International Journal of Digital Crime and Forensics
Volume 16 • Issue 1 • January-December 2024

Named Entity Recognition Method of Chinese Legal Documents Based on Parallel Instance Query Network

Rui Lu
Liaoning Police College, China

Linying Li
https://orcid.org/0009-0003-5100-444X
Dalian University of Foreign Languages, China

ABSTRACT

Legal Named Entity Recognition (NER) is crucial in intelligent judiciary systems, focusing on identifying case-specific entities in legal texts. It helps convert unstructured legal documents into structured data, improving e-discovery efficiency. However, challenges arise from insufficient understanding of legal terminology, leading to errors in identifying long and nested entity boundaries. To address this, a Legal NER method based on a parallel instance query network is proposed. This method uses learnable instance queries to extract entities in parallel, with a BERT+BiLSTM+attention structure to encode context and query information. Entity prediction is performed using a pointer network to identify span boundaries and entity types. A linear label assignment mechanism aligns legal entities with queries for more accurate labeling. Experimental results show that the model outperforms existing methods, and further validation through ablation experiments and case studies supports its effectiveness, offering valuable insights for advancing legal NER research.

KEYWORDS

Information Extraction, Named Entity Recognition, Parallel Instance Query Network, Legal Documents, Linear Label Assignment, Knowledge Extraction, Entity Nesting, E-discovery Automation

INTRODUCTION

Named entity recognition (NER) is a fundamental task in natural language processing (NLP), aimed at identifying general entities such as person, time, and location from text. In the judicial domain, legal named entity recognition (LNER) is a specialized task that focuses on case-specific entities closely related to legal proceedings. These entities typically consist of terms or phrases that hold significant meaning within the legal context, such as “suspect” and “victim,” which are subtypes of the “person” entity commonly identified in general NER. With advancements in NLP, extracting named entities from vast and unstructured legal texts has become a critical task for constructing legal knowledge graphs and developing intelligent justice systems (Correia et al., 2021; Guo et al., 2021). Additionally, LNER plays a foundational role in downstream tasks, such as judicial summarization, question answering, and case recommendation. However, the presence of specialized terminology, unclear entity boundaries, long entities, and nested entities in legal texts poses significant challenges. Most existing NER models struggle to effectively address these issues, resulting in suboptimal

DOI: 10.4018/IJDCF.367470

performance in legal entity recognition (Shen et al., 2022). Chinese legal texts, in particular, pose additional challenges due to the frequent occurrence of multi-word phrases or lengthy noun entities. This complexity complicates word segmentation, as Chinese lacks spaces between words, unlike English. Moreover, legal texts often contain nested entities, where one entity is embedded within another. For instance, in the phrase "a gold ring from the victim Lin's home," the entity "a gold ring" (stolen item) is nested within "the victim Lin's home" (location), which is further nested within "Lin" (victim). General NER methods may correctly identify "a gold ring" as a stolen item but may fail to recognize nested entities like "victim" or "location," resulting in incomplete recognition and a limited understanding of the relationships between entities such as location, person, and stolen item.

LNER faces the following main challenges:

- Due to the specificity of the legal domain, legal documents contain long entities and nested entities. Long entities are composed of multiple nouns or phrases, which complicates their segmentation. Nested entities, on the other hand, have multi-layer structures where entity boundaries intertwine and overlap, making their recognition particularly challenging.
- General NER methods primarily predict entity labels based on context, often overlooking the semantic relationships between the textual context and entity label types. While the machine reading comprehension (MRC) approach addresses some of these limitations, it is inefficient as it can only identify one entity type per inference. Furthermore, the quality of manually constructed queries in this approach can vary significantly, further affecting its accuracy.

To comprehensively address the issues mentioned above, this paper introduces a LNER method designed specifically for recognizing entities in Chinese legal documents. The method is based on the parallel instance query network-NER (PIQN-NER), which uses trainable queries to replace the fixed queries in MRC and extract entities simultaneously. Unlike previous methods, these queries can be constructed in advance without relying on external knowledge. A linear label assignment mechanism is employed to align gold entities with the instance queries. First, PIQN-NER fine-tunes bidirectional encoder representation from transformers (BERT) to encode character sequences. Then, a bidirectional long short-term memory (BiLSTM) combined with an attention mechanism is applied to assign different attention weights to both the context and instance queries, which improves the model's ability to correctly determine entity boundaries. Finally, the entity prediction component leverages a pointer network to capture both the span boundaries and types of legal entities. Experiments demonstrate that the proposed method outperforms related methods when applied to legal datasets.

The main contributions of this work are as follows:

- A PIQN-NER model is introduced for legal documents, where trainable instance queries replace MRC-specific queries, enabling the simultaneous identification of all entities. In this model, the BERT+BiLSTM+attention component encodes context and query instances, while the entity prediction component captures the boundaries and types of legal entities based on the pointer network.
- A novel loss function is designed based on the label assignment mechanism, which assigns entities to instance queries. This approach allows instance queries to learn richer semantics from legal documents without relying on external knowledge.
- Finally, the experimental results applied to legal documents and their practical implications are discussed, leading to the conclusions and limitations of the study.

RELATED WORK

With the rapid development of deep neural networks, various approaches have been introduced to address NER as a classification problem aimed at recognizing text spans related to specific predefined entity categories. However, most existing NER models are trained on general domain corpora, such as news articles, which limits their applicability to specialized domains or tasks, such as legal domains. Recently, some researchers have proposed models trained on specialized corpora to overcome these limitations.

NER

Flat NER has been extensively studied and is typically treated as a sequence labeling problem. However, nested entities, which carry multi-granular semantic meaning, are crucial in many real-world applications (Dang et al., 2023; Tan et al., 2021). The current mainstream methods for nested named entity recognition include four major approaches: hierarchical methods, sequence-to-sequence methods, hypergraph-based methods, and span-based methods.

The hierarchical method (Ju et al., 2018; Shibuya et al., 2020; Wang et al., 2020; Wang et al., 2021) achieves multi-level nested entity recognition by embedding multi-layer entity information within a hierarchical structure. Each layer is trained to recognize entities specific to its group, according to their depth in the dataset. However, hierarchical methods struggle to effectively process complex structures with multiple nested layers, leading to inaccurate recognition in certain cases. Additionally, this layered model architecture introduces the issue of error propagation, as the recognition results of one layer serve as input for the next. Consequently, errors in one layer can propagate and accumulate, negatively affecting the performance of the entire model. Moreover, the use of multiple stacked flat NER layers increases model complexity, requiring more computational resources and time for training and inference.

The sequence-to-sequence method (Straková et al., 2019; Yan et al., 2021) treats nested NER as a sequence generation task and utilizes a transformer-based model to predict entity labels for each position in the sequence. These models, especially those with attention mechanisms, show a superior ability to capture contextual information, which improves entity recognition accuracy. However, sequence-to-sequence models face challenges due to higher computational complexity, particularly when processing long sequences. Additionally, inconsistencies in sequence order can lead to unreasonable training losses, further diminishing model performance.

The hypergraph-based approach (Lu et al., 2015; Katiyar et al., 2018; Wang et al., 2018) identifies nested entities by encoding all possible entity mentions and their nested relationships within a hypergraph. While this approach offers a robust framework for modeling nested and overlapping entities, it suffers from high time complexity. Moreover, its performance is highly dependent on the domain and the design of unambiguous hypergraphs. The neural segmental hypergraph representation, although powerful, also faces challenges in training and optimization, particularly in more complex scenarios. Furthermore, the effectiveness of this method hinges on the quality of word-level and cross-level features extracted from neural networks. Inadequate or incomplete feature extraction can undermine the model's performance, affecting its overall effectiveness.

The span-based method (Sohrab et al., 2020; Tan et al., 2020; Yu et al., 2020) directly enumerates all possible entity spans in a sequence and predicts entity labels for each span based on semantic information, aiming to recognize nested named entities. The boundary detection model in this approach can generate high-quality candidate spans, thereby enhancing overall performance. However, this method relies heavily on external resources, which limit its adaptability across different domains or languages. Furthermore, while the span-based method demonstrates efficacy in recognizing nested entities, it struggles with exceptionally long entities or complex nested structures. To address these challenges, several studies have proposed improvements, such as optimizing span patterns and refining attention mechanisms (Shen et al., 2021; Su et al., 2023; Zhang et al., 2022; Zhang et al., 2023).

Shen et al. (2021) approached NER as a joint task of boundary regression and span classification, developing span proposals to reduce the number of candidate spans and improve the quality of span recognition. Su et al. (2022) introduced a global pointer method based on a multiplicative attention mechanism, which has several advantages, including providing a global perspective and addressing label imbalance issues in NER tasks. However, this method also presents challenges, such as dependence on the number of parameters and the quality of the training data. Zhang et al. (2022) proposed an MRC method, using biaffine attention to score the start and end of each entity in a text. Yuan et al. (2021) further extended this idea by introducing a triaffine attention mechanism, which fuses heterogeneous factors to achieve superior span representation. Nevertheless, both approaches face computational complexity issues, which can affect their efficiency. Other researchers (Dang et al., 2023; Tan et al., 2021) have utilized instance queries to adaptively learn semantics and extract all types of entities simultaneously. These queries can be pre-constructed without relying on external knowledge, offering a flexible solution to improve NER performance.

NER of Legal Documents

In recent years, the urgent demand for judicial data analysis has led to the emergence of NER research in the judicial field. Li et al. (2023) discussed the main factors restricting the development of legal intelligent systems. Wang et al. (2020) proposed a method utilizing a pre-trained model and an iterative dilated convolutional neural network, demonstrating its effectiveness on Brazilian legal documents, while Leitner et al. (2019) developed an NER method for German legal documents based on a BiLSTM-CRF (Conditional Random Field) model, achieving effective recognition of 19 entity types. Ji et al. (2020) developed a multi-task learning framework and proposed an end-to-end joint model to simultaneously extract evidence information and classify evidence in legal documents. Li et al. (2021) and Zhang et al. (2023) both focused on entity recognition for theft-related cases, albeit with different approaches. Li et al. (2021) introduced a BERT-ON-LSTM (Ordered Neuron-Long Short-Term Memory)-CRF model to address the polysemy issue in word vector embedding, aligning the NER system's performance with the specific needs of the judicial domain. Meanwhile, Zhang et al. (2023) incorporated char-level and word-level feature representations to enrich domain semantics, utilizing a global pointer method to accurately identify entity locations and boundaries. Wang et al. (2020), Guo et al. (2022), Wang et al. (2023), and Yang et al. (2024) have all proposed methods to improve NER for legal documents by incorporating multiple feature types, including sentence-level, word-level, and domain-specific legal terminology features. Zeng et al. (2022) integrated the NER task with the Chinese word segmentation task in a joint training framework, thereby improving the recognition rate of entity boundaries. Similarly, Mao et al. (2023) constructed a joint learning model based on BERT to identify judicial named entities, emphasizing the comprehensive utilization of both global text information and lexical local information.

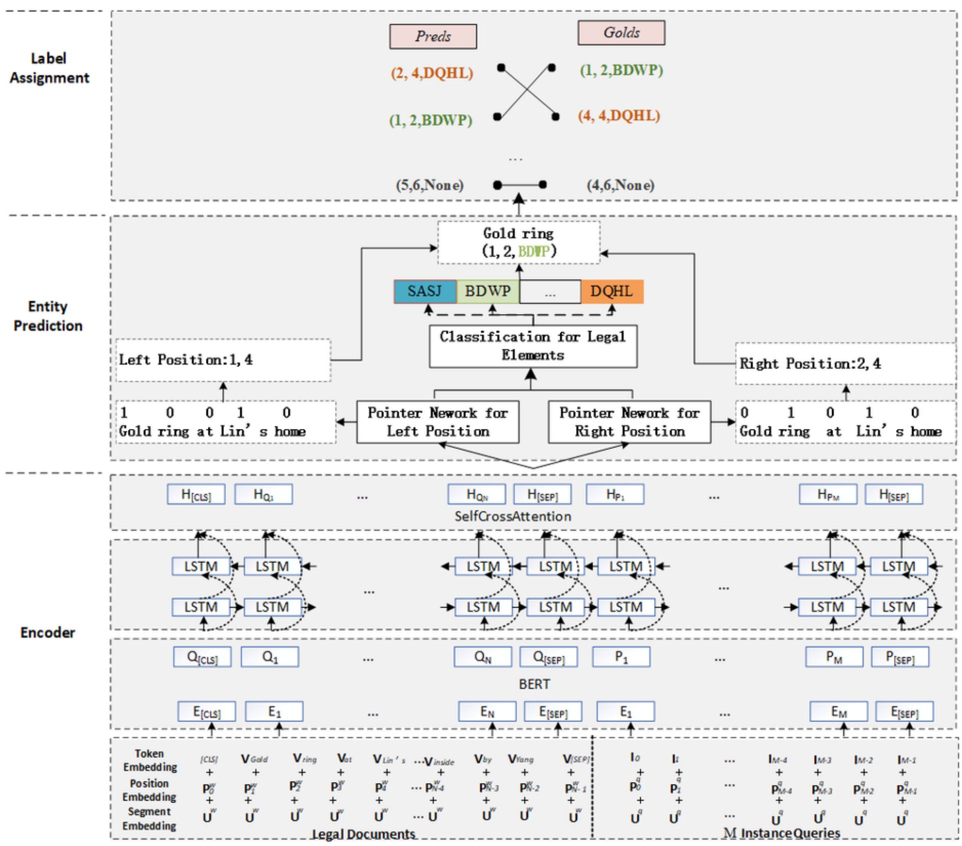
Although some progress has been made in LNER, including improvements in entity recognition accuracy, challenges such as overlooking semantic information in the vocabulary, unclear entity boundaries, difficulty in identifying long entities, and the complexity of nested entities still persist. Furthermore, directly applying entity recognition models designed for general domains to specialized fields like legal documents often leads to inaccurate results. This is due to the difficulty such models face in identifying domain-specific terminology, resulting in entity recognition that fails to capture precise semantic information. Therefore, it is crucial to develop more domain-specific models capable of accurately identifying and capturing the specialized terms and entities in the legal domain.

LNER is designed to automatically identify key information relevant to legal cases, thereby highlighting the essential elements of these cases. As a result, LNER has become a critical component in e-discovery. By implementing LNER technology, investigators are relieved of the labor-intensive tasks of manual browsing, selection, and analysis. Moreover, when integrated into electronic data forensic systems, LNER provides a solid foundation for accurate analysis and enables more efficient follow-up investigations.

METHODOLOGY

This section first introduces the named entity identification task and then elaborates on each component of the proposed method. The method consists of two main parts: the BERT+BiLSTM+attention encoder module and the legal entity prediction module based on the parallel instance query network. Additionally, a mechanism is introduced to dynamically assign labels to the loss function, aligning gold entities with instance queries to achieve more accurate labels. An overview of the model is shown in Figure 1. The following subsections provide a more detailed explanation.

Figure 1. Chinese legal documents based on parallel instance query network



Note. SASJ: symbol for the time of the incident; BDWP: stolen item; DQHL: criminal profit from theft. All symbols are abbreviations of Chinese Pinyin representations for corresponding entities.

Problem Definition

A sample is denoted by (X, Y), where $X = (x_1, x_2, \ldots, x_N)$. N words are labeled by triple sets $Y = \left\{ y_k^t, y_k^l, y_k^r \right\}_{k=0}^{G-1}$. $y_k^t \epsilon [1, \alpha], y_k^l \epsilon [0, N-1]$ and $y_k^r \epsilon [0, N-1]$ are indices for the entity type, start position, and end position of the k-th element, respectively, where α is the entity type. The task of entity identification is $f{:}X \rightarrow Y$, that is finding out the span of each entity and classify type with a predefined entity type set. Unlike flat NER, the words of nested named entities in legal documents have more than one tag.

Table 1 shows four examples of nested entities, which focus on victim, location, crime tool, stolen item, and organization. In example c, the organization type for "The Financial Market Supervisory Authority of Tianqiao District, Jinan City" is a long entity and is nested with the crime location type for "Tianqiao District, Jinan City."

Table 1. Examples of named entities in legal documents

#	Nested type	Examples (Translated from Chinese)
1	Victim and Location	Victim Victim Cao runs the flower shop located on * * Street Location
2	Victim and Stolen Item	Victim The defendant Gao stole Wang Moujia' s Thinkpad laptop Stolen Item
3	Location and Organization	Location The Financial Market Supervisory Authority of Tianqiao District, Jinan City Oragnization
4	Crime Tool and Stolen Item	Crime Tool Defendant Liang used a stolen kitchen knife to pry open the gold counter Stolen Item

Encoder

Input Embedding

The input of the BERT layer is a concatenation of token embeddingsE_t, segment embeddingsE_s, and position embeddings$E_p$from two sequences shown in Figure 1, where [CLS] is the beginning of the sentence and [SEP] denotes the end of the sentence. The input E is calculated as follows:

$$\begin{aligned} E_t &= Concat(V, I) \\ E_p &= Concat(P^w, P^q) \\ E_s &= Concat(U^w, U^q) \\ E &= E_t + E_p + E_s \end{aligned} \tag{1}$$

where$V$and$I$are token embeddings of the input sentence, and the instance query vectors, P^w, P^q, U^w and U^q are separate learnable position embeddings and segment embeddings, respectively.

The core of BERT is the bidirectional transformer encoding structure, and its self-attention part interacts with the word in a sentence and instance queries. Instance queries are initialized randomly according to a normal distribution, which can influence sentence encoding and affect sentence semantics. Therefore, one-way self-attention is adopted instead of normal self-attention, as shown in Equation (2):

$$Attention(Q, K, V) = softmax\left(\frac{QK^T}{\sqrt{h}} + M\right)V \tag{2}$$

where h is the dimensionality of the input vector, Q, K and V are parameter matrices, and the attention score is represented by the mask matrix $M \in \{0, -inf\}$. Sentence encoding cannot be applied to instances because the upper right sub-matrix of M is a full $-inf$ matrix of size $(N \times M)$ and has no other elements. For kept units, the elements in M are set to 0, while the removed ones are $-inf$.

BiLSTM and Transformer

After BERT encoding, BiLSTM layers and a transformer are used to obtain the sentence context vector. The calculation process is shown in equation (3):

$$H_t = \left[\overrightarrow{H_t}; \overleftarrow{H_t}\right] \tag{3}$$

where $\overrightarrow{H_t}$ and $\overleftarrow{H_t}$ are the outputs of forward LSTM and backward LSTM at time t, respectively, and $[;]$ means the concatenation of two hidden vectors.

The input of each transformer is based on the output of the preceding one. The output of the u-th transformer can be represented as:

$$H_u = TransformerBlock\left(H_{u-1}\right), 1 < u < L \tag{4}$$

$$H = H_{L-1} \tag{5}$$

where$L$is the layer number. Finally, $H \in \mathbb{R}^{(N+M)\times h}$ is split into two parts: sentence encoding $H_Q \in \mathbb{R}^{N\times h}$ and instance query encoding $H_P \in \mathbb{R}^{M\times h}$.

Entity Prediction

Entity prediction is treated as two subtasks: extraction and classification. The former predicts the element boundary (left and right positions) and the latter recognizes the label of each entity with instance queries. An entity boundary and an entity classifier are used for them, respectively.

Entity Boundary

For the i-th instance query H_P^i and j-th word H_Q^j, two feedforward layers are used, relying on the begin and end indices of the span, as shown in equations (6) and (7):

$$R_{ij}^l = ReLU\left(W_P^l H_P^i + W_Q^l H_Q^j\right) \tag{6}$$

$$R_{ij}^r = ReLU\left(W_P^r H_P^i + W_Q^r H_Q^j\right) \tag{7}$$

where l and r denote the left and right boundary and W_P^l *and* W_Q^l are learnable parameter matrices. The probability of the j-th word in the sentence being a boundary word is calculated in the following way:

$$P_{ij}^l = sigmoid\left(R_{ij}^l W_l + b_l\right) \tag{8}$$

$$P_{ij}^r = sigmoid\left(R_{ij}^r W_r + b_r\right) \tag{9}$$

where W_l, W_r, b_l, b_r are trainable parameter matrices.

Entity Classifier

The entity span is important for the entity classifier. P_i^l and P_i^r are applied to weight the words and then concatenate them with an instance query as follows:

$$R_i^t = ReLU\left(W_P^t H_P^i; P_i^l H_Q; P_i^r H_Q\right) \tag{10}$$

where W_P^t is a learnable parameter. The entity probability queried by the *j*-th instance query belonging to type α can then be calculated as follows:

$$P_{i\alpha}^t = \frac{exp\left(R_i^t W_\alpha^t + b_\alpha^t\right)}{\sum_{\alpha'} exp\left(R_i^t W_{\alpha'}^t + b_{\alpha'}^t\right)} \tag{11}$$

where $W_{\alpha'}^t$ and $b_{\alpha'}^t$ are learnable parameters. Finally, the prediction of legal entities is achieved via

$$\hat{Y} = < \hat{y}_i^t, \hat{y}_i^l, \hat{y}_i^r > \tag{12}$$

where $\hat{y}_i^t = argmax(P_{i\alpha}^t), \hat{y}_i^l = argmax\left(P_{ij}^l\right)$ *and* $\hat{y}_i^r = argmax\left(P_{ij}^r\right)$ are the type of entity, left boundary and right boundary, respectively.

Loss Function

Instance queries are used to learn more semantics from legal documents, rather than relying on external knowledge. During the training stage, instance queries are represented as vectors, and it is not possible to assign gold entities to them ahead of time. To address this issue, a loss function is designed based on the allocation of named entities to the instance queries. Prior to calculating the training loss, the optimal allocation between the instance query and the set of gold entities needs to be determined first. Specifically, this allocation is formulated as an optimal linear assignment problem (LAP) with a minimal assignment cost.

The optimization objective of the LAP is defined as:

$$min \sum_{i=0}^{M-1} \sum_{k=0}^{G} A_{ik} cost_{ik}$$

$$s.t. \sum_k A_{ik} \le 1 \tag{13}$$

$$\sum_i A_{ik} = q_k \tag{14}$$

$$A_{iG} = \begin{cases} 0, \sum_k A_{ik} = 1 \\ 1, \sum_k A_{ik} = 0 \end{cases} \tag{15}$$

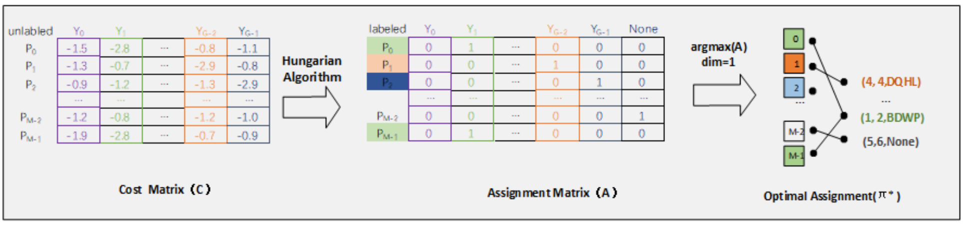
$$Cost_{ik} = -\left(P_{iy_k^t}^t + P_{iy_k^l}^l + P_{iy_k^r}^r\right) \tag{16}$$

$$\forall i, k, A_{ik} \in \{0, 1\} \tag{17}$$

where constraints (13), (14) sand (17) are for the assignment matrix A. Constraint (16) is the cost function of assigning the *k*-th entity $Y = \left\langle y_k^t, y_k^l, y_k^r \right\rangle$ to the *i*-th instance query, and y_k^t, y_k^l and y_k^r are defined in the 'Problem Definition' section above. q_k is the quantity that can be assigned to the *k*-th gold entity. Since the total number of instance queries surpasses the total assignable quantity of entity labels, some of them will not be assigned to any entity label. Column G is extended to set the None label by using constraint (15).

Then, the Hungarian algorithm is used to solve the LAP, yielding matrix A with cost matrix C in Figure 2. Based on A, the label $\hat{Y}$ index by (π^*) for M instance queries can be further obtained using π^*, where $\pi^* = \underset{dim=1}{argmax} A$ is the label index vector for instance queries under optimal assignment.

Figure 2. Label assignment



Note. BDWP: stolen item; DQHL: criminal profit from theft. All symbols are abbreviations of Chinese Pinyin representations for corresponding entities.

The probability of the entity boundary and entity classification has been obtained, and the labels $\hat{Y}$ have been computed based on the minimum value of the cost matrix. The boundary loss and classification loss are used to train this model. For boundary prediction, the task of predicting the left and right boundaries of a named entity is treated as a binary classification problem, with binary cross-entropy used to calculate the loss:

$$L_b = -\sum_{d\in\{l,r\}} \sum_{i=0}^{N-1} \sum_{j=0}^{M-1} F\left[\hat{Y}_i^d = j\right] \log P_{ij}^d + F\left[\hat{Y}_i^d \neq j\right] \log\left(1 - P_{ij}^d\right) \tag{18}$$

and cross entropy for entity classification is used to calculate the loss:

$$L_t = -\sum_{i=0}^{M-1} \sum_{z\in\varepsilon} F\left[\hat{Y}_i^t = z\right] \log P_{iz}^t \tag{19}$$

where $F[e]$ represents the indicator function, which takes a value of 1 when e is true and 0 otherwise.

Thus, the total loss on the train set D can be defined as:

$$L = \sum_D \sum_{\tau=0}^{L} L_t^\tau + L_b^\tau \tag{20}$$

where L_t^τ, L_b^τ are classification loss and boundary loss at the τ-th layer (Lin et al., 2017; Shen et al., 2022).

EXPERIMENTAL ANALYSIS

Dataset and Experiment Settings

Criminal judgments from real-life theft cases, as well as public datasets available on Chinese judgment websites (https://wenshu.court.gov.cn/), were collected for this study. These documents were converted into TXT files without losing any content and then manually annotated using the open-source annotation tool Doccano (http://doccano.herokuapp.com/). The annotation process was guided by techniques from both domestic and international research (Cai et al., 2020), ensuring effective and accurate labeling. To segment words and sentences, universal information extraction (UIE; Lu et al., 2022) provided by Baidu Paddle-Paddle was employed. Additionally, small samples were utilized to fine-tune the UIE model and then roughly label the samples. The data were annotated by three graduate students and two legal experts were invited to develop annotation standards and conduct sample surveys on the annotated data. These standards are based on the requirements for e-discovery. For example, the location of the incident refers to the geographical information involved in the case, which should be as detailed as possible. This includes, but is not limited to, administrative district names, street names, community names, building numbers, and floor numbers. The experts also conducted sample surveys on the annotated data to assess the initial quality of the annotations. The annotation process comprised three stages: In the first stage, the case texts were annotated based on UIE. Following inspection by the expert team, the annotation specifications were revised and provided to the annotators. In the second stage, the annotated texts were examined and revised using the updated specifications, and the verification process from the first stage was repeated. In the third stage, based on the feedback accumulated from the previous stages, all the case texts were annotated, with experts conducting sampling inspections. The annotation process was completed with test consistency maintained at over 93%. From this, a corpus of theft cases was constructed.

There are 10 categories of named entities, time, location, organization, suspect, victim, crime tool, stolen item, stolen money, high value, and theft profit as shown in Table 2. Due to various factors affecting sample collection, the distribution of various entity categories in the sample is not balanced, with a relatively small sample size of entity types, such as theft profit and crime tool.

Table 2. Entity description of the corpus of theft cases

Type	Label	Description	Entity Number	Nested Entity Number
Time	SASJ	Time of the incident refers to the time expression during the occurrence of the case, including calendar time (such as year, month, day) and non-calendar time (such as morning, afternoon, evening, etc.)	2765	2
Location	SADD	Location refers to the geographical information involved in the case and should be annotated as detailed as possible. It includes administrative district names, street names, community names, building numbers, floor numbers, etc. Additionally, it should include location indicators, such as “50 meters west from the house”.	3517	325
Organization	SAJG	Organization refers to the administrative organizations, business organizations, or non-governmental organizations involved in the case.	806	54
Suspect	FZXYR	Suspect is a subtype of the general 'Person' entity, referring to an individual alleged to have committed a crime.	6463	4
Victim	SHR	Victim is a subtype of the general 'Person' entity, representing an individual who suffered a loss in the cases.	3108	602
Crime Tool	ZAGJ	Item refers to both public and private property involved in the case. Crime Tool is a subtype of the general 'Item' entity, referring to instruments or objects used specifically during the commission of theft, such as crowbars, lockpicks, and other burglary tools.	735	30
Stolen Item	BDWP	Item refers to both public and private property involved in the case. Stolen Item is a subtype of the general 'Item' entity, referring to personal property unlawfully taken from the victim. This can range from small items, like jewelry, to larger items, such as vehicles. To distinguish items more accurately, their attributes (such as quantity, color, brand) are also specified.	5781	747

continued on following page

Table 2. Continued

Type	Label	Description	Entity Number	Nested Entity Number
Stolen Money	BDJE	Stolen money is a subtype of the general 'Currency' entity, referring to the currency obtained by the suspect through theft. The stolen currency attributes (such as RMB, USD, etc.) also need to be annotated to differentiate between types of currency.	915	82
High Value	WPJZ	High Value is a subtype of the general 'Currency' entity, referring to the value of the stolen item. The value of the stolen item attributes (such as RMB, USD, etc.) also need to be annotated to differentiate between types of currency.	2090	124
Theft Profit	DQHL	Theft Profit is a subtype of the general 'Currency' entity, referring to the criminal profit from theft. The profit from the stolen item attributes (such as RMB, USD, etc.) also need to be annotated to differentiate between types of currency.	481	0

Note. SASJ: symbol for the time of the incident; SADD: geographical information involved in the case; SADD: geographical information involved in the case; SAJG: organizations involved in the case; FZXYR: suspect; SHR: victim; ZAGJ: crime tool; BDWP: stolen Item; BDJE: currency obtained by the suspect through theft; WPJZ: value of the stolen item; DQHL: criminal profit from theft. All symbols are abbreviations of Chinese Pinyin representations for corresponding entities.

The dataset, which is in JSON(JavaScript Object Notation) format, was randomly divided into training, validation, and test sets in an 8:1:1 ratio. Each sample is represented as a dictionary with the following specific format. Notably, the text in the “tokens” field is translated from Chinese:

```
{
"tokens": ["After", "the", "case", "was", "solved", ",", "the", "Public", "Security", "Bureau", "returned", "the", "seized", "mobile", "phones", "to", "the", "victims", "of", "Yan", "and", "Xiao", "."],
"entities": [
{"type": "SHZ", "start": 20, "end": 20 },
{"type": "SHZ", "start": 22, "end": 22 },
{"type": "BDWP", "start": 8, "end": 10},
],
"id": "8818"
}
```

To validate the effectiveness of the proposed solution, experiments were conducted using legal documents from theft cases, which included 6,690 data samples, totaling 335,010 characters across 10 categories and 6,463 named entities. Within the dataset, 84.26% of the sentences contain fewer than 100 words, 14.94% range from 100 to 200 words, 0.72% are between 200 and 300 words, and 0.08% exceed 300 words.

Unlike the judicial smooth entity dataset in BIO(Begin, Inside, Outside) format, the dataset was formatted in the “entity type - start - end” format for model learning. Each sentence was segmented into characters and placed into the “token” field, with the corresponding labels stored in the “entities” field.

Experiments were conducted on a single Graphics Processing Unit (GPU), GeForce RTX3060, using PyTorch to build the model with a pre-trained network. The training iterations ranged from 30 to 60, with a maximum sequence length of 512, a batch size of 8, a learning rate of 0.001, a dropout rate of 0.5, and a BiLSTM hidden layer dimension of 384. Following previous work, the results were analyzed using micro precision, recall, and F1 scores.

Results and Analysis

To ensure a fair and comprehensive comparison with previous work, the proposed model was compared to six baseline models.

- BERT+CRF (Hu et al., 2022): word vectors are encoded with BERT and vectors are fed into the CRF layer to generate a state transition matrix that considers the current token's contextual information when decoding (Shen et al., 2021).
- BERT+BiLSTM+CRF (Guo et al., 2021): a BiLSTM layer is added to the BERT+ CRF model to learn contextual features.
- BERT+MRC (Li et al.,2019): type-specific queries for entity categories are constructed using prior semantic information in the NER task, which is a question-answering task.
- BERT+SPAN (Yuan et al., 2021): a span prediction model, which acts as a system combiner to recognize named entities across different system outputs, is designed.
- W2NER (Li et al., 2022): unified NER is formulated as a word–word relation classification problem. This model tackles the neighboring relations between entity words with next neighboring word and tail–head–word relations.
- Global pointer (Su et al., 2022): a global pointer network is designed, utilizing a multiplicative attention mechanism to capitalize on information from relative positions.

The methods used in these six models were replicated and executed on our dataset for experimentation, with the results presented in Table 3.

Table 3. Results comparison (%)

Model	P	R	F1
W2NER (Li et al., 2022):	91.83	69.26	78.98
BERT+CRF (Hu et al., 2022)	88.68	89.45	89.06
BERT+BiLSTM+CRF (Guo et al., 2021)	84.28	90.08	87.08
BERT+MRC (Li et al., 2019)	88.13	87.95	88.03
BERT+SPAN (Yuan et al., 2021)	87.75	90.11	88.91
BERT+ Global pointer (Su et al., 2022)	89.39	88.98	89.18
Our model	88.87	90.26	89.56

The F1 scores of our proposed PIQN-NER model consistently exceed those of the other six models, indicating the superior overall effectiveness of our model. The PIQN-NER, BERT+SPAN, and BERT+ Global pointer models all utilize span-based pointer networks for entity recognition, where two multi-classification networks are employed to determine the start and end boundaries between different entities.

This approach proves superior to BERT+BiLSTM+CRF and W2NER, which do not utilize pointer networks, demonstrating the beneficial impact of pointer networks in entity recognition tasks. The F1 scores of the PIQN-NER model are, respectively, 0.38 and 1.53 percentage points higher than those of the BERT+ Global pointer and BERT+MRC models, indicating that the parallel instance query network could dynamically learn query semantics. There are three factors that may explain this:

- As a variant of pointer networks and MRC, the parallel instance query network reduces the dependence on external knowledge injection, thus avoiding the influence of quality fluctuations in manually constructed queries.
- Each query only predicts an entity instead of a group of entities of a particular type, thus enhancing query refinement with refined query semantics at the entity level.
- The model can be encoded and perform predictions by adding instance queries in parallel, leveraging the inherent relationships between entities presented by different instance queries.

To further assess the effectiveness of the proposed model, two models with superior performance were selected for comparison. The F1 scores for each model were calculated separately for each entity type. The results of these experiments are presented in Table 4. As shown, the PIQN-NER model consistently outperforms the comparison models across all 10 entity categories, demonstrating superior recognition accuracy.

Table 4. Results of detailed legal entities F1

Model	SASJ	SADD	SAJG	FZXYR	SHZ	ZAGJ	BDWP	BDJE	WPJZ	DQHL
BERT+CRF	**92.64**	**81.22**	**84.68**	**96.00**	**93.31**	**79.28**	**81.11**	**81.73**	**97.01**	**85.86**
BERT + Global pointer	**92.18**	**79.85**	**86.95**	**96.27**	**93.22**	**79.61**	**81.32**	**81.94**	**97.25**	**83.94**
PIQN-NER	**93.01**	**84.19**	**86.96**	**96.45**	**93.66**	**79.90**	**81.72**	**82.58**	**97.82**	**87.05**

Table 5 shows the performance of entity identification using the PIQN-NER model. Among 10 entity categories, the P, R, and F1 values on six entities (suspects, high value, etc.) exceed 80%, highlighting the model's advantages in fine-grained entity recognition. However, the recognition performance for certain entities, such as crime tool, stolen item, stolen money, and location, is relatively low, with F1 scores of 79.90%, 81.72%, 82.58%, and 84.19%, respectively.

Table 5. Results of detailed legal entities (%)

Type	P	R	F1
SASJ	94.22	91.84	93.01
SADD	84.77	83.62	84.19
SAJG	85.00	89.01	86.96
FZXYR	95.50	97.41	96.45
SHZ	92.90	94.44	93.66
ZAGJ	77.72	82.20	79.90
BDWP	80.05	83.45	81.72
BDJE	85.22	80.09	82.58
WPJZ	97.18	98.47	97.82
DQHL	81.55	93.33	87.05

Unlike victim and suspect entities, which have explicit boundary features, the crime tool and stolen item entities lack characteristic sentence structures, making them difficult to identify. The expression of the crime time and the organization involved in the judicial elements is more complex and prone to semantic interference, resulting in a reduced recognition performance.

Ablation Experiment

In the ablation experiment, the effects of various components of the PIQN-NER model were examined; the results are shown in Table 6. The following experiments were performed: (a) without PIQN, that is, the parallel instance query network was replaced with CRF, and (b) without CE, that is, cross-entropy loss was replaced with the focus loss.

Table 6. Ablation experiment (%)

Model	Loc.F1	Cls.F1	P	R	F1
w/o PIQN	88.83	93.36	85.59	88.96	87.25
w/o CE	90.39	93.92	88.38	89.73	89.05
Our model	90.49	94.23	88.87	90.26	89.56

Note. PIQN = Parallel Instance Query Network; CE = Cross Entropy; Loc.F1 = Location F1; Cls.F1 = Classification F1; P = Precision; R = Recall.

Compared to the CRF, the PIQN significantly improved the localization and classification by +1.66% and +0.87%, respectively, and resulted in P, R, and F1 score increases of +3.28, +1.30%, and +2.31, respectively. This suggests that the PIQN can capture the semantics between the sentence and instance query, rather than relying on external knowledge or pre-specifying entity labels. Thus, we replaced focal loss with cross-entropy loss as the loss function. Compared to the model without CE, our model exhibits an improved F1 score by 0.51%.

Case Study

The results of a case study on predictions are presented in Table 7. The left column contains labels indicating the type of entity in the lower right corner, with the superscripts showing the start and end position of the entity. The named entity that was predicted is displayed in the right column.

Table 7. Case study

#	Sentence with Entities (Translated from Chinese)	Prediction
1	$[^{0}$Wang$^{0}]_{FZXYR}$ stole $[^{2}$a black Apple phone$^{5}]_{BDWP}$from $[^{7}$the bedroom of $[^{10}$Yang's$^{10}]_{SHZ}$ home$^{11}]_{SADD}$ and used $[^{14}$a chisel$^{15}]_{ZAGJ}$ to break the padlock on the bedside wooden cabinet, stealing $[^{27}$20,000 RMB cash$^{29}]_{BDJE}$ placed inside by $[^{33}$Yang$^{33}]_{FZXYR}$.	√ (0,0, FZXYR) √ (2,5, BDWP) √ (7,11, SADD) √ (10,10, SHZ) √ (14,15, ZAGJ) √ (27,29, BDJE) √(33,33, FZXYR)
2	As appraised by [3rde Price Certification Center of Nanhu District, Nong'an City$^{12}]_{SAJG}$, it was determined that $[^{18}$the stolen electric bicycle belonging to $[^{26}$Mr. Long$^{27}]_{SHZ}]]_{BDWP}$ was valued at $[^{30}$2930 RMB$^{31}]_{WPJZ}$.	× (8,12, SADD) √ (26,27, SHZ) × (18,21, BDWP) √ (30,31, BDJE)
3	In summary, $[^{3}$Deng$^{3}]_{SHZ}$ committed two thefts, with a total amount of $[^{13}$6250 RMB$^{14}]_{WPJZ}$.	√ (3,3, SHZ) × (13,14, BDJE)
4	$[^{0}$On June 22, 2018, at approximately 11:40 a.m.$^{9}]_{SASJ}$, $[^{11}$Wang$^{11}]_{FZXYR}$ stole $[^{13}$a pack of Golden Anhui cigarettes$^{18}]_{BDWP}$. $[^{20}$On the same day$^{23}]_{SASJ}$, he sold the stolen $[^{29}$gold jewelry$^{30}]_{BDWP}$.	√ (0,9, SASJ) √ (11,11, FZXYR) √ (13,18, BDWP) × (0,9, SASJ) × (27,30, BDWP)

Our model demonstrates a good performance in recognizing both nested entities and long entities. In case 1 in Table 7, the model accurately predicts the five-word-long entity and its two-level nested structure; “the bedroom of Yang’s home” and “Yang’s” were correctly classified as SADD and SHZ, respectively. Complete entity identification effectively indicates the involved persons, locations, and items, thereby revealing the intrinsic dependencies in the semantic domain.

The model's understanding of legal documents remains limited, possibly due to the following reasons:

- The model concentrates too much on local semantics. In case 2, the PIQN-NER incorrectly labels “Nanhu District, Nong'an City” as SADD, failing to recognize the long phrase ‘the Price Certification Center of Nanhu District, Nong'an City’, which should be a long entity of SAJG. Similarly, the model concentrates on the local context of “the stolen electric bicycle” while disregarding the preceding element “Mr. Long.”
- There is confusion regarding similar entity types. The model confuses the classifications of WPJZ and BDJE and misidentifies ‘2930 RMB’ as BDJE in case 3. This misclassification primarily stems from the relatively lower amount of samples, hindering the model from fully grasping the features of these types. Statistical analysis reveals that nearly 50% of WPJZ instances were identified as BDJE.
- Some specific phrases are misunderstood. The model fails to identify the element ‘On the same day’ in case 4, which should be recognized as SASJ.

CONCLUSIONS

This paper presents a PIQN-NER method designed to extract entities from legal documents using learnable instance queries. Unlike previous approaches, it does not rely on external knowledge for pre-designed queries. Instead, the semantics of an instance query are implicit, necessitating a label assignment mechanism to match gold entities with instance queries. Experimental results validate the effectiveness of the proposed method, demonstrating its superior performance on legal datasets compared to related approaches.

The recognition of nested entities in legal documents is challenging due to the complexity of their semantic structures. Consequently, refining recognition methods and evaluating the model on diverse crime case documents will be key areas of focus in future research. Furthermore, future research will extend the PIQN-NER method to address the recognition of complex entities in long legal text. This is crucial as legal documents often contain extensive narratives and detailed descriptions, requiring robust methods to accurately identify and extract relevant information.

COMPETING INTERESTS

The authors have declared that no competing interests exist.

FUNDING STATEMENT

This research was supported by the Program for Liaoning Provincial Department of Science and Technology Applied Basic Research Program Project [grant number 2023JH2/101300134], the Key Laboratory of Big Data Intelligent Application for Liaoning Provincial Public Security Project [grant number 2022JH13/10200047], the Key Research Project of the Liaoning Provincial Ministry of Education [grant number JYTZD2023088], the Basic Research Project of Liaoning Provincial Ministry of Education [grant number LJKMZ20221549], and the Liaoning Graduate Education Teaching Reform Research Project [grant number LNYJG2022423].

PROCESS DATES

December 31, 2024
Received: September 3, 2024, Revision: December 2, 2024, Accepted: December 15, 2024

CORRESPONDING AUTHOR

Correspondence should be addressed to Rui Lu (China, luruilly@sina.com)

REFERENCES

Cai, L., Wang, S., & Liu, J. (2020). Survey of data annotation. *Journal of Software*, *31*(2), 302–320. DOI: 10.13328/j.cnki.jos.005977

Correia, F., Almeida, A., Nunes, J., Santos, K., Hartmann, I., Silva, F., & Lopes, H. (2022). Fine-grained legal entity annotation: A case study on the Brazilian supreme court. *Information Processing & Management*, *59*(1), 102794. Advance online publication. DOI: 10.1016/j.ipm.2021.102794

Dang, X., Wang, L., Dong, X., Li, F., & Deng, H. (2023). DdERT: Research on named entity recognition for mine hoist using a Chinese BERT model. *Electronics (Basel)*, *12*(19), 4037. Advance online publication. DOI: 10.3390/electronics12194037

Guo, L., Li, Y., & Wang, S. (2022). Named entity recognition in legal instruments based on matching strategy and community attention mechanism. *Journal of Chinese Information Processing*, *36*(2), 85–92. DOI: 10.3969/j.issn.1003-0077.2022.02.010

Guo, Z., & Deng, X. (2021). Intelligent identification method of legal case entity based on BERT-BiLSTM-CRF. *Journal of Bejing University of Posts and Telecommunications*, *44*(4), 129–134. DOI: 10.13190/j.jbupt.2020-241

Hu, S., Zhang, H., Hu, X., & Du, J. (2022). Chinese named entity recognition based on BERT-CRF model. In *Proceedings of 22nd international conference on computer and information science* (pp. 105–108). IEEE. DOI: 10.1109/ICIS54925.2022.9882514

Ji, D., Tao, P., Fei, H., & Ren, Y. (2020). An end-to-end joint model for evidence information extraction from court record document. *Information Processing & Management*, *57*(6), 102305. Advance online publication. DOI: 10.1016/j.ipm.2020.102305

Ju, M., Miwa, M., & Ananiadou, S. (2018). A neural layered model for nested named entity recognition. In *Proceedings of NAACL-HLT* (pp. 1446–1459). ACL. DOI: 10.18653/v1/N18-1131

Katiyar, A., & Cardie, C. (2018). Nested named entity recognition revisited. In *Proceedings of 2018 conference of the North American chapter of the Association for Computational Linguistics: Human language technologies* (*Vol. 1*, pp. 861–871). ACL. DOI: 10.18653/v1/N18-1079

Leitner, E., Rehm, G., & Moreno-Schneider, J. (2019). Fine-grained named entity recognition in legal instruments. In Acosta, M., Cudré-Mauroux, P., Maleshkova, M., Pellegrini, T., Sack, H., & Sure-Vetter, Y. (Eds.), Lecture notes in computer science: *Vol. 11702*. *Semantic systems. The power of AI and knowledge graphs* (pp. 272–287). Springer., DOI: 10.1007/978-3-030-33220-4_20

Li, C., Wang, L., Sun, Y., & Lin, H. (2021). BERT based named entity recognition legal texts on theft cases. *Journal of Chinese Information Processing*, *35*(8), 73–81. DOI: 10.3969/j.issn.1003-0077.2021.08.010

Li, J., Fei, H., Liu, J., Wu, S., Zhang, M., Teng, C., Ji, D., & Li, F. (2022). Unified named entity recognition as word-word relation classification. *Proceedings of the AAAI Conference on Artificial Intelligence*, *36*(10), 10965–10973. DOI: 10.1609/aaai.v36i10.21344

Li, J., Li, Y., & Ge, F. (2023). Survey of research on intelligent system for legal domain. *Computer Engineering and Applications*, *59*(7), 31–50. DOI: 10.1007/978-981-99-0799-1_3

Li, X., Feng, J., Meng, Y., Han, Q., Wu, F., & Li, J. (2019). *A unified MRC framework for named entity recognition*. arXiv. https://doi.org//arXiv.1910.1147DOI: 10.48550

Lin, T., Goyal, P., Girshick, R., He, K., & Dollar, P. (2017). Focal loss for dense object detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *42*(2), 318–327. DOI: 10.1109/TPAMI.2018.2858826 PMID: 30040631

Lu, W., & Roth, D. (2015). Joint mention extraction and classification with mention hypergraphs. In *Proceedings of the 2015 conference on empirical methods in natural language processing* (pp. 857–867). ACL. DOI: 10.18653/v1/D15-1102

Lu, Y., Liu, Q., & Dai, D. (2022). *Unified structure generation for universal information extraction.* arXiv. / arXiv.2203.12277DOI: 10.18653/v1/2022.acl-long.395

Mao, X., Chen, X., & Ning, K. (2023). Global and local information integration for recognizing key named entity. *Journal of Software*, *34*(12), 5724–5736. DOI: 10.13328/j.cnki.jos.006903

Shen, Y., Ma, X., Tan, Z., Zhan, S., Wang, W., & Lu, W. (2021). Locate and label: A two-stage identifier for nested named entity recognition. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing* (*Vol. 1*, pp. 2782–2794). ACL. DOI: 10.18653/v1/2021.acl-long.216

Shen, Y., Wang, X., Tan, Z., Xu, G., Xie, P., Huang, F., Lu, W., & Zhuang, Y. (2022). Parallel instance query network for named entity recognition. In *Proceedings of the 60th annual meeting of the Association for Computational Linguistics* (*Vol. 1*, pp. 947–961). ACL. DOI: 10.18653/v1/2022.acl-long.67

Shibuya, T., & Hovy, E. (2020). Nested named entity recognition via second-best sequence learning and decoding. *Transactions of the Association for Computational Linguistics*, *8*, 605–620. DOI: 10.1162/tacl_a_00334

Sohrab, M., & Miwa, M. (2018). Deep exhaustive model for nested named entity recognition. In *Proceedings of the 2018 conference on empirical methods in natural language processing* (pp. 2843–2849). ACL. DOI: 10.18653/v1/D18-1309

Straková, J., Straka, M., & Hajic, J. (2019). Neural architectures for nested NER through linearization. In *Proceedings of the 57th annual meeting of the Association for Computational Linguistics* (pp. 5326–5331). ACL. https://arxiv.org/abs/1908.06926 DOI: 10.18653/v1/P19-1527

Su, J., Murtadha, A., Pan, S., Hou, J., Sun, J., Huang, W., Wen, B., & Liu, Y. (2022). *Global Pointer: Novel efficient span-based approach for named entity recognition.* arXiv. https://arxiv.org/pdf/2208.03054

Tan, C., Qiu, W., Chen, M., Wang, R., & Huang, F. (2020). Boundary enhanced neural span classification for nested named entity recognition. *Proceedings of the AAAI Conference on Artificial Intelligence*, *34*(5), 9016–9023. DOI: 10.1609/aaai.v34i05.6434

Tan, Z., Shen, Y., Shuai, Z., Lu, W., & Zhuang, Y. (2021). A sequence-to-set network for nested named entity recognition. *Proceedings of the thirtieth international joint conference on artificial intelligence (IJCAI-21)* (pp. 3936–3942). DOI: 10.24963/ijcai.2021/542

Wang, B., & Lu, W. (2018). Neural segmental hypergraphs for overlapping mention recognition. In *Proceedings of the 2018 conference on empirical methods in natural language processing* (pp. 204–214). ACL. DOI: 10.18653/v1/D18-1019

Wang, D., Wang, S., & Pei, W. (2020). Named entity recognition based on JCWA-DLSTM for legal instrument. *Journal of Chinese Information Processing*, *34*(10), 51–58. DOI: 10.3969/j.issn.1003-0077.2020.10.007

Wang, J., Shou, L., Chen, K., & Chen, G. (2020). Pyramid: A layered model for nested named entity recognition. In *Proceeding of the 58th annual meeting of the Association for Computational Linguistics* (pp. 5918–5928). ACL. DOI: 10.18653/v1/2020.acl-main.525

Wang, X., & Wan, Y. (2023). A named entity identification method for legal documents. *Computer Applications and Software*, *40*(8), 180–186. DOI: 10.3969/j.issn.1000-386x.2023.08.030

Wang, Y., Shindo, H., Matsumoto, Y., & Watanabe, T. (2021). Nested named entity recognition via explicitly excluding the influence of the best path. *In Proceedings of the 59th annual meeting of the Association for Computational Linguistics* (pp. 3547–3557). ACL. DOI: 10.18653/v1/2021.acl-long.275

Wang, Z., Wu, Y., Lei, P., & Peng, C. (2020). Named entity recognition method of Brazilian legal text based on pre-training model. *Journal of Physics: Conference Series*, *1550*(3), 032149. Advance online publication. DOI: 10.1088/1742-6596/1550/3/032149

Yan, H., Gui, T., Dai, J., Guo, Q., Zhang, Z., & Qiu, X. (2021). A unified generative framework for various NER subtasks. In *Proceedings of the 59th annual meeting of the Association for Computational Linguistics and the 11th international joint conference on natural language processing* (*Vol. 1*, pp. 5808–5822). ACL. DOI: 10.18653/v1/2021.acl-long.451

Yang, S., Liu, T., & Huang, W. (2024). Chinese legal document named entity recognition based on MVBCN-FLW. *Computer Engineering and Applications*, 1–10. http://kns.cnki.net/kcms/detail/11.2127.tp.20231228.1537.016 .html

Yu, J., Bohnet, B., & Poesio, M. (2020). Named entity recognition as dependency parsing. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics* (pp. 6470–6476). ACL. DOI: 10.18653/v1/2020.acl-main.577

Yuan, Z., Tan, C., Huang, S., & Huang, F. (2021). Fusing heterogeneous factors with triaffine mechanism for nested named entity recognition. In *Findings of the Association for Computational Linguistics: ACL 2022* (pp. 3174–3186). ACL. DOI: 10.18653/v1/2022.findings-acl.250

Zeng, L., Wang, Y., & Chen, P. (2022). Named entity recognition based on BERT and joint learning for judgment documents. *Jisuanji Yingyong*, *42*(10), 3011–3017. DOI: 10.11772/j.issn.1001-9081.2021091565

Zhang, X., & Luo, X. (2022). A Machine-Reading-Comprehension method for named entity recognition in legal documents. In M. Tanveer, S. Agarwal, S. Ozawa, A. Ekbal, & A. Jatowt (Eds.), *Communications in computer and information science: Vol. 1793. Neural information processing* (pp. 224–236). Springer. DOI: 10.1007/978-981-99-1645-0_19

Zhang, X., Luo, X., & Wu, J. (2023). A RoBERTa-GlocalPointer-based method for named entity recognition of legal documents. In *Proceedings of International Joint Conference on Neural Networks* (pp. 1–8). IEEE. DOI: 10.1109/IJCNN54540.2023.10191275

Zheng, H., Qin, B., & Xu, M. (2021). Chinese medical named entity recognition using crf-mt-adapt and ner-mrc. In *Proceedings of 2nd International Conference on Computing and Data Science* (pp. 362–365). IEEE. DOI: 10.1109/CDS52072.2021.00068

Lu Rui received her M.S. and Ph.D. degrees from Dalian Maritime University in 2005 and Northeastern University in 2009, respectively. From 2017 to 2018, she was a visiting scholar at the University of Tennessee, USA. She is currently a professor at Liaoning Police College, where she is involved in teaching and research. Her research interests include data analysis and natural language processing.

Liying Li received his Ph.D. degree from the Shenyang Institute of Automation, Chinese Academy of Sciences, in 2010, and his M.S. degree from Dalian University of Technology in 2004. From 2017 to 2018, he was a visiting scholar at Swansea University, UK. He is currently a professor at the School of Software Engineering, Dalian University of Foreign Languages. His research interests include applied machine learning, federated learning, data mining, natural language processing, and optimization methods.