

Article

Kcr-FLAT: A Chinese-Named Entity Recognition Model with Enhanced Semantic Information

Zhenrong Deng ^{1,2}, Yong Tao ², Rushi Lan ^{2,*}, Rui Yang ² and Xueyong Wang ³¹ Guangxi Key Laboratory of Images and Graphics Intelligent Processing, Guilin University of Electronic Technology, Guilin 541004, China² Nanning Research Institute, Guilin University of Electronic Technology, Guilin 541004, China³ Guilin Xintong Technology Co., Ltd., Guilin 541004, China

* Correspondence: rslan@guet.edu.cn

Abstract: The performance of Chinese-named entity recognition (NER) has improved via word enhancement or new frameworks that incorporate various types of external data. However, for Chinese NER, syntactic composition (in sentence level) and inner regularity (in character-level) have rarely been studied. Chinese characters are highly sensitive to sentential syntactic data. The same Chinese character sequence can be decomposed into different combinations of words according to how they are used and placed in the context. In addition, the same type of entities usually have the same naming rules due to the specificity of the Chinese language structure. This paper presents a Kcr-FLAT to improve the performance of Chinese NER with enhanced semantic information. Specifically, we first extract different types of syntactic data, functionalize the syntactic information by a key-value memory network (KVMN), and fuse them by attention mechanism. Then the syntactic information and lexical information are integrated by a cross-transformer. Finally, we use an inner regularity perception module to capture the internal regularity of each entity for better entity type prediction. The experimental results show that with F1 scores as the evaluation index, the proposed model obtains 96.51%, 96.81%, and 70.12% accuracy rates on MSRA, resume, and Weibo datasets, respectively.

Keywords: Chinese-named entity recognition; semantic information; syntactic information; cross-transformer; regularity perception



Citation: Deng, Z.; Tao, Y.; Lan, R.; Yang, R.; Wang, X. Kcr-FLAT: A Chinese-Named Entity Recognition Model with Enhanced Semantic Information. *Sensors* **2023**, *23*, 1771. <https://doi.org/10.3390/s23041771>

Academic Editor: Giorgio Fumera

Received: 4 January 2023

Revised: 27 January 2023

Accepted: 31 January 2023

Published: 4 February 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

With the popularity of smart mobile terminals, the number of internet users in developing countries is growing rapidly, and on 31 August, China's internet Network Information Center (CNNIC) released its 50th statistical report on the development status of China's internet. According to the report, as of June 2022, there were 1.051 billion internet users in China; a large number of China's text data will be generated when people obtain information, search for content, and interact with information on the internet. How to make full use of large amounts of Chinese text (to quickly extract and mine valuable information from it) is a popular research direction.

To improve the processing speed of Chinese text, many applications in the field of Chinese natural language processing (NLP) need the support of Chinese NER, including intelligent recommendations [1], question-answering [2], text understanding [3,4], and text generation [5]. Chinese NER is one of the most important and basic tasks in NLP, and it plays an important role in many downstream tasks. Chinese NER mainly identifies entities from raw text and classifies the detected entities into a pre-defined category, such as people names, place names, organizations, and other proper nouns [6]. For example, in Chinese, “*Dong ao hui zai bei jing shun li ju xing*. (The Winter Olympic Games were successfully held in Beijing)”, in which the named entities include “*Dong ao hui* (Winter Olympic Games)” and “*bei jing* (Beijing)”.

Compared to English NER, Chinese NER is more difficult for several reasons. First, the development of Chinese NER is relatively late compared to the development of English NER, the corpus of Chinese NER is less, which restricts the development of Chinese NER [7,8]. Second, there are spaces between English words as separators, so that preprocessing word segmentation is not necessary. However, Chinese does not have these obvious features [9], The formation of the Chinese word is flexible. A Chinese sequence can be divided into different words according to different contexts. For example, in “*huang he nan an* (south bank of the Yellow River)”, the entity is “*huang he* (Yellow River)”, but in other contexts, “*he nan* (Henan province)” is a place name as another named entity.

To address the aforementioned issues and handle the complexity of the Chinese language structure, this paper presents a Kcr-FLAT for Chinese NER. We use the Flat-Lattice transformer (FLAT) model as the base model to take advantage of its lightweight and parallel computing. Based on word enhancement, the proposed model extracts three different types of syntactic data and corresponding context features, encodes the syntactic information with their context features by a KVMN, and fuses them by attention mechanism. To solve the segmentation error introduced by lexical information, we used a cross-transformer to fully integrate lexical information and syntactic information. Finally, we explored the internal regularity of each entity to improve the recognition accuracy of the named entities through an inner regularity perception module. The main contributions of the proposed model include:

1. We propose a Chinese NER model Kcr-FLAT, which can achieve better performance in three Chinese datasets—MSRA, resume, and Weibo.
2. We extract and encode three types of syntactic data and fuse them with lexical information to solve the segmentation error.
3. We combine the internal regularity of the entity mentions by an inner regularity perception module with external information to improve the performance of the designed Kcr-FLAT model.

The remainder of this article is organized as follows. Section 2 describes the related work of our model. Section 3 introduces the specific composition structure of Kcr-FLAT. Section 4 describes the datasets, and evaluates the index and experimental settings. Section 5 presents the experimental results in detail. Finally, the article is summarized in Section 6.

2. Related Works

Ever since NER was proposed as a subtask at the sixth Message Understanding Conference (MUC-6) in 1995, this topic has attracted the attention of many experts and scholars in China and abroad, and has become a clear research task in the field of NLP. There are three main development stages of NER: rule-based NER, statistical machine learning-based NER, and deep learning-based NER.

The early and mid-term mainly involve rule-based NER and statistical machine learning-based NER. The rule-based method mainly involves naming rules manually, and then matching the characters inside the rules by custom rules to find the entities inside the text [10]; it relies heavily on manual efforts and takes a lot of time to build up the relevant knowledge base and lexicon. It is difficult to match the entities in all scenarios and it has very large scenario limitations without migration. Among the NER techniques in Chinese, Wang et al. [11] formulated named entity rules related to company names for datasets in the financial news domain, considered the structural features of company names and the contextual semantic associations in the text files, and established six knowledge bases for recognizing company names in the financial news domain. To improve the efficiency of specifying rules, scholars began to study how to formulate named entity rules through machine assistance and implemented named entity recognition through unsupervised classification of a large number of datasets [12]. Rule-based named entity recognition methods require a lot of human and material resources and specified rules have

great limitations. There are often named entity rules in a certain application field, with low generalization ability.

Statistically-based named entity recognition methods are mainly implemented by machine learning algorithms; the most frequently used in named entity recognition are maximum entropy model (ME) [13], hidden Markov model (HMM) [14], support vector machine (SVM) [15], conditional random field (CRF) [16], etc. Statistical-based named entity recognition methods require a large amount of corpus support and manual set-ups of feature templates that match the specific application scenario.

With the development of neural networks, deep learning methods have become research hotspots. For Chinese NER, various lexicon-based models have been proposed that incorporate external lexicon information and obtain better results. A typical method is Lattice-LSTM [17], which incorporates word lexicons into the character-based NER model. However, limited by the construction method, the lexical representation can only be added to the last word and the lattice structure fails to compute in parallel. To address this problem, Sui et al. [18] proposed a collaborative graph network-based lexical knowledge to achieve automatic lexicon construction through large-size automatic segmentation text pre-training and graph convolution. In the same year, Ding et al. [19] proposed a named entity recognition method based on graph neural networks and combining multiple lexicons, allowing the model to automatically learn the features of lexicons, which helps to alleviate the problem of false matches.

These methods are necessary to manually constructed lexicon resources, which are expensive and time-consuming, and the quality of the lexicon is critical to the tasks. Xu et al. [20], inspired by these works, proposed a character-level word embedding Chinese-named entity recognition model, which enables semantic information to be utilized at multiple granularities from the root, character, and word level. Ma et al. [21] improved Chinese word separation accuracy by merging lexical information into a vector representation, thus avoiding the introduction of complex sequence structures to characterize lexical information. Hu et al. [22] proposed the SLK-NER Chinese-named entity recognition model, which still faces the challenge of matching word boundary conflicts due to the simple first-order lexical knowledge that provides insufficient lexical information, and the authors proposed new insights into the second-order lexical knowledge of each character in a sentence to provide more lexical information, including semantic and word boundary features.

Several works focus on glyph structure information in Chinese NER. Meng et al. [23] proposed the Glyce network, which is the first model to apply Chinese glyph information in the field of Chinese-named entity recognition, improving the generalization ability of computer vision to text data. Wu et al. [24] extracted the radicals of Chinese characters through convolutional neural networks and then incorporated the radicals and character information into the model through a cross-transformer module and random attention mechanism to improve the accuracy of Chinese-named entity recognition. In these works, the structural components of Chinese characters have been proven to enrich the semantics of the characters, which achieved better results without using any external word embeddings and word resources. However, there are two major drawbacks of the glyph models. First, the glyph information codes independently. The interactive knowledge between the glyph and context is ignored. Second, the entity type and quantity of modern Chinese are far richer and more complex than those of ancient Chinese, it has been proven that historical texts are meaningless to NER, to some extent [25].

Recently, more Chinese characters and text features have been used in NER. Nie et al. [26] improved named entity recognition by using the syntactic information with the attentive ensemble. The RICON [27] model was published by Huawei Cloud in 2022, which enhances the prediction of entities by exploring the naming regularity within the entities. In the above work, the Chinese-named entity recognition has reached a very high accuracy rate, which also represents a more mature period for the development of the Chinese-named entity. Thus, it is more desirable to use these semantic data and effectively integrate them together.

3. Proposed Method

The overall structure of the proposed Kcr-FLAT model is shown in Figure 1, which mainly consists of three modules, i.e., the KVMN module for encoding the syntactic information with its context features, the cross-transformer module for incorporating lexical information and syntactic information, and the regularity perception module for analyzing inner regularity.

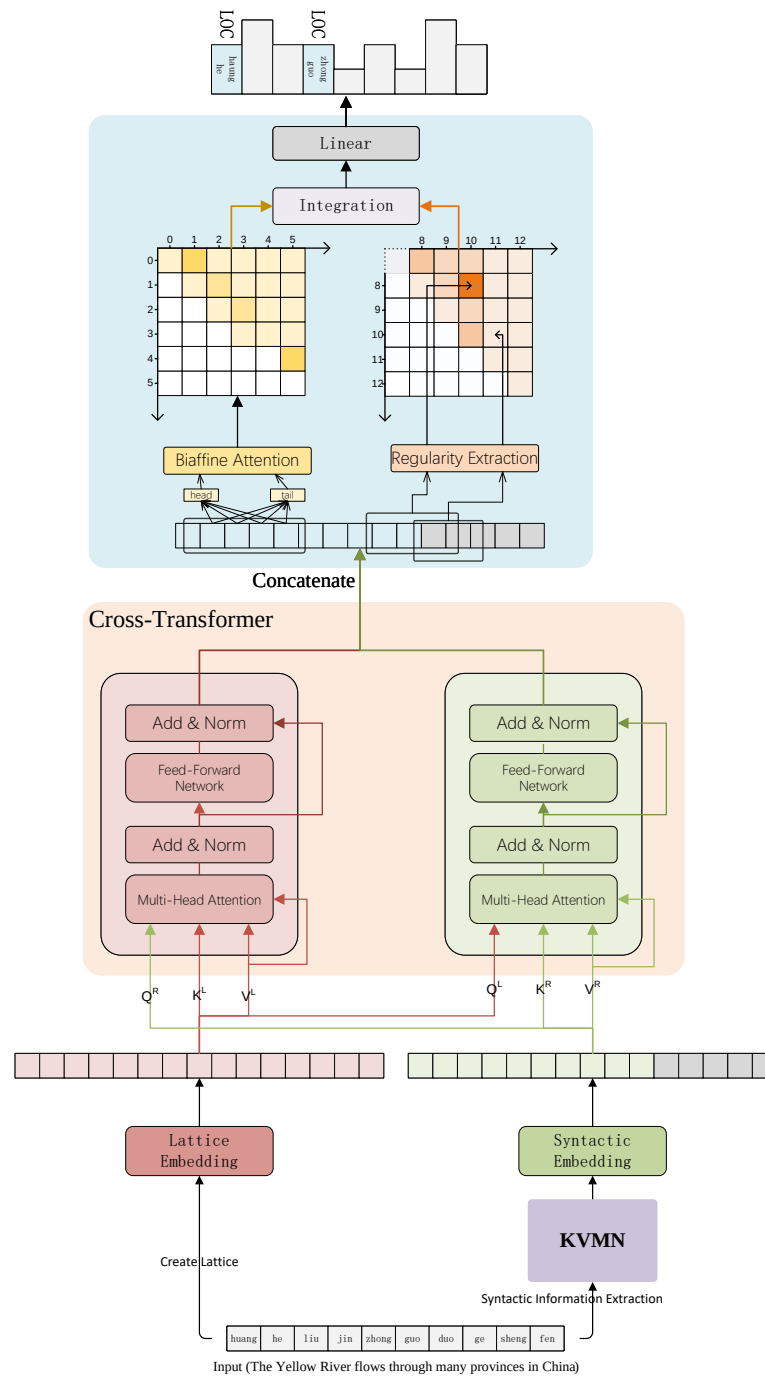


Figure 1. Overall structure of Kcr-FLAT.

3.1. KVMN for Syntactic Embedding

Considering Chinese NER tasks, the character-level NER methods are usually superior to word-level NER methods [28–31]. In this paper, the character-level NER method is used. Chinese characters may have different meanings in different contexts, and there is no separation between characters, which makes it difficult to obtain entity boundary information. Since lexical boundaries usually play crucial roles in entity boundaries; this paper uses lexical information as external information to enhance the input to obtain entity boundary information. Although the introduction of lexical information does not require splitting the input, it is still affected by segmentation errors. For example, “*huang he liu jin zhong guo duo ge sheng fen* (The Yellow River flows through many provinces in China)”, where “*huang he liu jin zhong guo* (the Yellow River flows through China)” can be decomposed into “*huang he* (Yellow River)”, “*he liu* (river)”, “*liu jin* (flow through)” and so on. However, after syntactic analysis, due to the syntactic constraint, “*huang he* (Yellow River)” and “*zhong guo* (China)” are the inherent names, and “*liu jin* (flow through)” is the only verb, so it is easy to be correctly classified as “*huang he* (Yellow River) + *liu jin* (flow through)”, eliminating the ambiguity caused by the word “*he liu* (river)”. Therefore, this paper introduces syntactic information to eliminate ambiguity; it encodes different types of contextual features and their syntactic data, including POS labels [32], syntactic constituents [33], and dependency relations [34], to improve the recognition accuracy of named entities in different contexts. To obtain the embedding of syntactic information, the input sequence X is firstly parsed using off-the-shelf NLP parsing tools to obtain three types of syntactic data and corresponding text features [26], the extraction of syntactic information is illustrated in Figure 2.

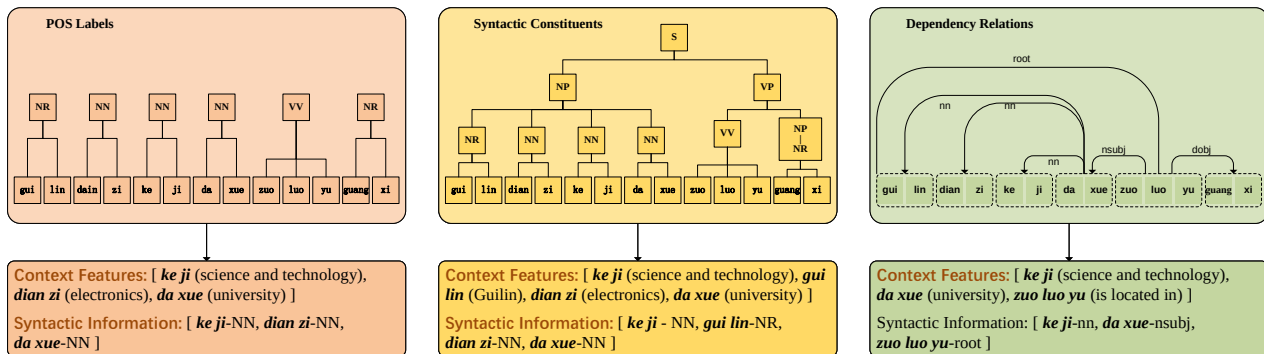


Figure 2. The extracted syntactic information in POS labels, syntactic constituents, and dependency relations for ‘ke’ (science).

Each syntactic information is encoded using KVMN and integrated by the attention mechanism to obtain the feature embedding of Chinese characters at the sentence level. The encoding of syntactic information of “*gui lin dian zi ke ji da xue zuo luo yu guang xi*. (Guilin University of Electronic Science and Technology is located in Guangxi)” is shown in Figure 3.

After parsing the input sequence X , each character x_i in X is contained in a word, using the word as the center word and mapping its context features and syntactic information into a set of keys and values denoted by $K_i^c = [k_{i,1}^c, \dots, k_{i,j}^c, \dots, k_{i,m_i}^c]$ and $V_i^c = [v_{i,1}^c, \dots, v_{i,j}^c, \dots, v_{i,m_i}^c]$, respectively. Where $c \in C = \{P, S, D\}$, P, S, D denotes a syntactic type respectively, m_i is the number of text features of x_i in syntactic type c , $k_{i,j}^c$ is the j th context feature of x_i for the syntactic type c , and $v_{i,j}^c$ is the syntactic information of $k_{i,j}^c$. Then mapping K_i^c and V_i^c into matrix form as $e_{i,j}^{k_c}$ and $v_{i,j}^{k_c}$. Therefore, $e_{i,j}^{k_c}$ refer to the context features of x_i , $v_{i,j}^{k_c}$ refer to the corresponding syntactic data. The syntactic information is calculated by

$$p_{i,j}^c = \frac{\exp(h_i \cdot e_{i,j}^{k_c})}{\sum_{j=1}^{m_i} \exp(h_i \cdot e_{i,j}^{k_c})} \quad (1)$$

where $p_{i,j}^c$ is the weight assigned to the syntactic information. Since Chinese syntax is word-based and multiple characters in the same word share the same syntactic information, For example, “*da xue* (university)”, in which “*da*” and “*xue*” have the same vector encoded by syntactic information; therefore, h_i is introduced to distinguish the representation of different characters in the same word, h_i is obtained by

$$h_i = E(x_i) \quad (2)$$

where E is the character-level embedding layer. Then, the weights $p_{i,j}^c$ were assigned to the corresponding syntactic data $v_{i,j}^{k_c}$, which is computed by

$$s_i^c = \sum_{j=1}^{m_i} p_{i,j}^c v_{i,j}^{k_c} \quad (3)$$

where s_i^c is the output of KVMN, containing the weighted syntactic information of type c . Therefore, the syntactic data are weighted and encoded according to their corresponding context features so that useful information could be leveraged accordingly.

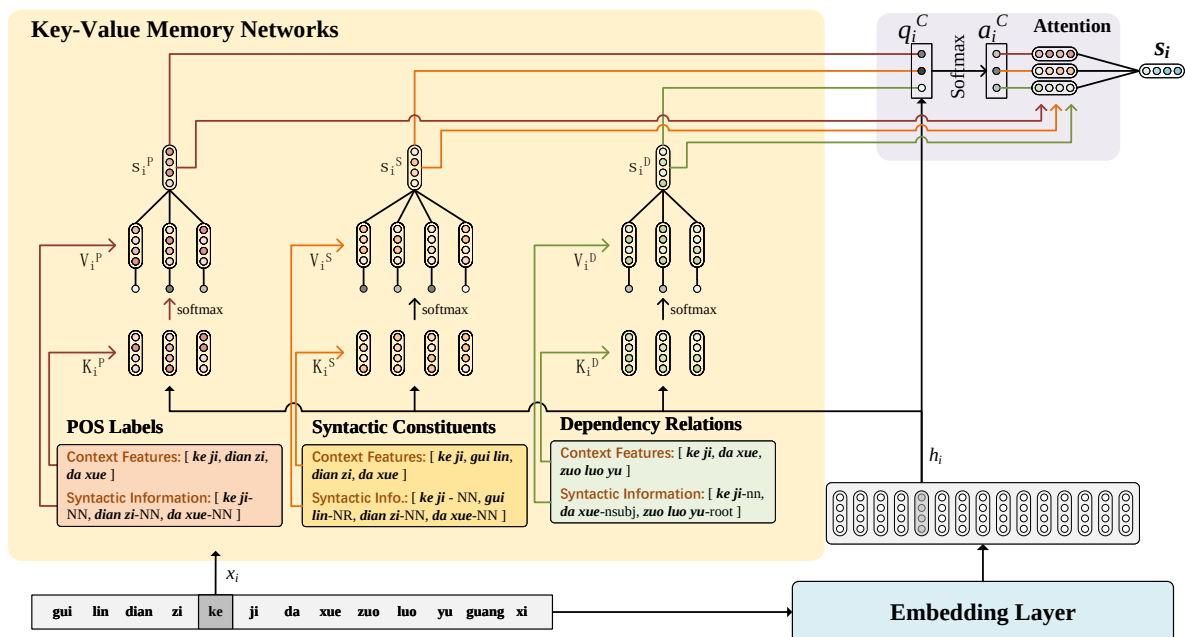


Figure 3. KVMN for encoding syntactic information.

After the syntactic information vector s_i^c is obtained, to integrate the three types of syntactic data and resolve the conflicts between different syntactic information, different weights are assigned to each syntactic information. The different syntactic information is integrated through the attention mechanism. The weight q_i^c of the syntactic information s_i^c is calculated by

$$q_i^c = \sigma(W_q^c \cdot (h_i \oplus s_i^c) + b_q^c) \quad (4)$$

where W is the trainable vector, b is the bias, $\oplus$ is the concatenate operation, and σ is the sigmoid function. The attention of the syntactic information is subsequently calculated using the *softmax* function, and the attention is calculated by

$$a_i^c = \frac{\exp(q_i^c)}{\sum_{c \in C} \exp(q_i^c)} \quad (5)$$

where a_i^c is the attention of the corresponding syntactic information of type c . Finally, the attention is assigned to the corresponding syntactic information and the three different types of syntactic data are fused into s_i by

$$s_i = \sum_{c \in C} a_i^c s_i^c \quad (6)$$

Compared with simply connecting the three kinds of syntactic information, the attention mechanism selectively leverages different features, resolving the conflicts between different syntactic data by distinguishing more important ones from others. Therefore, different types of syntactic data are comprehensively and selectively encoded into s_i .

3.2. Cross-Transformer for Semantics Fusion

Since lexical information is more concerned with the relationship between words and characters, it is easier to identify local information, such as word position and boundaries, while syntactic information is more concerned with the overall information of the sentence, and the segmentation errors introduced by lexical information are corrected by syntactic constraints. After obtaining the embedding of syntactic information, the lexical and syntactic data are fused using a cross-transformer. The cross-transformer network is shown in Figure 4.

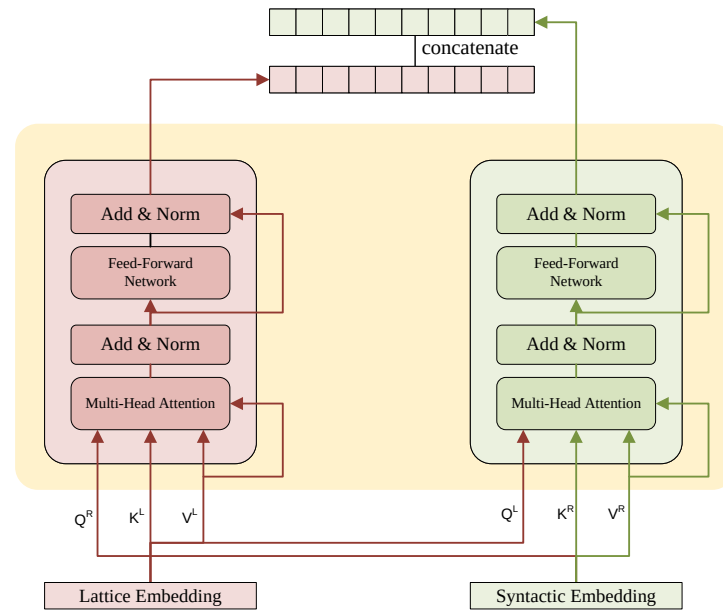


Figure 4. The cross-transformer module.

The input (Q_x^L, K_x^L, V_x^L) of the left transformer encoder is obtained from lattice embedding, which contains lexical information by linear transformation as follows:

$$[Q_x^L, K_x^L, V_x^L] = E_x^L [W_Q^L, W_K^L, W_V^L] \quad (7)$$

where E_x^L is the lattice embedding, $E_x^L \in \{x_1, \dots, x_i, \dots, x_N\}$, x_i is the lexical representation of the input and N is the length of the input sequence X , each W^L is a learnable parameter. The Q_s^R, K_s^R , and V_s^R of the left transformer encoder are obtained from the embedding of syntactic information and by linear transformation as follows:

$$[Q_s^R, K_s^R, V_s^R] = E_s^R [W_Q^R, W_K^R, W_V^R] \quad (8)$$

where E_s^R is the syntactic embedding, $E_s^R \in \{s_1, \dots, s_i, \dots, s_N\}$, s_i is the syntactic representation of the input, we keep the length of the input sequence at N to ensure that the input of the cross-transformer has the same length at both ends; each W^R is a learnable parameter. The cross-transformer consists of two transformer encoders. Each encoder is composed of self-attention and feed-forward network (FFN) layers, followed by residual connection and layer normalization. FFN is a position-wise multi-layer perceptron (MLP) with the nonlinear transformation of the semantic space. The self-attention layer is used to extract semantic-level information. The attention score of the vanilla transformer is calculated as:

$$Att(A, V) = softmax(A)V \quad (9)$$

$$A_{i,j} = \left(\frac{Q_i K_j^T}{\sqrt{d_k}} \right) \quad (10)$$

where d_k is the dimension of K . To cross the syntactic information and lattice information obtained from Equations (7) and (8), we use a variant of self-attention [35] with the relative position encoding in the FLAT model, which is computed as follows:

$$Att^L = softmax(A^R)V^L \quad (11)$$

$$Att^R = softmax(A^L)V^R \quad (12)$$

where A^R is the syntactic attention score and A^L is the lattice attention score. A^R is computed by

$$A_{i,j}^R = (Q_i^R + u^R)^T K_j^R + (Q_i^R + v^R)^T (R_{i,j}^R W_r^R) \quad (13)$$

where u^R and v^R are attention bias, W_r^R are learnable parameters. $R_{i,j}^R$ is the relative position encoding, which is computed by

$$R_{i,j}^R = ReLU(W_r(p_{h_i-h_j} \oplus p_{t_i-h_j} \oplus p_{h_i-t_j} \oplus p_{t_i-t_j})) \quad (14)$$

The relative position encoding $R_{i,j}$ is used to avoid the loss of directionality caused by the inner dot-product of the vector. Each p denotes a relative distance. The calculation of Att^L is essentially the same.

3.3. Regularity Perception Module

In the patterns of the word formation of Chinese and its language structure, there are category-to-category named patterns for entities, e.g., “XX-textitxue xiao (school)” and “XX-fa yuan (court)” usually represent place names, but also regularity within words, e.g., “ma nao he (agate river)”, “huang he (yellow river)”, “liu yang he (liuyang river)”, etc., are all in the format of “XX-he (river)” to express locations. To capture the regularity of these entities, this paper uses the regularity perception module proposed by Huawei Cloud in 2022 to analyze the internal regularity. For the NER task, there are two classical solutions, one is to handle it as a sequence-labeling task, but sequence labeling does not easily identify nested entities, e.g., “gui lin dian zi ke ji da xue (Guilin University of Electronic Science and Technology)”, which identifies the complete “Guilin University of Electronic Science and Technology” but not a place entity “gui lin (Guilin)” and a textitdian zi ke ji da xue (University of Electronic Science and Technology of China) which means a totally different university”. This problem can be effectively avoided by using the span method, which can be used to search, reorganize, and classify words on the span. In this paper, we adopt the span method to process the named entity, use the regularity perception module to capture

the internal regular features of each span $s_{i,j}$, and replace the linear and CRF at the top of the cross-transformer with the regularity perception module, as shown in Figure 5.

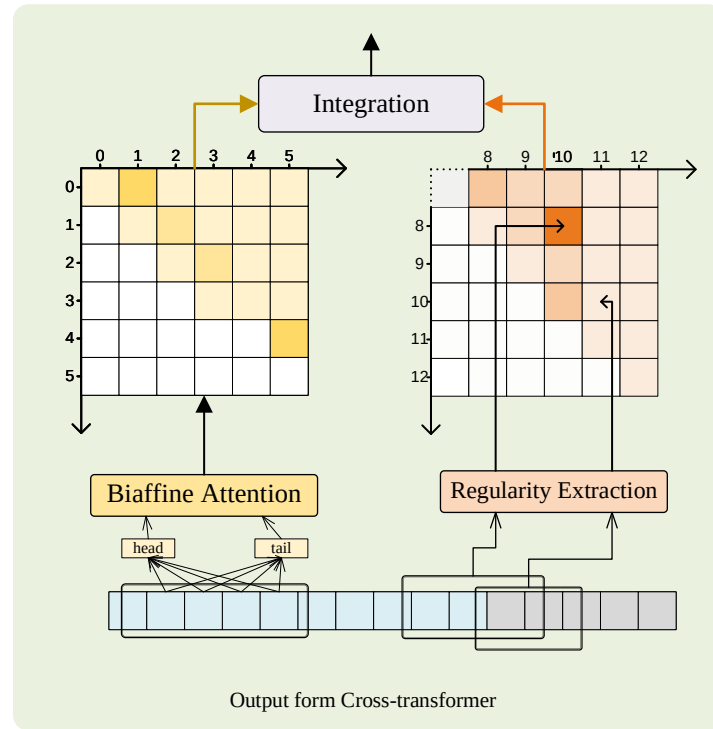


Figure 5. The regularity perception module.

In Figure 5, the regularity representation of each span is obtained by a linear attention layer calculated as follows:

$$h_{s_{i,j}}^{rec} = \sum_{t=i}^j \alpha_t \cdot h_t \quad (15)$$

where h_t is the concatenated output from cross-transformer, $t \in \{i, i+1, \dots, j\}$ refers to the index of span $s_{i,j}$, α_t is the attention score of h_t , which is calculated as follows:

$$\alpha_t = \frac{\exp(a_t)}{\sum_{k=i}^j \exp(a_k)} \quad (16)$$

$$a_t = W_{reg}^T h_t + b_{reg} \quad (17)$$

where $W_{reg}^T \in \mathbb{R}^{d \times 1}$ and $b_{reg} \in \mathbb{R}^1$ denote the learnable weight and bias, respectively. The model uses a biaffine attention mechanism to learn the interactions of head and tail features to integrate the regularity between entities into the representation of span which is computed by

$$h_{s_{i,j}}^{(span)} = h_i^T U^{(1)} h_j + (h_i \oplus h_j) U^{(2)} + b_1 \quad (18)$$

where h_i and h_j are the head and tail representations of the span $s_{i,j}$, respectively, $U^{(1)} \in \mathbb{R}^{2d \times 2d \times 2d}$ and $U^{(2)} \in \mathbb{R}^{4d \times 2d}$, integrating the regularity and span representation through a gate network:

$$g_{s_{i,j}} = \sigma \left(U^{(3)} \left[h_{s_{i,j}}^{(span)}; h_{s_{i,j}}^{(reg)} \right] + b_2 \right) \quad (19)$$

$$h_{s_{i,j}} = g_{s_{i,j}} \odot h_{s_{i,j}}^{(span)} + (1 - g_{s_{i,j}}) \odot h_{s_{i,j}}^{(reg)} \quad (20)$$

where $U^{(3)} \in \mathbb{R}^{4d \times 1}$ is the learnable weight, b_2 is the bias, σ is the *sigmoid* function, $\odot$ refers to the element-wise multiplication operation. Then, the prediction of the entity type for each span is obtained from a *softmax* linear classifier:

$$\hat{y}_{s_{i,j}} = \text{Softmax}(W_{type}^{(T)} h_{s_{i,j}} + b_3) \quad (21)$$

where $W_{type}^{(T)} \in \mathbb{R}^{2d \times c}$ is the trainable weight and b_3 is the bias. The loss function is represented by the cross entropy:

$$L_{aware} = -\frac{1}{n} \sum_{n=1}^N \sum_{i=1}^l \sum_{j=1}^l y_{s_{i,j}}^{(n)} \log(\hat{y}_{s_{i,j}}^{(n)}), i \leq j \quad (22)$$

where $\hat{y}_{s_{i,j}}^{(n)}$ is the predicted type of span, $y_{s_{i,j}}^{(n)}$ is the ground truth type of span, and N is the number of training samples in the regularity perception module.

4. Experimental Settings

In this section, we evaluate the proposed Kcr-FLAT method on three Chinese datasets. We use the span method to calculate F1-score (F1), precision (P), and recall (R) as the evaluation metrics.

4.1. Datasets

In this paper, we use three Chinese NER datasets to evaluate our model, including MSRA [36], resume [17], and Weibo [37], respectively. The MSRA dataset comes from the Third International Chinese Language Processing Bake-Off: word segmentation and named entity recognition, which mainly contains news datasets and is one of the most widely used datasets in the NER task. The resume dataset was proposed by Zhang et al.; it contains a large number of names of people, places, and companies. The Weibo dataset was proposed by Peng et al.; it contains named entities for multiple scenarios from micro-blogging and can extend the application of named entity models. Among them, the Weibo dataset has four types of entities—PER (person), ORG (organization), LOC (location), and GPE (government-wide point of entry). Resume has eight types of entities—CONT (country), EDU (education), LOC, PER, ORG, PRO (profession), RACE (race), and TITLE (job title). The MSRA dataset contains three types of entities—ORG, PER, and LOC.

4.2. Evaluation Index

The weighted summed average F1 is used as the evaluation index in this paper. F1 is calculated as:

$$F1 = \frac{2 \times P \times R}{P + R} \quad (23)$$

where P stands for the precision rate and R stands for the recall rate; P and R are calculated as follows:

$$P = \frac{TP}{TP + FP} \quad (24)$$

$$R = \frac{TP}{TP + FN} \quad (25)$$

where TP is the number of entities correctly identified by the model, FP is the number of incorrectly identified entities, $TP + FP$ represents the number of all predicted entities. FN is the number of entities that are false negatives, $TP + FN$ is the number of gold entities.

4.3. Parameter Setting

This experiment was developed via PyTorch 1.8.2, using the Python 3.6 language on Win10 operating system with the following hardware environment: Intel (R) Xeon (R) CPU E5-2698 v4 @ 2.20 GHz processor, 16 GB running memory (RAM), GPU for NVIDIA

Tesla K80, and 16 G memory. Experimental parameters are as follows: The hidden layer nodes of the FFN layer in the cross-transformer on the Weibo and resume datasets are 384, the dropout parameter is 0.3, the number of multi-head attention layer heads is 8, the dimension of each head is 16, the total number of nodes is 128, the learning rate is set to 0.0018, and the epoch is set to 50. For the MSRA dataset, the number of hidden layer nodes in the FFN layer of the cross-transformer is 480, the dropout parameter is set to 0.3, the number of heads in the multi-head attention layer is 8, the dimension of each head is 20, the total number of nodes is 160, the learning rate is set to 0.0014, and the epoch is set to 100.

5. Results

In this section, we compare the different methods applied and discuss the results. To make the experimental results more reasonable, we set up additional working methods for assessing the performance of syntactic information and inner regularity.

5.1. The Impacts of Different Improvement Strategies on the F1 Score

The three improved strategies proposed above are, respectively, calculated for the F1 of the model, and the original FLAT model is used as a reference. The results are shown in Table 1, where the baseline is the base model FLAT, KNMN is the first improved strategy added in our model for syntactic information extraction and the endcode, CT represents the cross-transformer module, RP represents the regularity perception module, Kcr-FLAT is the model proposed in this paper.

Table 1. Ablation experiments

Model	KVMN	CT	RA	MSRA	Resume	Weibo
Baseline				93.45	94.93	63.42
+KNMN	✓			95.98	96.23	69.27
+KVMN & CT	✓	✓		96.24	95.98	69.78
+RP			✓	96.14	96.02	66.98
Ours (Kcr-FLAT)	✓	✓	✓	96.51	96.81	70.12

As can be seen from Table 1, each improved strategy used in this paper improved the performance of the original FLAT model to varying degrees. Among them, adding the KVMN had the most significant improvement of the model, about 6% higher on the Weibo dataset and nearly 2% higher on MSRA and resume datasets because adding the syntactic information improves the accuracy of Chinese NER in different semantic scenarios. By adding the cross-transformer module, the performance was 0.5% higher on MSRA and Weibo datasets, but experienced a 0.25% drop on the resume dataset, as the highly formatted text in this dataset may have limited the performance of the integration of syntactic and lexical information through cross-transformer. The regularity perception module improved the model by 2% on average. The designed Kcr-FLAT Chinese-named entity recognition model was 3% higher than the original Chinese NER model FLAT on the MSRA dataset, 2% higher on the resume dataset, and about 6.5% higher on the Weibo dataset. The addition of syntactic information was fully fused with lexical information by the cross-transformer module, which greatly improved the accuracy of NER in contextually rich datasets, such as Weibo.

5.2. Comparison with Other Methods

Taking F1 as the evaluation index, the proposed model was compared with Sui et al. [18], Ding et al. [19], Glyce [23], Xu et al. [20], Ma et al. [21], SLK-NER [22], Nie et al. [26], MECT [24], RICON [27], Zhu et al. [38], Mao et al. [39], and the non-flat-lattice transformer (NFLAT) [40]. The results are shown in Table 2.

Table 2. Comparison with other Chinese-named entity recognition models.

Models	MSRA	Resume	Weibo
Sui et al. [18]	-	-	63.09
Ding et al. [19]	-	-	59.50
Glyce [23]	93.26	96.54	67.60
Xu et al. [20]	-	-	68.93
Ma et al. [21]	-	95.59	61.24
SLK-NER [22]	-	95.80	64.00
Nie et al. [26]	81.18	96.62	69.78
MECT [24]	96.24	95.98	70.43
RICON [27]	96.14	96.02	66.89
Zhu et al. [38]	96.26	96.66	72.66
Mao et al. [39]	-	96.33	70.85
NFLAT [40]	94.55	95.58	61.94
Ours (Kcr-FLAT)	96.51	96.81	70.12

Compared with Chinese-named entity recognition models in recent years, the model designed in this paper has obvious advantages in the three different datasets: MSRA, resume, and Weibo. Compared with the RICON model, the designed model does not improve significantly with the MSRA and resume datasets, but with the Weibo dataset, it improves nearly 4% due to the syntactic information encoded by KNMN in the model, which demonstrates the effectiveness of augmenting the input information by syntactic information to improve the recognition accuracy of Chinese-named entities in contextually rich texts. However, the designed model in Weibo data is 2.5% points lower than that of Zhu et al. The boundary-smoothing method of their work helped the trained NER model maintain calibration so that the generated confidence can better represent the accuracy of the prediction entity and have better generalization, thus achieving remarkable results with the Weibo dataset. For the text of variable scenes, regularity may limit the recognition accuracy of named entities in different contexts, but in specific scenarios, the addition of the regularity perception module can improve the recognition accuracy of named entities. For example, the model performs better than MECT, RICON, and Mao et al., on the MSRA and resume datasets.

5.3. Case Study

To better understand how our model brings improvement, we conducted a case study on two examples selected from the MSRA and Weibo datasets, respectively. The base model FLAT was used for the comparison and analysis with the designed Kcr-FLAT model. The Chinese-named entity recognition results are shown in the following table.

Where B, M, E represent the beginning, the middle, and the end of an entity, respectively; O shows that the character is not an entity; ORG represents the type of the entity; B-LOC represents the beginning of an entity, which type, the *location*, etc.

Where B-LOC.NAM represents the beginning of a specific location, E-LOC.NOM represents the end of a general location, etc.

As can be seen from Tables 3 and 4, the designed Chinese-named entity recognition model Kcr-FLAT is improved compared with the original model FLAT. In the example of “*huang jia hai jun fu zhu dui* (Royal Navy Auxiliary Corps)”, the recognition results of FLAT are “*huang jia hai jun* (Royal Navy)” only, and “*fu zhu dui* (Auxiliary Corps)” is ignored since it is not an entity. However, in this case, although “*huang jia hai jun* (Royal Navy)” is identified, it is not the proper entity type according to the context. In regard to syntactic information, both “*huang jia hai jun* (Royal Navys)” and “*fu zhu dui* (Auxiliary Corps)” are the same syntactic components, the weights assigned to their syntactic data are equal. The Kcr-FLAT identifies the whole “*huang jia hai jun fu zhu dui* (Royal Navy Auxiliary Corps)” as an independent entity, which is more accurate.

Table 3. Comparison of Chinese-named entity recognition results on the MSRA dataset.

MSRA	<i>Huang Jia Hai Jun Fu Zhu Dui (Royal Navy Auxiliary Corps)</i>						
Entity	huang	jia	hai	jun	fu	zhu	dui
gold label	B-ORG	M-ORG	M-ORG	M-ORG	M-ORG	M-ORG	E-ORG
Baseline (FLAT)	B-ORG	M-ORG	M-ORG	E-ORG	O	O	O
Ours (Kcr-FLAT)	B-ORG	M-ORG	M-ORG	M-ORG	M-ORG	M-ORG	E-ORG

Table 4. Comparison of Chinese-named entity recognition results on the Weibo dataset.

Weibo	<i>Long Zhi Meng Jiu Dian (Longemont Hotel)</i>				
Entity	long	zhi	meng	jiu	dian
gold label	B-LOC.NAM	M-LOC.NAM	M-LOC.NAM	M-LOC.NAM	E-LOC.NAM
Baseline(FLAT)	O	O	O	B-LOC.NOM	E-LOC.NOM
Ours(Kcr-FLAT)	B-LOC.NAM	M-LOC.NAM	M-LOC.NAM	M-LOC.NAM	E-LOC.NAM

In the example of “*long zhi meng jiu dian (longemont hotel)*”, the result of FLAT recognition is “*jiu dian (hotel)*”. Due to the limitation of the lexical method, the entity name consisting of single characters that do not form a word is hard to be identified. The proposed model Kcr-FLAT learns the regularity of “XX + *zhi* (of) + XX” due to the regularity perception module paying its highest attention score on three character “*zhi* (of)”, so that patterns, such as “*long zhi meng (longemont)*”, can be accurately identified, or “*gui guai zhi di (land of ghost)*” in other cases.

6. Conclusions

In this paper, we propose a Chinese NER model namely Kcr-FLAT. The proposed method uses a key-value memory network to encode syntactic information, a cross-transformer module to integrate the syntactic information with lexical information, and a regularity perception module to explore the internal regularity of an entity. Experimental results on three datasets demonstrate that the proposed model can effectively solve the segmentation error and improve the performance of Chinese NER. In future research, we will search for a more efficient way to extract syntactic information without other NLP tools, and consider how to integrate external lexical information and internal regularity in a better way to improve the performance of Chinese NER and extend it to other NLP tasks.

Author Contributions: Conceptualization, Z.D.; methodology, Y.T.; funding acquisition, R.L.; supervision, R.Y.; project administration, X.W.; formal analysis, X.W.; investigation, Z.D. All authors have read and agreed to the published version of the manuscript.

Funding: This work was partially supported by the Guangxi Science and Technology Project (nos. AB20238013, AD20159034), the National Natural Science Foundation of China (no. 62172120), the Open Funds from Guilin University of Electronic Technology, Guangxi Key Laboratory of Image and Graphic Intelligent Processing Project (nos. GIIP2211, GIIP2003), and the Innovation Project of GUET Graduate Education (no. 2022YCX087).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Mu, R. A survey of recommender systems based on deep learning. *IEEE Access* **2018**, *6*, 69009–69022. [CrossRef]
2. Mollá, D.; Van Zaanen, M.; Smith, D. Named entity recognition for question answering. In Proceedings of the Australasian Language Technology Workshop 2006, Sydney, Australia, 30 November–1 December 2006; pp. 51–58.
3. Wang, Y.; Sun, C.; Wu, Y.; Zhou, H.; Li, L.; Yan, J. ENPAR: Enhancing entity and entity pair representations for joint entity relation extraction. In Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume, Online, 19–23 April 2021; pp. 2877–2887.

4. Cheng, P.; Erk, K. Attending to entities for better text understanding. In Proceedings of the AAAI Conference on Artificial Intelligence, New York, NY, USA, 7–12 February 2020; Volume 34, pp. 7554–7561.
5. Iqbal, T.; Qureshi, S. The survey: Text generation models in deep learning. *J. King Saud-Univ.-Comput. Inf. Sci.* **2022**, *34*, 2515–2528. [\[CrossRef\]](#)
6. Nadeau, D.; Sekine, S. A survey of named entity recognition and classification. *Linguisticae Investig.* **2007**, *30*, 3–26. [\[CrossRef\]](#)
7. Yang, J.; Teng, Z.; Zhang, M.; Zhang, Y. Combining discrete and neural features for sequence labeling. In *Proceedings of the International Conference on Intelligent Text Processing and Computational Linguistics*; Springer: Berlin, Germany, 2016; pp. 140–154.
8. He, H.; Sun, X. F-Score Driven Max Margin Neural Network for Named Entity Recognition in Chinese Social Media. In Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 2, Short Papers, Valencia, Spain, 3–7 April 2017; pp. 713–718.
9. Zhu, P.; Cheng, D.; Yang, F.; Luo, Y.; Huang, D.; Qian, W.; Zhou, A. Improving Chinese Named Entity Recognition by Large-Scale Syntactic Dependency Graph. *IEEE/ACM Trans. Audio Speech Lang. Process.* **2022**, *30*, 979–991. [\[CrossRef\]](#)
10. Mikheev, A.; Moens, M.; Grover, C. Named Entity Recognition without Gazetteers. In Proceedings of the Ninth Conference on European Chapter of the Association for Computational Linguistics, Bergen, Norway, 8–12 June 1999; pp. 1–8.
11. Wang, N.; Ge, R.; Yuan, C.; Wong, K.; Li, W. Company name identification in Chinese financial domain. *J. Chin. Inf. Process.* **2002**, *16*, 1–6.
12. Collins, M.; Singer, Y. Unsupervised models for named entity classification. In Proceedings of the 1999 Joint SIGDAT Conference on Empirical Methods in Natural Language Processing and very Large Corpora, College Park, MD, USA, 21–22 June 1999.
13. Berger, A.; Della Pietra, S.A.; Della Pietra, V.J. A maximum entropy approach to natural language processing. *Comput. Linguist.* **1996**, *22*, 39–71.
14. Hu, W.; Tian, G.; Kang, Y.; Yuan, C.; Maybank, S. Dual sticky hierarchical Dirichlet process hidden Markov model and its application to natural language description of motions. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *40*, 2355–2373. [\[CrossRef\]](#) [\[PubMed\]](#)
15. Chen, P.H.; Lin, C.J.; Schölkopf, B. A tutorial on ν -support vector machines. *Appl. Stoch. Model. Bus. Ind.* **2005**, *21*, 111–136. [\[CrossRef\]](#)
16. Lee, C.; Hwang, Y.G.; Oh, H.J.; Lim, S.; Heo, J.; Lee, C.H.; Kim, H.J.; Wang, J.H.; Jang, M.G. Fine-grained named entity recognition using conditional random fields for question answering. In *Proceedings of the Asia Information Retrieval Symposium*; Springer: Berlin, Germany, 2006; pp. 581–587.
17. 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), Association for Computational Linguistics, Melbourne, Australia, 15–20 July 2018; pp. 1554–1564. [\[CrossRef\]](#)
18. Sui, D.; Chen, Y.; Liu, K.; Zhao, J.; Liu, S. Leverage Lexical Knowledge for Chinese Named Entity Recognition via Collaborative Graph Network. 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), Hong Kong, China, 3–7 November 2019; pp. 3830–3840.
19. Ding, R.; Xie, P.; Zhang, X.; Lu, W.; Li, L.; Si, L. A Neural Multi-digraph Model for Chinese NER with Gazetteers. In Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, Florence, Italy, 28 July–2 August 2019; pp. 1462–1467.
20. Xu, C.; Wang, F.; Han, J.; Li, C. Exploiting multiple embeddings for chinese named entity recognition. In Proceedings of the 28th ACM International Conference on Information and Knowledge Management, Beijing, China, 3–7 November 2019; pp. 2269–2272.
21. Ma, R.; Peng, M.; Zhang, Q.; Wei, Z.; Huang, X. Simplify the Usage of Lexicon in Chinese NER. In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, Association for Computational Linguistics, Online, 5–10 July 2020; pp. 5951–5960.
22. Hu, D.; Wei, L. SLK-NER: Exploiting Second-order Lexicon Knowledge for Chinese NER. In Proceedings of the SEKE, Pittsburgh, PA, USA, 1–10 July 2020.
23. Meng, Y.; Wu, W.; Wang, F.; Li, X.; Nie, P.; Yin, F.; Li, M.; Han, Q.; Sun, X.; Li, J. Glyce: Glyph-vectors for Chinese Character Representations. In *Proceedings of the Advances in Neural Information Processing Systems*; Curran Associates, Inc.: Red Hook, NY, USA, 2019; Volume 32.
24. Wu, S.; Song, X.; Feng, Z. MECT: Multi-Metadata Embedding based Cross-Transformer for Chinese 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 (Volume 1: Long Papers), Virtual Event, 1–6 August 2021; Association for Computational Linguistics: Stroudsburg, PA, USA, 2021; pp. 1529–1539.
25. Song, C.H.; Sehanobish, A. Using Chinese Glyphs for Named Entity Recognition (Student Abstract). *Proc. AAAI Conf. Artif. Intell.* **2020**, *34*, 13921–13922. [\[CrossRef\]](#)
26. Nie, Y.; Tian, Y.; Song, Y.; Ao, X.; Wan, X. Improving Named Entity Recognition with Attentive Ensemble of Syntactic Information. In *Proceedings of the Findings of the Association for Computational Linguistics: EMNLP 2020*; Association for Computational Linguistics: Stroudsburg, PA, USA, 2020; pp. 4231–4245.
27. Gu, Y.; Qu, X.; Wang, Z.; Zheng, Y.; Huai, B.; Yuan, N.J. Delving Deep into Regularity: A Simple but Effective Method for Chinese Named Entity Recognition. *arXiv* **2022**, arXiv:2204.05544.

28. Soutner, D.; Müller, L. Application of LSTM neural networks in language modelling. In *Proceedings of the International Conference on Text, Speech and Dialogue*; Springer: Berlin, Germany, 2013; pp. 105–112.
29. Xiaoyan, Z.; Ting, W.; Huowang, C. A mixed statistical model-based method for Chinese named entity recognition. *Comput. Eng. Sci.* **2006**, *28*, 135–139.
30. Wang, J.; Shou, L.; Chen, K.; Chen, G. Pyramid: A layered model for nested named entity recognition. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*; Association for Computational Linguistics: Stroudsburg, PA, USA, 2020; pp. 5918–5928.
31. Chen, C.; Kong, F. Enhancing entity boundary detection for better chinese 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 (Volume 2: Short Papers)*, Virtual Event, 1–6 August 2021; pp. 20–25.
32. McCallum, A. Efficiently Inducing Features of Conditional Random Fields. In *Proceedings of the Nineteenth Conference on Uncertainty in Artificial Intelligence*; Morgan Kaufmann Publishers Inc.: Burlington, MA, USA, 2002; pp. 403–410.
33. Li, P.H.; Dong, R.P.; Wang, Y.S.; Chou, J.C.; Ma, W.Y. Leveraging linguistic structures for named entity recognition with bidirectional recursive neural networks. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, Copenhagen, Denmark, 7–11 September 2017; pp. 2664–2669.
34. Cetoli, A.; Bragaglia, S.; O’Harney, A.; Sloan, M. Graph Convolutional Networks for Named Entity Recognition. In *Proceedings of the 16th International Workshop on Treebanks and Linguistic Theories*, Prague, Czech Republic, 23–24 January 2017; pp. 37–45.
35. Dai, Z.; Yang, Z.; Yang, Y.; Carbonell, J.G.; Le, Q.V.; Salakhutdinov, R. Transformer-XL: Attentive Language Models beyond a Fixed-Length Context. *arXiv* **2019**, arXiv:1901.02860.
36. Levow, G.A. The Third International Chinese Language Processing Bakeoff: Word Segmentation and Named Entity Recognition. In *Proceedings of the Fifth SIGHAN Workshop on Chinese Language Processing*, Association for Computational Linguistics, Sydney, Australia, 22–23 July 2006; pp. 108–117.
37. Peng, N.; Dredze, M. Named entity recognition for chinese social media with jointly trained embeddings. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, Lisbon, Portugal, 17–21 September 2015; pp. 548–554.
38. Zhu, E.; Li, J. Boundary Smoothing for Named Entity Recognition. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, Dublin, Ireland, 22–27 May 2022; Association for Computational Linguistics: Dublin, Ireland, 2022; pp. 7096–7108. [[CrossRef](#)]
39. Mao, Q.; Li, J.; Meng, K. Improving Chinese Named Entity Recognition by Search Engine Augmentation. *arXiv* **2022**, arXiv:2210.12662.
40. Wu, S.; Song, X.; Feng, Z.; Wu, X. NFLAT: Non-Flat-Lattice Transformer for Chinese Named Entity Recognition. *arXiv* **2022**, arXiv:2205.05832.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.