to capture cross-paragraph attack associations. To address sample sparsity, we introduce a class-imbalance-aware adaptive decoding strategy, significantly enhancing model robustness in recognizing long-range dependencies and sparse relations.
- Experiments on the CDTier dataset demonstrate that RoGloRE improves the entity extraction F1 score by 9.45% over the state-of-the-art baseline. Furthermore, relation extraction precision and recall reach 93.3% and 95.5%, respectively, fully verifying the framework's effectiveness in complex contexts.

2 Related Work

This section reviews related work from two dimensions: document-level relation extraction and threat intelligence information extraction, focusing on existing solutions for nested entities and semantic fragmentation.

2.1 General Document-Level Relation Extraction Techniques

Document-level relation extraction aims to identify cross-paragraph entity associations, requiring long-range dependency processing and complex reasoning[9]. Early graph neural network-based methods (e.g., the heterogeneous graph reasoning network by Wang et al.[10]) capture cross-sentence dependencies but rely heavily on static syntactic parsing, making them susceptible to error propagation. Recently, Transformer-based architectures have become mainstream. Zhou et al.[11] introduced adaptive thresholds and localized context pooling. Liu et al.[12] explicitly injected syntactic structures to guide the attention mechanism. Although performing excellently on general datasets,

these methods often struggle to aggregate scattered cross-paragraph clues in Chinese CTI reports due to lacking specific designs for fragmented semantics.

Furthermore, document-level relation extraction generally faces severe class imbalance, where negative samples of irrelevant relations far outnumber positive ones. To this end, Tan et al.[13] combined an adaptive focal loss to effectively mitigate the negative impact of long-tail distributions on model training. Loss function optimization or data augmentation has become a key direction for solving this problem[14]. Our proposed RoGloRE framework draws on these ideas, introducing a class-imbalance-aware adaptive decoding strategy to address relation sparsity in CTI data.

2.2 CTI Relation Extraction Techniques

CTI information extraction aims to identify attack entities and their associations from unstructured reports[15]. Compared to general domains, CTI texts are replete with numerous professional terms (e.g., CVE numbers) and complex nested descriptions. Early regular expression matching methods struggle with unknown threats, whereas deep learning models have become mainstream recently. For instance, the KnowCTI framework by Wang et al.[16] introduces external knowledge to enhance entity representations. Targeting the Chinese context, Zhou et al.[7] constructed the CDTier dataset and highlighted cross-paragraph extraction challenges. Liu et al.[17] and Wang et al.[18] explored end-to-end joint extraction to optimize multi-task performance.

Furthermore, traditional sequence labeling models struggle to effectively address prevalent nested entities within CTI reports. The GlobalPointer model by Su et al.[19] transforms entity recognition into a matrix region classification problem and integrates rotary position embedding[20] to achieve unified overlapping entity parsing. However, most existing methods focus on introducing external knowledge or graph structures, paying little attention to semantic fragmentation caused by high-entropy features (e.g., hash values). They also lack effective aggregation of fragmented cross-paragraph attack clues. This is precisely the core research gap this paper aims to fill.

3 Methodology

3.1 Overall Framework Architecture

As Fig. 2 illustrates, we propose RoGloRE, an adaptive global-local joint extraction framework addressing the non-linear distribution and semantic fragmentation of CTI clues. To avoid traditional pipeline error propagation, it operates end-to-end via three collaborative modules. First, the Representation Learning Module employs a RoPE-enhanced RoBERTa encoder to inject adaptive spatial topology, constructing a long-range context representation. Second, the Entity Recognition Module utilizes an entity-type-decoupled GlobalPointer, applying 2D grid decoding to parallelly parse high-density nested entities and overcome 1D sequence labeling limitations. Finally, the Relation Extraction Module leverages a dual-stage attention mechanism to dynamically fuse global spatial structures and local neighborhood features. Subsequently, a class-imbalance-aware decoder outputs high-confidence triplets, transforming unstructured text into structured threat intelligence graphs.

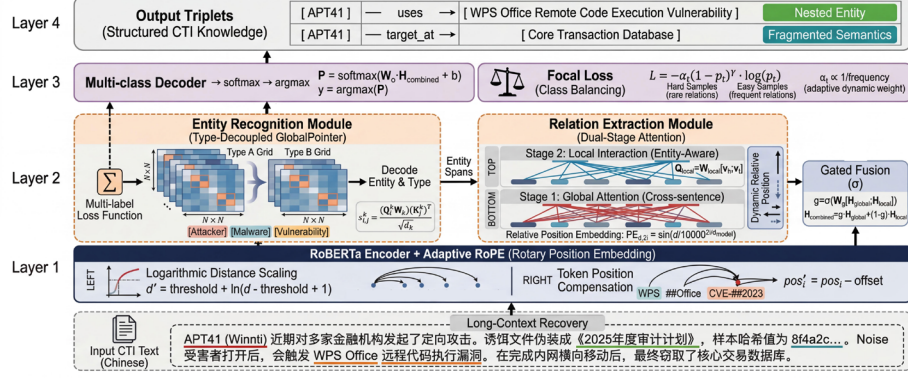


Fig. 2. Overall Architecture of RoGloRE Integrating Rotary Position Embedding and Type-Decoupled GlobalPointer.

3.2 The Recognition of Entities with Dynamic Awareness of Positions

Adaptive Rotary Position Embedding for High-Entropy Entities and Ultra-Long Span Awareness. Original RoPE lacks security-domain topological adaptation, causing token fragmentation (via high-entropy entities) and long-range position noise (via cross-paragraph clues) in CTI texts. To address this, we propose two domain-specific RoPE correction mechanisms: token position compensation and logarithmic distance scaling.

Token Position Compensation for High-Entropy Indicators of Compromise. Conventional tokenizers often split high-entropy strings (e.g., MD5 hashes) into fragmented subwords. Standard RoPE assigns them incrementally increasing rotation angles, causing severe intra-entity position shifts and disrupting semantic integrity. We introduce mean pooling to recalibrate position offsets, anchoring scattered subwords as a unified semantic entity. The formulation is:

$$p[w_i:w_j] = \frac{1}{j-i+1} \sum_{k=i}^j p_k \quad (1)$$

Here, p_k is the original position index for subword k , while w_i and w_j are the entity's start and end indices. This correction term eliminates relative distance variance caused by subword tokenization, ensuring intra-entity vector space continuity and enhancing the model's perception of long-span security entities.

Logarithmic Distance Scaling for Long-Span Attack Chains. Chinese CTI reports often contain detailed long-text forensic descriptions, where key entities (e.g., attack clues, damage conclusions) span multiple paragraphs, causing extremely large inter-entity relative distances. In standard relative position encoding, directly using excessively large original distance indices m in rotation matrix calculations causes violent distant-region position embedding fluctuations and significant noise, leading to long-range attention

weight decay. To address the long-tail inter-entity distance distribution in CTI, we design a logarithmic smoothing mechanism. Specifically, a piecewise continuous spatial mapping function compresses the long-range relative position index into $\tilde{m}$:

$$\tilde{m} = \begin{cases} m, & \text{if } |m| \leq L_{max} \\ \text{sgn}(m) \times (L_{max} + \log(|m| - L_{max} + 1)), & \text{if } |m| > L_{max} \end{cases} \quad (2)$$

We empirically set the threshold $L_{max} = 512$, as over 98% of valid associated entity pairs in real-world CTI corpora (e.g., CDTier) fall within this token distance. This piecewise mechanism balances local and global feature extraction. For relative distance $|m| \leq L_{max}$, the mapping remains linear to accurately capture local syntactic structures. When $|m| > L_{max}$, the logarithmic term suppresses distance value inflation, smoothly compressing long spans into a compact interval. This prevents outlier-induced position space sparsification, enabling the model to stably capture distant fragmented attack clues in long texts and overcome long-range sparsity.

Type-Decoupled GlobalPointer. In CTI texts, the original GlobalPointer's parameter-sharing mechanism causes cross-category feature interference in dense nested entity regions. To address boundary ambiguity, we propose a Type-Decoupled GlobalPointer, which introduces representation decoupling to construct independent feature projection subspaces for each of the K target entity categories. This isolates category-specific boundary features, significantly reducing false positives in complex nested structures. For the k -th category, the confidence score $S_k(i, j)$ for an entity span from position i to j is calculated as:

$$S_k(i, j) = \frac{(R_i q_{i,k})^T (R_j q_{j,k})}{\sqrt{d_k}} \quad (3)$$

Here, R_i and R_j represent the rotation transformation matrices incorporating relative position information at positions i and j , respectively. The terms $q_{i,k}$ and $q_{j,k}$ denote the category-specific feature vectors generated at the start and end positions, i and j , for the k -th entity category. The variable d_k is the dimension of these feature vectors, and $\sqrt{d_k}$ serves as a scaling factor to prevent gradient anomalies caused by excessively large inner product values.

During 2D grid decoding, the vast majority of coordinate points correspond to non-entity spans. This causes a severe class imbalance problem, where the number of positive samples is $O(N)$ while the number of negative samples reaches $O(N^2)$. To address this issue, we introduce a generalized cross-entropy loss function. The loss function $\mathcal{L}_{ner}$ for a specific entity category k is defined as:

$$\mathcal{L}_{ner} = \log \left(1 + \sum_{(i,j) \in \mathcal{P}_k} e^{-S_k(i,j)} \right) + \log \left(1 + \sum_{(i,j) \in \mathcal{N}_k} e^{S_k(i,j)} \right) \quad (4)$$

Here, $\mathcal{P}_k$ and $\mathcal{N}_k$ denote the ground-truth positive and negative span sets, respectively. This loss function globally balances positive and negative gradients and inherently mines hard negatives. Since the negative sample confidence score $S_k(i, j)$ resides

in the exponential term, backpropagation assigns larger penalty weights to high-scoring negative spans (typically easily confused hard negatives). This adaptive optimization effectively drives the model to distinguish subtle entity boundary differences during training.

3.3 Global-Local Adaptive Relation Extraction Network

Global Vision: Relative Space-Aware Attention. To enhance long-range structural context perception, we propose a global attention mechanism incorporating dynamic relative position encoding. For a target entity pair (E_1, E_2) , their relative text distance is $\Delta = pos(E_2) - pos(E_1)$. To capture entity span directionality and features, sine and cosine functions construct the relative position embedding P_Δ :

$$P_\Delta^{(2i)} = \sin\left(\frac{\Delta}{10000^{2i/d}}\right), P_\Delta^{2i+1} = \cos\left(\frac{\Delta}{10000^{2i/d}}\right) \quad (5)$$

Here, i is the dimension index and d is the total encoding dimension. Subsequently, the input hidden state sequence H , the entity type embedding E_{type} , and the generated relative position embedding P_Δ are concatenated along the feature dimension and projected into the Query space via a learnable weight matrix W_g^Q , thereby guiding the global Multi-Head Attention calculation:

$$H_g = \text{MultiHead}(Q_g, K_g, V_g), Q_g = [H; E_{type}; P_\Delta]W_g^Q \quad (6)$$

Here, H_g is the feature of relations at a global level for the output. By explicitly encoding the prior of relative positions into the vector of queries, the weights of attention can be adaptively adjusted based on the span in space between entities. This mechanism effectively alleviates the problem of the dispersion of attention caused by the sparsity of features in texts that are long and improves the accuracy for the capture of relations between entities in contexts that are complex.

Local Vision: Entity Interaction Feature Extraction. During high-dimensional feature aggregation, fine-grained entity-pair semantic details are easily diluted. To address this, we introduce Local Entity-Aware Attention, guiding the model to refocus on the local context of target entities. Specifically, average pooling the hidden states of the head (E_{head}) and tail (E_{tail}) entities yields entity vectors v_h and v_t . These are concatenated and linearly mapped to form a local query vector Q_{loc} . Finally, attention is computed using the base hidden state sequence H as keys/values:

$$Q_{loc} = [v_h; v_t]W_q^L, H_e = \text{Attention}(Q_{loc}, H, H) \quad (7)$$

Here, W_q^L is a matrix of projection that is learnable, and H_e represents the features of context at a local level that are extracted. Following this, we design a mechanism for adaptive gating to dynamically fuse the features at a global level H_g and the features at a local level H_e :

$$h_{pair} = \lambda \cdot H_g + (1 - \lambda) \cdot H_e, \lambda = \sigma(W_\lambda[H_g; H_e]) \quad (8)$$

Here, W_λ is the matrix of weights of the network for gating, and σ is the function of activation of Sigmoid. Through the coefficient for gating λ , this mechanism adaptively allocates the weights for information of topology at a global level and features of neighborhoods at a local level. The resulting representation of relations that is joint h_{pair} directly serves as the input for the module for decoding downstream.

3.4 Adaptive Decoding Aware of Imbalances in Classes

To alleviate the decision boundary shift caused by the extreme class imbalance in CTI datasets, this paper designs an adaptive optimization strategy. First, the joint relation representation h_{pair} is mapped to the relation category space to calculate the probability distribution $P(r|h_{pair})$, and the maximum a posteriori prediction y_{pred} is output:

$$P(r|h_{pair}) = \text{Softmax}(W_{rel} \cdot h_{pair} + b_{rel}) \quad (9)$$

$$y_{pred} = \text{argmax}(P(r|h_{pair})) \quad (10)$$

Here, W_{rel} and b_{rel} are the parameters of the projection layer. Furthermore, to prevent the gradients of sparse categories in the long-tail distribution from being overwhelmed by majority categories, this paper introduces a dynamic re-weighting loss function based on Focal Loss:

$$\mathcal{L}_{re} = -\sum_{t=1}^K \mathbb{I}(y = t) \cdot \alpha_t (1 - p_t)^\gamma \log(p_t) \quad (11)$$

Here, K is the total relation count (including NA), y is the ground-truth label, and p_t is the predicted probability for category t . This function optimizes via two mechanisms: at the sample level, a decay factor $(1 - p_t)^\gamma$ down-weights easy, high-confidence samples to focus on hard ones; at the category level, a balancing parameter α_t provides an adaptive weight inversely proportional to training frequency. This dual dynamic weighting effectively corrects extreme class imbalance, improving long-tail relation recall.

4 Experiments

To verify the effectiveness of the framework of RoGloRE in scenarios of CTI in Chinese, experiments were conducted based on the dataset of CDTier. CDTier is the first dataset of CTI in Chinese that is publicly annotated, and multiple works of research that are recent have conducted experiments based on this dataset[7, 23, 24]. CDTier contains:

- Subset for the extraction of entities: 100 reports, 3,744 sentences of threats, and 4,259 entities that are annotated. (attacker, tool, industry, region, etc.)
- Subset for the extraction of relations: 2,598 sentences and 2,562 triplets, covering 11 types of relations such as uses, target_at, etc.

The dataset features characteristics that are complex such as terminology of mixed Chinese and English and entities with long spans, making it highly challenging. The

partitioning of data is performed using sampling that is stratified at a ratio of 8:1:1 (training/validation/testing).

Table 1. Comparison of RoGloRE with Baseline Models on Entity Recognition Performance.

Model	P	R	F1
ConsistNER[21]	60.28%	63.00%	61.61%
LLM-TIKG[22]	83.78%	50.69%	63.17%
BERT+LSTM[7]	51.11%	61.50%	55.82%
BERT+BiLSTM[7]	58.83%	75.79%	66.24%
BERT+BiLSTM+CRF[7]	69.25%	75.79%	72.37%
BERT+BiLSTM+GRU+CRF[7]	69.25%	79.18%	77.95%
RoGloRE(Ours)	86.67%	88.15%	87.40%

4.1 Setup for Experiments and Metrics

All models are implemented based on the framework of PyTorch in an environment of GPU of NVIDIA RTX 4090. The optimizer of AdamW is employed with a rate of learning that is initial of $3e-5$ and a size of batches of 8. A strategy for stopping early is controlled based on the score of F1 on the set for validation to prevent overfitting. The metrics for evaluation are Precision (P), Recall (R), and the score of F1.

4.2 Recognition of Entities

Table 2. Performance Comparison on Long Threat Entity Recognition.

Model	P	R	F1
BERT+BiLSTM+CRF	72.21%	61.72%	66.55%
BERT+BiLSTM+GRU+CRF	74.62%	65.56%	69.80%
RoGloRE(Ours)	86.24%	85.47%	85.85%

To evaluate RoGloRE in entity recognition, we compare it with classic BERT-centric sequence labeling networks, ConsistNER[21] which emphasizes prediction consistency, and LLM-TIKG[22] which integrates large language model features. As Table 1. shows, RoGloRE's F1-score outperforms the best baseline by 9.45%. Specifically, although ConsistNER and LLM-TIKG introduce novel feature fusion paradigms, when processing highly specialized, long-span, and structurally complex CTI entities in the CDTier dataset, they are highly prone to entity truncation or localization deviation due to the lack of an explicit fine-grained span boundary modeling mechanism. In contrast, RoGloRE utilizes rotary position encoding to enhance long-range dependency perception and synergizes with global GlobalPointer decoding to accurately delineate complex boundaries. This achieves a high precision-recall balance in long and cross-sentence scenarios, significantly boosting overall robustness.

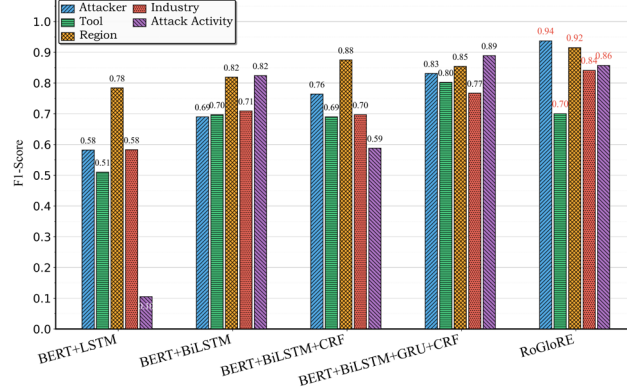


Fig. 3. Performance Comparison of RoGloRE and Various Baseline Models on Entity Recognition for Five Categories of Chinese Threat Intelligence.

Figure 3 shows RoGloRE achieves excellent recognition across five threat entity categories. Significant improvements in Attacker and Attack Activity verify its advantages in handling nested, complex, and long-distance entities, while Industry and Region remain stable. Although Tool metrics slightly trail the optimal baseline, likely because short spans and sparse context limit RoPE, RoGloRE demonstrates strong adaptability and generalization across diverse entities.

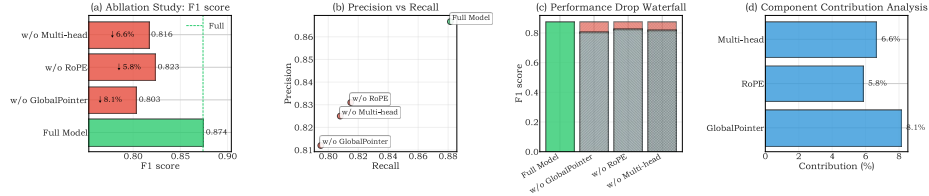


Fig. 4. Ablation Study of the Entity Recognition Module Incorporating RoPE and Type-Decoupled GlobalPointer.

To investigate the contributions of the base encoder and entity recognition modules, we conducted ablation experiments (Fig. 4). The category-decoupled global pointer network plays a core role; its removal drops the F1-score by 8.1%, verifying its effectiveness in handling overlapping, nested, and complex entities within the 2D grid decoding space. Furthermore, stripping rotary position encoding causes a 5.8% performance drop, confirming that explicit relative position information enhances long-distance entity dependency perception. Removing multi-head self-attention from the base encoder yields a 6.6% decrease, indicating that global context semantic capture by the base language model is an indispensable prerequisite. Ultimately, the entity recognition performance improvement primarily stems from the effective synergy between dynamic base position perception and upper-level global decoding strategies.

Case 1	Input:	在行动期间, APT29 针对哈萨克斯坦共和国外交部的鱼叉式网络钓鱼攻击被网络安全分析员检测。 During the operation, the spear-phishing attack launched by APT29 against the Ministry of Foreign Affairs of the Republic of Kazakhstan was detected by cybersecurity analysts.	
	Baseline:	APT29 针对哈萨克斯坦共和国外交部 哈萨克斯坦共和国	— attack-type (✗) — location-GPE (✓)
	RoGloRE(Ours):	APT29 针对哈萨克斯坦共和国外交部的鱼叉式网络钓鱼攻击 哈萨克斯坦共和国	— attack-event (✓) — location-GPE (✓)
Case 2	Input:	DarkHydrus 凭证窃取行动于2018年首次被发现, 随后被确认与通过Excel宏投放 RogueRobin 木马程序 有关。 The DarkHydrus credential-stealing campaign was first uncovered in 2018 and was subsequently confirmed to be associated with the delivery of the RogueRobin Trojan via Excel macros.	
	Baseline:	DarkHydrus凭证窃取行动 — attack-type (✓) 通过Excel宏投放 Rogue — malware (✗)	
	RoGloRE(Ours):	DarkHydrus凭证窃取行动 — attack-type (✓) 通过Excel宏投放RogueRobin木马程序 — malware (✓)	

Fig. 5. Qualitative Comparison of RoGloRE and Baseline Models in Correcting Long Entity Boundary Truncation Errors and Preserving Semantic Integrity.

To further evaluate long entity recognition, we utilized a large language model to generate variant sentences with complex, long threat entities, strictly preserving CTI attack scenarios and semantics. To ensure evaluation objectivity and prevent data leakage, these manually reviewed samples formed an independent long-entity challenge test set. As Table 2 shows, RoGloRE’s F1-score exceeds the optimal baseline by 16%, with substantial simultaneous improvements in precision and recall. This performance leap primarily stems from two-stage attention and position compensation mechanisms, enabling stable capture of deep, cross-sentence, and long-range entity features. This demonstrates exceptional generalization and robustness in complex CTI contexts. Furthermore (Fig. 5), RoGloRE more precisely delineates long threat boundaries and entity types than baselines, effectively reducing false positive and negative rates.

4.3 Relation Extraction

To verify RoGloRE’s effectiveness in cross-sentence dependency handling and long-distance relation modeling in CTI texts, we conducted a comprehensive baseline comparison (Fig. 6). Traditional deep learning models suffer from finite context windows or static syntax parsing error propagation, yielding generally low accuracy. Pre-trained BERT increases accuracy to 89.4% but remains insufficient for long documents. ATLOP[11], introducing document-level features, and KnowPrompt[25], integrating prompt learning, further improve recognition performance. However, lacking explicit relative position compensation for cross-sentence entities, they remain prone to local semantic decay and feature loss when processing highly fragmented, extremely large-span CTI threat relations. In contrast, RoGloRE leverages rotary position encoding to effectively strengthen long-distance dependency perception. Although KnowPrompt performs adequately on the single accuracy metric, RoGloRE achieves 93.3% precision and 95.5% recall, comprehensively surpassing cutting-edge baselines and fully verifying its outstanding robustness in complex long-distance relation reasoning.

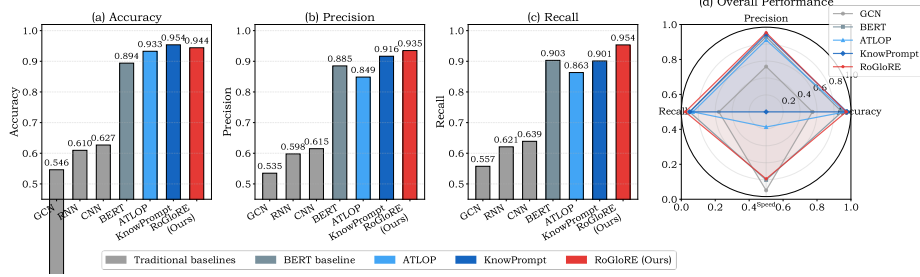


Fig. 6. Comparative performance analysis of RoGloRE against existing baseline models on relation extraction tasks.

Regarding the efficiency-performance balance, although RoGloRE infers slightly slower than BERT due to the additional computational overhead of the global pointer and dual-stage attention mechanisms, it trades this marginal time sacrifice for comprehensive precision and recall advantages. As the radar chart (Fig. 6(d)) illustrates, RoGloRE achieves a superior efficiency-performance balance, fully satisfying the real-time requirements of most CTI applications.

To analyze RoGloRE's performance improvement in relation extraction, we conducted an ablation study on its core components (Fig. 7). The dual-stage attention mechanism is most critical; its removal drops accuracy by 7.3%, confirming that global and local semantic fusion is essential for integrating fragmented clues across different CTI report paragraphs. Furthermore, removing the adaptive optimization strategy reduces accuracy by 5.5%. This verifies that for extreme long-tail CTI data distributions, dynamically adjusting loss weights via Focal Loss and category frequencies corrects majority-class decision boundary bias, enhancing sparse attack relation identification. Finally, removing both core modules causes substantial performance degradation, demonstrating their deep synergy—rather than linear superposition—in ensuring model robustness.

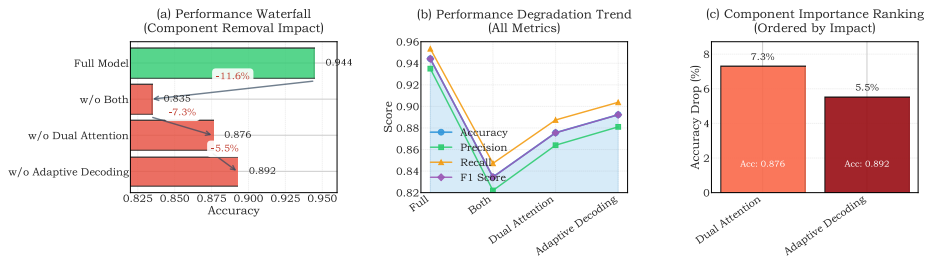


Fig. 7. Ablation Study on the Relation Extraction Module: Evaluating the Contribution of Dual Attention and Adaptive Decoding.

4.4 Analysis of Scalability and Robustness

To evaluate the reliability and adaptability of the model in practical application environments, we evaluate RoGloRE from the dimensions of data efficiency, computational resource overhead, and cross-domain generalization capability.

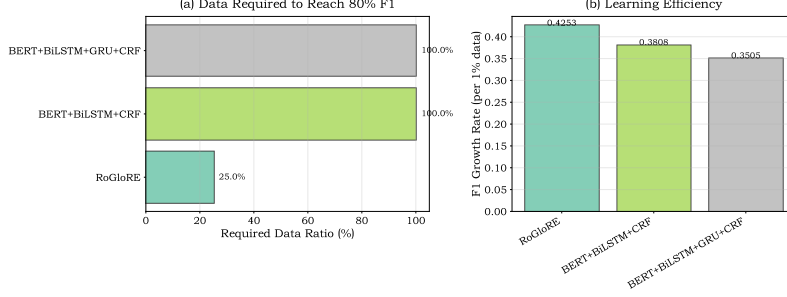


Fig. 8. Comparison of Data Efficiency and Learning Performance across Different Model Architectures in Low-Resource Scenarios.

To address the scarcity of annotated CTI data, we evaluate the model in low-resource environments. Figure 8(a) shows that RoGloRE requires only 25% of the training data to achieve an F1-score of 80%, whereas the BERT+BiLSTM+CRF baseline requires the full dataset. Furthermore, the learning efficiency of RoGloRE reaches 0.4253 (Fig. 8(b)), which outperforms the score of 0.3505 of the baseline. This rapid capture of key features from limited samples demonstrates strong cold-start capabilities, effectively reducing the reliance on massive expert annotations during deployment.

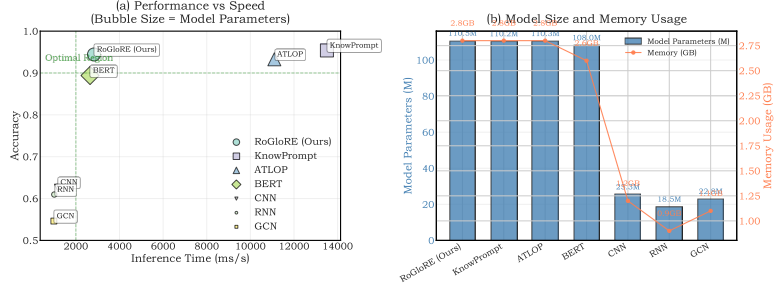


Fig. 9. Comprehensive Comparison of Model Accuracy, Inference Speed, and Computational Cost.

Regarding computational resources and deployment feasibility, fig. 9(a) demonstrates the accuracy-inference speed trade-off across models. RoGloRE occupies the optimal upper-left region, maintaining excellent accuracy and highly competitive inference latency. Conversely, despite high accuracy, ATLOP and KnowPrompt exhibit sharply increasing inference times (>10,000 ms/sentence), restricting practical application efficiency. Furthermore, resource overhead analysis (Fig. 9(b)) shows RoGloRE's

parameter count and GPU memory consumption match mainstream pre-trained baselines. This proves the model achieves a significant performance leap through high architectural efficiency rather than parameter expansion. Consequently, RoGloRE exhibits excellent deployability on mainstream GPU hardware, satisfying strict computational constraints for real-time threat intelligence analysis.

To verify RoGloRE avoids overfitting CTI data, we evaluated its entity extraction submodule on four general Chinese NER datasets: OntoNotes 4.0, MSRA, Resume, and Weibo. Table 3 shows robust performance, marginally inferior only to the general-domain state-of-the-art W^2NER [26]. This is because W^2NER is optimized for general scenarios (e.g., news, social media) with balanced entity distributions. Conversely, RoGloRE is customized for the CTI domain; its category-decoupled global pointer network targets complex nested and overlapping entities. Transferring this decoding mechanism without fine-tuning to general NER datasets featuring short, flat entities introduces structural redundancy, causing slight performance degradation.

Nevertheless, RoGloRE’s entity recognition module achieves F1 scores of 82.22% and 95.23% on OntoNotes 4.0 and MSRA, respectively, remaining on par with advanced mainstream methods. This demonstrates its encoder and global pointer architecture successfully capture fundamental natural language semantics. By trading marginal general-scenario performance for substantial CTI domain advantages, the module exhibits excellent cross-domain generalization and robustness, establishing a solid foundation for complex threat relation extraction tasks.

Table 3. Cross-Dataset Performance Comparison.

	OntoNotes4.0			MSRA			Resume			Weibo		
	P	R	F1	P	R	F1	P	R	F1	P	R	F1
Zhang et al.[27]	76.35	71.56	73.88	93.57	92.79	93.18	94.81	94.11	94.46	53.04	62.25	58.79
Yan et al.[28]	-	-	72.43	-	-	92.74	-	-	95.00	-	-	58.17
Gui et al.[29]	76.40	72.60	74.45	94.50	92.93	93.71	95.37	94.84	95.11	57.14	66.67	59.92
Li et al.[30]	-	-	81.82	-	-	96.09	-	-	95.86	-	-	68.55
Ma et al.[31]	83.41	82.21	82.81	95.75	95.10	95.42	96.08	96.13	96.11	70.94	67.02	70.50
W^2NER [26]	82.31	83.36	83.08	96.12	96.08	96.10	96.96	96.35	96.65	70.84	73.87	72.32
RoGloRE(ours)	81.56	82.89	82.22	95.12	95.34	95.23	95.78	96.23	96.00	68.34	69.78	69.05

5 Conclusion

To address semantic fragmentation and complex nested entities in Chinese CTI texts, we propose RoGloRE, an adaptive global-local relation extraction framework. For entity extraction, adaptive rotary position embeddings and an entity-type-decoupled GlobalPointer overcome recognition difficulties caused by word segmentation fragmentation and complex long-span nesting. For relation extraction, a dual-stage attention mechanism dynamically fuses global spatial structures and local neighborhood features. Combined with a class-imbalance-aware decoding strategy, this significantly enhances the capture of sparse, long-distance relations. Extensive experiments on the CDTier dataset demonstrate our method surpasses mainstream baselines in overall performance, exhibiting excellent robustness and practical engineering value regarding cross-domain generalization, data efficiency, and inference latency. Future work will

explore few-shot learning via large pre-trained language models to reduce reliance on high-quality annotations, extending RoGloRE to automated multilingual cybersecurity knowledge graph construction.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Ahmed, K., Khurshid, S.K., Hina, S.: Cyberentrel: Joint extraction of cyber entities and relations using deep learning. *Computers & Security* **136**, 103579 (2024)
2. Chen, X., Zhang, N., Xie, X., Deng, S., Yao, Y., Tan, C., Huang, F., Si, L., Chen, H.: Knowprompt: Knowledge-aware prompt-tuning with synergistic optimization for relation extraction. In: *Proceedings of the ACM Web Conference 2022*, pp. 2778–2788. ACM, New York (2022)
3. Gui, T., Zou, Y., Zhang, Q., Peng, M., Fu, J., Wei, Z., Huang, X.J.: A lexicon-based graph neural network for chinese ner. In: *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pp. 1040–1050. ACL, Stroudsburg (2019)
4. Guo, Y., Liu, Z., Huang, C., Liu, J., Jing, W., Wang, Z., Wang, Y.: Cyberrel: Joint entity and relation extraction for cybersecurity concepts. In: Gao, D., Li, Q., Guan, X., Liao, X. (eds.) *Information and Communications Security – ICICS 2021, LNCS*, vol. 12918, pp. 447–463. Springer, Cham (2021)
5. Han, X., Gao, T., Lin, Y., Peng, H., Yang, Y., Xiao, C., Liu, Z., Li, P., Zhou, J., Sun, M.: More data, more relations, more context and more openness: A review and outlook for relation extraction. In: *Proceedings of the 1st Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the 10th International Joint Conference on Natural Language Processing*, pp. 745–758. ACL, Stroudsburg (2020)
6. Hu, Y., Zou, F., Han, J., Sun, X., Wang, Y.: Llm-tikg: Threat intelligence knowledge graph construction utilizing large language model. *Computers & Security* **145**, 103999 (2024)
7. Jo, H., Lee, Y., Shin, S.: Vulcan: Automatic extraction and analysis of cyber threat intelligence from unstructured text. *Computers & Security* **120** (14), 102763 (2022)
8. Kim, H., Kim, J.E., Kim, H.: Exploring nested named entity recognition with large language models: Methods, challenges, and insights. In: *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, pp. 8653–8670. ACL, Stroudsburg (2024)
9. Li, J., Fei, H., Liu, J., Wu, S., Zhang, M., Teng, C., Ji, D., Li, F.: Unified named entity recognition as word-word relation classification. In: *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 36, pp. 10965–10973. AAAI Press, Palo Alto (2022)
10. Li, X., Yan, H., Qiu, X., Huang, X.J.: Flat: Chinese ner using flat-lattice transformer. In: *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pp. 6836–6842. ACL, Online (2020)
11. Li, Y., Guo, Y., Fang, C., Liu, Y., Chen, Q.: A novel threat intelligence information extraction system combining multiple models. *Security and Communication Networks* **2022** (1), 8477260 (2022)
12. Liu, W., Zhou, L., Zeng, D., Qu, H.: Document-level relation extraction with structure enhanced transformer encoder. In: *2022 International Joint Conference on Neural Networks (IJCNN)*, pp. 1–8. IEEE, Piscataway (2022)

13. Liu, Y., Han, X., Zuo, W., Lv, H., Guo, J.: Cti-je: a joint extraction framework of entities and relations in unstructured cyber threat intelligence. In: 2024 27th International Conference on Computer Supported Cooperative Work in Design (CSCWD), pp. 2728–2733. IEEE, Piscataway (2024)
14. Ma, B., Zhou, Y., Wu, S., Wang, Z., Xiao, Y., Cui, Y., Liu, Y., Tian, Z.: Apt-kg2qa: An intelligent fine-tuning strategy for large language models utilizing the apt knowledge graph. *IEEE Internet of Things Journal* **12***(18), 38493–38508 (2025)
15. Ma, R., Peng, M., Zhang, Q., Wei, Z., Huang, X.J.: Simplify the usage of lexicon in chinese ner. In: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, pp. 5951–5960. ACL, Online (2020)
16. Siracusano, G., Sanvito, D., Gonzalez, R., Srinivasan, M., Kamatchi, S., Takahashi, W., Kawakita, M., Kakumaru, T., Bifulco, R.: Time for action: Automated analysis of cyber threat intelligence in the wild. *arXiv preprint arXiv:2307.10214* (2023)
17. Su, J., Ahmed, M., Lu, Y., Pan, S., Bo, W., Liu, Y.: Roformer: Enhanced transformer with rotary position embedding. *Neurocomputing* **568***, 127063 (2024)
18. Su, J., Murtadha, A., Pan, S., Hou, J., Sun, J., Huang, W., Wen, B., Liu, Y.: Global pointer: Novel efficient span-based approach for named entity recognition. *arXiv preprint arXiv:2208.03054* (2022)
19. Tan, Q., He, R., Bing, L., Ng, H.T.: Document-level relation extraction with adaptive focal loss and knowledge distillation. In: Findings of the Association for Computational Linguistics: ACL 2022, pp. 1672–1681. ACL, Stroudsburg (2022)
20. Varghese, V., Mahalakshmi, S., Kb, S.: Extraction of actionable threat intelligence from dark web data. In: 2023 International Conference on Control, Communication and Computing (ICCC), pp. 1–5. IEEE, Piscataway (2023)
21. Wang, D., Hu, W., Cao, E., Sun, W.: Global-to-local neural networks for document-level relation extraction. In: Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP), pp. 3711–3721. ACL, Stroudsburg (2020)
22. Wang, G., Liu, P., Huang, J., Bin, H., Wang, X., Zhu, H.: Knowcti: Knowledge-based cyber threat intelligence entity and relation extraction. *Computers & Security* **141***, 103824 (2024)
23. Wang, H., Zhang, S., Wang, Y., Sun, Y., Liu, Q.: Joint extraction of cyber threat intelligence entity relationships based on a parallel ensemble prediction model. *Sensors* **25***(16), 5193 (2025)
24. Wang, R., Liu, Y., Sun, Y., Su, S., Fang, B., Tian, Z.: H4: A software-defined deception defense system in safeguard defense mode. *IEEE Transactions on Dependable and Secure Computing* (2025)
25. Wu, C., Ke, W., Wang, M., Luo, Z., Li, G., Chen, W.: Consistner: towards instructive ner demonstrations for llms with the consistency of ontology and context. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 38, pp. 19234–19242. AAAI Press, Palo Alto (2024)
26. Yan, H., Deng, B., Li, X., Qiu, X.: Tener: adapting transformer encoder for named entity recognition. *arXiv preprint arXiv:1911.04474* (2019)
27. Yuan, Y., Zhou, X., Pan, S., Zhu, Q., Song, Z., Guo, L.: A relation-specific attention network for joint entity and relation extraction. In: Proceedings of the 29th International Joint Conference on Artificial Intelligence (IJCAI), pp. 4054–4060. IJCAI, Yokohama (2020)
28. Zhang, Y., Yang, J.: Chinese ner using lattice lstm. In: Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pp. 1554–1564. ACL, Stroudsburg (2018)

29. Zhao, Y., Deng, Y., Yang, L., Wang, M., Zhang, R., Cheng, H., Lam, W., Shen, Y., Xu, R.: A comprehensive survey on relation extraction: Recent advances and new frontiers. *ACM Computing Surveys* **56**(11), 1–39 (2024)
30. Zhou, W., Huang, K., Ma, T., Huang, J.: Document-level relation extraction with adaptive thresholding and localized context pooling. In: *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 35, pp. 14612–14620. AAAI Press, Palo Alto (2021)
31. Zhou, Y., Ren, Y., Yi, M., Xiao, Y., Tan, Z., Moustafa, N., Tian, Z.: Ctdier: A chinese dataset of threat intelligence entity relationships. *IEEE Transactions on Sustainable Computing* **8**(4), 627–638 (2023)