

Article

# Named-Entity Recognition in Sports Field Based on a Character-Level Graph Convolutional Network

Xieraili Seti <sup>1,2</sup>, Aishan Wumaier <sup>1,2</sup>, Turgen Yibulayin <sup>1,2</sup>, Diliyaer Paerhati <sup>2</sup>, Lulu Wang <sup>1,2</sup> and Alimu Saimaiti <sup>1,2,\*</sup>

<sup>1</sup> College of Information Science and Engineering, Xinjiang University, Urumqi 830046, China; 107551700945@stu.xju.edu.cn (X.S.); hasan1479@163.com (A.W.); turgun@xju.edu.cn (T.Y.); wanglulu@stu.xju.edu.cn (L.W.)

<sup>2</sup> Multilingual Information Technology Laboratory of Xinjiang University, Urumqi 830046, China; 107551701971@stu.xju.edu.cn

\* Correspondence: 107556516097@stu.xju.edu.cn; Tel.: +86-181-6065-2952

Received: 30 October 2019; Accepted: 31 December 2019; Published: 5 January 2020



**Abstract:** Traditional methods for identifying naming ignore the correlation between named entities and lose hierarchical structural information between the named entities in a given text. Although traditional named-entity methods are effective for conventional datasets that have simple structures, they are not as effective for sports texts. This paper proposes a Chinese sports text named-entity recognition method based on a character graph convolutional neural network (Char GCN) with a self-attention mechanism model. In this method, each Chinese character in the sports text is regarded as a node. The edge between the nodes is constructed using a similar character position and the character feature of the named-entity in the sports text. The internal structural information of the entity is extracted using a character map convolutional neural network. The hierarchical semantic information of the sports text is captured by the self-attention model to enhance the relationship between the named entities and capture the relevance and dependency between the characters. The conditional random fields classification function can accurately identify the named entities in the Chinese sports text. The results conducted on four datasets demonstrate that the proposed method improves the F-Score values significantly to 92.51%, 91.91%, 93.98%, and 95.01%, respectively, in comparison to the traditional naming methods.

**Keywords:** character graph convolutional network; named-entity recognition; self-attention mechanism

## 1. Introduction

Named-entity recognition refers to the identification of entities with a specific meaning. This includes the name and place associated with a person or the name of an organization from a large amount of unstructured or structured text. This research involves knowledge graphs, machine translation technology, entity relation extraction [1], and automatic question answering. To obtain a more effective translation with machine translation technology, large enterprises often extract the entities in sentences for customized processing, which increases the use of keywords and improves the quality of sentence translation. However, traditional named-entity recognition methods rely heavily on linguistic knowledge and feature engineering. These ignore the hidden information of the entities in the text, thus increasing the difficulty of named-entity recognition in the text. Therefore, using the effective features and neural network technology to improve the accuracy of the named-entity recognition in the text is a hot research topic. With the improvement in human living standards, sports have become an indispensable part of our lives. Extracting the content of sporting events is of interest

since there is a significant amount of sports information that appears on the Internet every day; hence, this is an urgent problem that needs to be solved. The identification of named entities in the sports field is an important part of information extracted from sports events, which is also the topic of this paper.

In recent years, with the development of deep learning technology, numerous deep learning methods have been applied to extract and recognize named entities. For example, Gu et al. [2] accurately recognized the complex sports events named entities in Chinese text. Gu et al. proposed a method of named-entity recognition based on the cascaded conditional random fields. Feng et al. [3] used word vector features and bidirectional long-short-term memory (Bi-LSTM) networks to obtain the correlation of text sequence tags, as well as the context semantic information of named entities. This improved the recognition accuracy of the named entities. Habibi et al. [4] used word embedding technology to extract the word vector features of the named entities in biomedical text, which replaced the manual features. Subsequently, they identified and classified automatically extracted features via deep learning technology.

To improve the recognition accuracy of named entities in Vietnamese texts, Pham et al. [5] introduced conditional random fields (CRF) and convolutional neural networks (CNNs) based on Bi-LSTM networks. Pham et al. used word and sentence vector features as input to increase the discriminability of named entities in Vietnamese text. Augenstein et al. [6] combined feature sparse representation with deep learning technology to extract and recognize user-named entities from Web text, and they explained the feasibility of this method. Unanue et al. [7] extracted the word vector features of named entities in medical text using word embedding techniques such as Word2vec. Unanue et al. further extracted the context semantic information of text with recurrent neural networks (RNNs), which achieved the accurate recognition of named entities in the text. In the process of extracting named entities from patient reports, the scarcity of their text labels poses a great challenge. To solve the problem, Lee et al. [8] introduced transfer learning in entity extraction. Lee et al. transferred the data trained on the artificial neural network model to other case reports, which achieved good test results. To dispense with manual features completely, Wang et al. [9] used the gated CNN to extract the global and local semantic information of the text. To verify the feasibility of this method further, three named-entity datasets were tested and satisfactory results were obtained. To extract the relevant named entities from unstructured texts of drug compounds, Pei et al. [10] added models with an attention mechanism to the combined framework of bidirectional short and long memory networks and CRF to enhance the weights of key features in a text. This was validated using the CHEMDNER corpus. To capture the correlation between the named entities in the text further, Cetoli et al. [11] introduced the graph CNN into the traditional framework of named-entity recognition to depict the influence of the sentence syntax for named-entity recognition.

Song et al. [12] proposed a joint learning method of Chinese vocabulary and its components based on a ladder structure network. As a result, Yan Song et al. achieved the joint learning of different features and achieved good results.

The deep learning algorithms used by the researchers cited above eliminates the errors caused by artificial participation in setting features to improve the accuracy of named-entity recognition. However, most of these algorithms are based on simple word embedding technologies such as Word2vec or Word Embedding for extracting the word vector features of the text. This not only loses the semantic information of the named entities in the text but also ignores the hierarchical information between them. To solve these problems, this paper proposes a character-level graph convolutional self-attention network (CGCN-SAN) based on the Chinese sports text named-entity recognition method. The character-level graph convolutional network (GCN) is used to extract the character features and the internal structural information of the named entities in text. The self-attention mechanism is used to capture the global semantic information of the text and to describe the hierarchical structure of the named entities in the text.

The main contributions of this paper are as follows:

The character features of the named entities in sports text are obtained using the character-level GCN. The internal structural information of the named entities in the text is further characterized using graph convolution nodes.

The self-attention mechanism model is used to capture the global semantic information of sports text to enhance the correlation between the named entities in the text and to achieve an accurate named entities recognition.

## 2. Feature Learning

Effective features play a vital role in the correct recognition of named entities in sports text. That is, the more effective information the extracted features contain, the stronger the representation ability and vice versa. Based on simple word vector techniques such as Glove and Word2vec, traditional features of character, word vector, and sentence generation correspond to low-dimensional space vectors, which are used to represent text. These methods can effectively map sports texts into low-dimensional space to increase the use of potential information in sports text. However, they not only ignore the related information existing between the named entities in sports text, but also lose the global semantic information and details. The text information is filtered when the word vector is mapped. As a result, the unimportant information is deleted and only the important information is retained. This method seems to capture significant information about the sports text; however, the information that is often lost also has a certain value. Therefore, to better capture the correlation between named entities in sports text, as well as the global semantic information of sports text and to avoid the loss of detailed information, a character-level GCN is constructed in this study. The GCN extracts the character feature information of sports text and semantically models it to improve the character representation ability. The characters in the Chinese text can also be understood as a single word. This is the basic unit of the text sentence; the character features that are captured can also be understood as syntactic information of the sports text. When extracting character features, the adjacent nodes in the GCN, as well as the individual words that make up the entity, are connected. Therefore, when the character structure of the sports text is extracted by the character-level GCN, the internal structural information of the named-entity is further captured. In the two-layer GCN used in the paper, when the information is exchanged between the layers, the upper layer information is transferred to the next layer. This forms the hierarchical relationship, namely the hierarchical structure information, between the named entities in the sports text. The specific character feature-extraction network structure is shown in Figure 1.

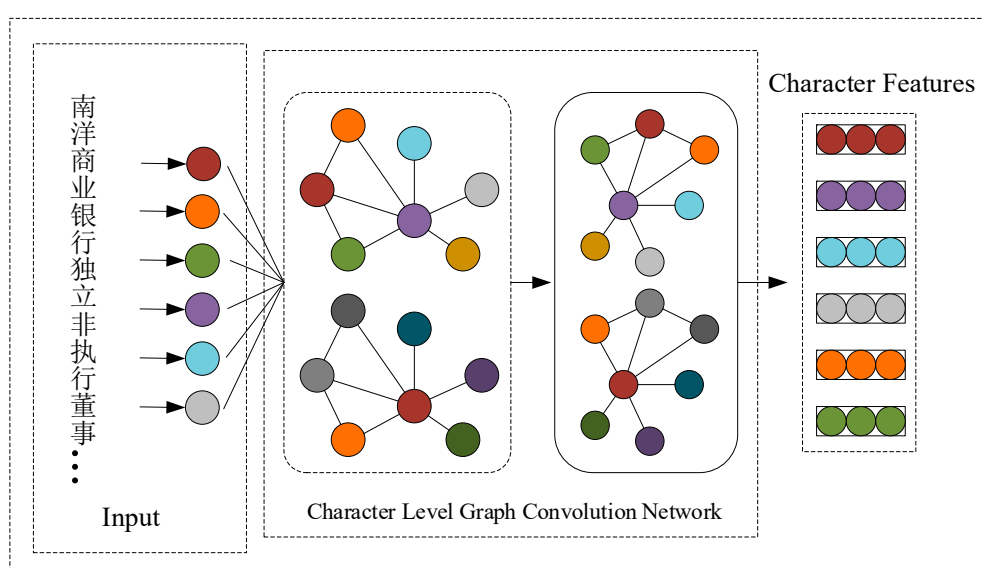


Figure 1. Network structure of the character-level graph convolution network.

In Figure 1, each node in the GCN represents a character. A total of 2545 characters are used in this paper, which contains the commonly used 2,500 basic Chinese characters, 10 numeric characters (0123456789) and 35 special symbols, such as ~! @#¥%.....&\*()——+ -=? "" ,/ ; “ ” | ~ []. One-hot is used to encode each character in the sports text. This is the input to character-level GCNs for learning how to obtain the character feature information of the sports text. The “character features” in the figure represents the character features that we have learned through the character-level GCN.

Since characters are the basic building blocks of words, and words compose texts, there is a close relationship between characters, words, and text. The traditional CNN [13] prioritizes the local characteristics of the text when capturing the character information of the text; thus, a lot of information is lost. While the long short-term memory network [14] considers the global semantic information of the text from the character information, it only expands in time; thus, it cannot effectively capture the deeper level abstract features. Each node in the character-level convolutional layer [15–17], when extracting the character features of the named-entity, transmits the feature information obtained by itself to the next adjacent node after nonlinear variation. This is then passed onto multiple nodes nearby to achieve the accumulation of character information. The node itself can self-loop; thus, the internal structural information can be captured further through the node itself. The hierarchical structural information between the entity and the non-entity in the text can be acquired from the transfer between the convolutional layers. Finally, through the above methods, the accurate acquisition of character information, internal structural information, and hierarchical structural information in the sports text can be realized. The specific calculation steps are as follows:

(1) To capture the character information of the named-entity and the association between the characters in the text, character similarity  $S_{ij}^c$  is used to construct the edge between the character nodes. If the value of  $S_{ij}^c$  is greater than 0, an edge is added to the character nodes  $i$  and  $j$ , then the weight of the edge is 1. If the value of  $S_{ij}^c$  is less than or equal to 0, the character nodes  $i$  and  $j$  are not edge-connected. The calculation is presented in Equation (1):

$$A_{ij} = \begin{cases} 1, S_{ij}^c > 0 \\ 0, S_{ij}^c \leq 0, i \neq j \end{cases} \quad (1)$$

In Equation (1),  $A_{ij}$  represents an adjacency matrix of the character nodes  $i$  and  $j$ .

(2) The character vectors of nodes  $i$  and  $j$  are captured by Word2vec and are denoted as  $C_i$  and  $C_j$ , respectively. The similarity between nodes  $i$  and  $j$  is calculated by using the cosine similarity equation. The specific equation  $S_{ij}^c$  is depicted in Equation (2):

$$S_{ij}^c = \cos(\theta) = \frac{C_i \cdot C_j}{\|C_i\| \|C_j\|} \quad (2)$$

In Equation (2),  $\theta$  represents the angle between the vectors  $C_i$  and  $C_j$ .

(3) From the above steps (1) and (2), a character graph structure is obtained, and the form satisfies  $G = (V, E)$ , where  $V$  and  $E$  respectively represent a set of character nodes and a set of edges between nodes. The matrix input into the character-level GCNs is  $X \in R^{N \times M}$ , where  $N$  represents the number of character nodes and  $M$  represents the feature vector dimension of the node.

The input and training process of the text in the graph convolution layer is demonstrated as follows:

$$\begin{cases} X = h^{(0)} \\ h^{(l+1)} = f(h^{(l)}, A) \end{cases} \quad (3)$$

where  $h$  represents the hidden layers, and  $l$  signifies the number of hidden layers.

The optimization process of the node matrix and the adjacency matrix of the graph convolution layer is provided as follows:

$$\begin{cases} h^{(l+1)} = f(h^{(l)}, A) \\ f(h^{(l)}, A) = \sigma(Ah^{(l)}\omega^{(l)}) \end{cases} \quad (4)$$

where  $\sigma$  and  $\omega$  represent the weight of the activation function and the upper hidden layer, respectively.

Considering that not all of the named entities in the dataset of sports text are labeled—that is, the classification label can only be used for a part of the nodes—the Laplacian regularization loss function [18] is introduced into the graph convolution layer. In this way, the label information of the node can be transmitted to the adjacent node, the feature information of the adjacent node can be pooled, and the internal structure information of the node can be captured. The Laplace regularization function is calculated as follows:

$$\begin{cases} \delta = \delta_0 + \lambda\delta_{reg} \\ \delta_{reg} = \sum_{i,j} A_{ij} \|f(x_i) - f(x_j)\|^2 = f(x)^T f_{\Delta}(x) \end{cases} \quad (5)$$

where  $\delta_0$  represents the supervision loss of some classification labels,  $f(\cdot)$  represents the differentiable function of the GCN,  $\lambda$  is a weighted factor of a loss function, and  $x$  is the eigenvector matrix of the nodes of the graph.

Finally, the hierarchical relationship between the named entities and non-entities is obtained. This further enhances the feature information representation ability.

(4) Finally, through the character-level GCN, we can capture the character information of the named-entity in the text, which is denoted as  $X^C$ .

### 3. Methodology

In the joint extraction of entity relations, the sequence labeling method is generally used to jointly extract and identify the entity relationships in the text. Therefore, to compensate for the deficiency of the description of the text semantic information by the features, this paper transforms the problem of named-entity recognition into a problem of sequence annotation. Consider the following classifications: the initial entity unit is labeled as B; the internal entity unit is I; other non-entity words are designated as O; the name of an athlete is SPER; a team is referred to as Steam; and an organization is referred to as Sport organization (SORG). For instance, Yao Ming is B-I-SPER, the organizer of the event is B-I-SORG, which represents the Chinese Basketball Association, and the team name is B-I-S, which is the Chinese Basketball Team. Taking ResumeNER as an experimental sample, the annotation result of named-entity of ResumeNER is given, as shown in Figure 2.

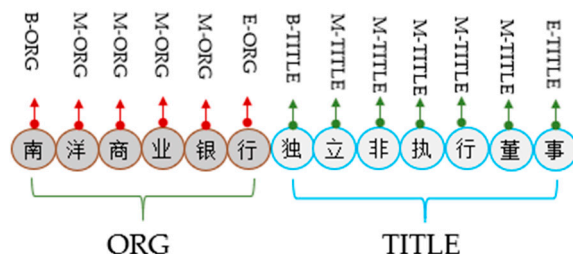


Figure 2. Named-entity annotation results for ResumeNER.

In Figure 2, “B” represents the start unit of the named-entity; “M” represents the middle unit of the named-entity; “E” represents the end unit of the named-entity; “ORG” represents the organization; “TITLE” represents the name class entity.

To further reduce the accumulated errors in the transmission of semantic information between layers, this paper adopts the self-attention model [19,20] to capture the spatial relationship between long-distance named entities in sports text. Because of the use of the transformer framework in

encoding and decoding conversions, the self-attention mechanism model not only effectively solves the long-distance dependency problem of cyclic neural networks (RNN) [21,22], but also improves the overall operation efficiency of the model. However, the correlation between the adjacent named entities is independent of each location when extracting the semantic information of the text is ignored. Therefore, a gated bidirectional long-term and short-term memory network is used based on the model of the self-attention mechanism. The detailed network structure is shown in Figure 3. The character features in the figure are the character information extracted by the character-level GCN. Word features are the word vector information captured by Word2vec, that is,  $X^W$ .

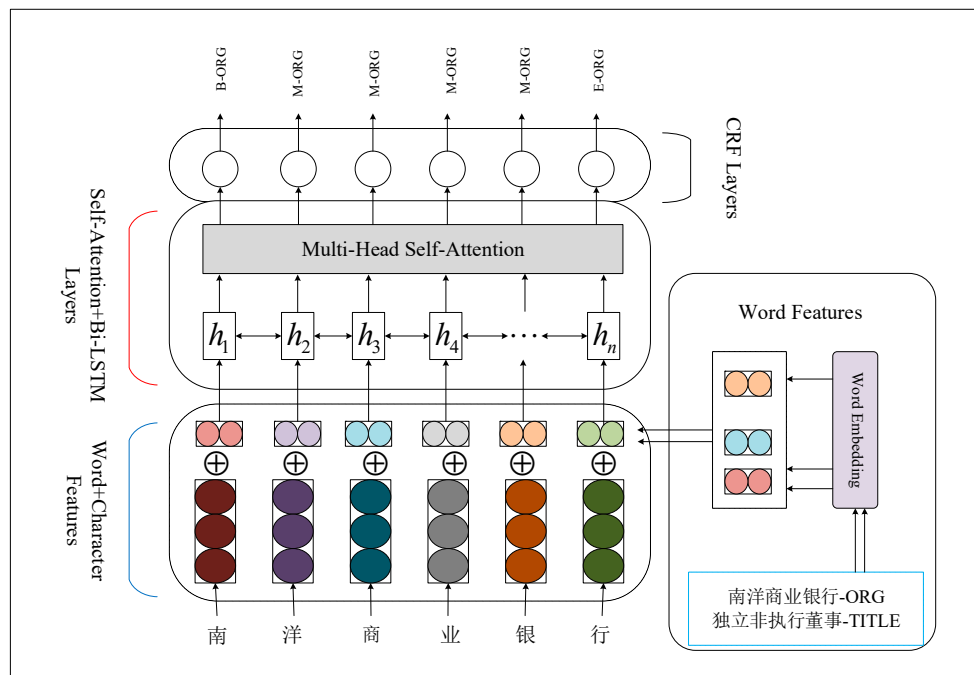


Figure 3. Network structure.

In Figure 3, the network framework is divided into three parts. These include the feature representation layer (character feature and word vector feature), the Bi-LSTM layer of self-attention, and the output layer of the CRF. The example “南洋商业银 行独立非执行董事 (means: nanyang commercial bank independent non-executive director)” in the figure is from the ResumeNER benchmark data set; “-ORG” represents the named-entity related to the organization. For the character entities in “-ORG,” we can respectively represent “B-ORG,” “M-ORG,” and “E-ORG,” where “B” represents the initial unit of the entity, “M” represents the intermediate unit of the entity, and “E” represents the end unit of the entity. “-TITLE” represents the named-entity associated with a TITLE.

The feature presentation layer combines the character information with the word vector information as  $\tilde{X}^{C+W}$ . The specific calculation is shown in Equation (6):

$$\tilde{X}^{C+W} = X^C \oplus X^W \tag{6}$$

In Equation (6),  $\oplus$  represents the series operation. For the Bi-LSTM layer of self-attention, the character and word vector features  $\tilde{X}^{C+W}$  of the named entities are inputted into a bidirectional long short-term memory network. This can characterize the dependencies between the long-term named entities in the text further; the relevant encoding and decoding operations are performed. The transformer framework [23,24], with the self-attention mechanism, is used in the process of the coding-decoding conversion. This can improve the overall operation efficiency of the model while



avoiding the loss of the location information of the named entities. The specific calculation of the input Bi-LSTM layer is described with Equation (7):

$$h_t^{\leftrightarrow} = LSTM^{\leftrightarrow}(w_t, h_{t-1}^{\leftrightarrow}) \tag{7}$$

where:  $h_t^{\leftrightarrow}$  represents the hidden layer of the bidirectional long-term and short-term memory network at time  $t$ ;  $LSTM^{\leftrightarrow}$  represents a bidirectional network of the long short-term memory; and  $w_t$  represents the feature vector input at time  $t$ . When  $t = 0, h_t^{\leftrightarrow} = \widetilde{X}^{C+W}$ .

$LSTM^{\leftrightarrow}$  in Equation (7) consists of a plurality of gate structures and memory cells [25,26]. The equations of each gate and memory unit are given as follows:

$$\begin{cases} i_t = \lambda[w_i(x_t + h_{t-1} + c_{t-1}) + b_i] \\ f_t = \lambda[w_f(x_t + h_{t-1} + c_{t-1}) + b_f] \\ z_t = \tanh(w_{\omega c}x_t + w_{hc}h_{t-1} + b_c) \\ c_t = f_t c_{t-1} + i_t z_t \\ o_t = \lambda[w_o(x_t + h_{t-1} + c_{t-1}) + b_o] \end{cases} \tag{8}$$

where the input gate, forget gate, output gate, offset, and the memory unit are represented as  $i, f, o, b,$  and  $c,$  respectively.

In order to further capture the correlations between the characters in the text, between the characters and the named entities, and between the entity character locations, a multi-head self-attention mechanism was developed in the Bi-LSTM layer [27,28]. This was done to enforce the dependency between the characters and the words. First, for the single-head self-attention mechanism, the expression form of the time series  $h_t^{\leftrightarrow}$  of  $t$  is shown in Equation (9):

$$Attention(h_t^{\leftrightarrow}) = Attention(Q, K, V) = SoftMax\left(\frac{W_Q^T h_t^{\leftrightarrow} (W_K^T h_t^{\leftrightarrow})^T}{\sqrt{d_k}}\right) W_V^T h_t^{\leftrightarrow} \tag{9}$$

where  $Q, K,$  and  $V$  represent the query, key, and value vectors. The calculation of the multi-head self-attention is shown in Equation (10):

$$\begin{cases} MultiHead(h_t^{\leftrightarrow}) = Concat(head_1, head_2, \dots, head_n) W^O \\ head_i = Attention(QW_i^Q, KW_i^K, VW_i^V) \end{cases} \tag{10}$$

where  $W_i^Q, W_i^K, W_i^V,$  and  $W^O$  represent the relevant weight matrix.

CRF layer: Considering the dependency between the adjacent tags, we used the CRF to predict the probability of the tags [29]. The probability prediction value of the tag sequence  $y = \{y_1, y_2, \dots, y_n\}$  in the sentence is shown in Equation (11):

$$P(y/s) = \frac{\exp(\sum_{i=1}^n (F(y_i) + L(y_{i-1}, y_i)))}{\sum_{y' \in \ell(s)} \exp(\sum_{i=1}^n (F(y'_i) + L(y'_{i-1}, y'_i)))} \tag{11}$$

where  $\ell(s)$  represents the possible value of a tag  $s$  in a sequence of sentences;  $F(y_i)$  represents the score of  $y_i$  and  $L(y_{i-1}, y_i)$ .

In order to reduce the over-fitting phenomenon in the learning process of the model and to confirm the tag sequence  $y$  with the highest score in the text sentences, the regularized loss function was applied [30]. The specific calculation for this is shown in Equation (12):

$$J_{Loss}(\Theta) = -\sum_{i=1}^n \log(P(y_i/s_i)) + \frac{\lambda}{2} \|\Theta\|^2 \quad (12)$$

where we used the regularized loss function,  $L_2$ .

In summary, the proposed framework can encode and decode the named entities in the text according to the contextual semantic information. It can also further capture the relationship between the long-distance named entities in sports text. Finally, complete accurate identification of the named entities was obtained.

## 4. Experimental Evaluation

### 4.1. Datasets

SportsNER: This is a custom dataset that primarily consists of information collected from major Chinese sports news websites. To verify the feasibility of the model, in this study, we used the data of 10,000 sports texts for the experimental verification, including 10 types of entities, such as the sporting event names, event levels, stadium names, team names, player names, time, and results. The ratio of the training set and the testing set was 6:4. The verification set was randomly extracted as 10% from the training set. To further ensure the accuracy of the experimental results, the P-, R-, F-Score, and Loss values were used to evaluate the experimental results to verify the representation ability of the GCNs and the validity of the features for the named-entity recognition framework. The data annotation specifications are listed in Table 1.

**Table 1.** Data annotation specifications using SportsNER.

| Entity Category                 | Parameter                             |
|---------------------------------|---------------------------------------|
| Name of competition             | Sport competition (SCOM)              |
| Team name                       | Sport team (STeam)                    |
| place name                      | Sport LOC (SLOC)                      |
| Name of athlete Athletes duties | Sport PER (SPER) Sport job name (SJM) |
| Competition level               | Sport level (SLevel)                  |
| Competition ranking             | Sport name (SN)                       |
| Match time                      | Sport time (STime)                    |
| Report to the media             | Sport media (SM)                      |
| Organizers                      | Sport organization (SORG)             |

Bakeoff-3 [31]: This is a publicly published named-entity identification library for MSRA. It contains 46,346 entities for training and 4365 entities for the test datasets. It has three types of named entities (person, location, and organization). To better verify the identification algorithm proposed in this paper, we used the cross-validation method. To this end, we randomly selected 4365 items in the training set as the verification set.

OntoNotes [32]: The data set contains 15.7K entities for training, 4.3K entities for testing, and 4.3K entities for verification. It mainly includes 18 kinds of named entities in the field of news (Chinese).

ResumeNER [33]: There are 13,438 entities for training, 1630 for testing, and 1497 for verification. There are eight types of entities, including country, location, personal name, and profession.

To better test and verify the experiment on the extraction of the character features and model training, the original text was pre-processed before the experiment. The number of nodes in Table 2 is the sum of the text, characters, and words. The statistics of the dataset are summarized in Table 2.



**Table 2.** Summary of the statistics of different datasets.

| Datasets  | Training | Testing | Valid | Nodes  | All    | Classes |
|-----------|----------|---------|-------|--------|--------|---------|
| Sports    | 6000     | 4000    | 600   | 12,535 | 10,000 | 10      |
| Bakeoff-3 | 46,346   | 4365    | 4365  | 80,718 | 55,076 | 3       |
| OntoNotes | 15,700   | 4300    | 4300  | 30,756 | 24,300 | 18      |
| ResumeNER | 13,438   | 1630    | 1497  | 21,569 | 16,565 | 8       |

## 4.2. Experimental Results and Discussion

### 4.2.1. Experimental Environment and Algorithm Parameters

To ensure the smooth progress of the experiment, an experimental environment with the specifications listed in Table 3 was implemented.

**Table 3.** Experimental environment specifications.

| Configuration Item    | Parameter                |
|-----------------------|--------------------------|
| Operating system      | Windows 10               |
| GPU                   | GTX 1050                 |
| Python                | 3.6                      |
| Development tools     | Visual Studio 2017       |
| Deep learning library | NumPy, keras, TensorFlow |

Reasonable parameter settings play an important role in the overall performance of the named-entity recognition framework. Otherwise, not only will it affect the effective use of the features by the recognition framework but it will also increase the deduplication of the named-entity extraction and recognition. Therefore, the parameters of the algorithm need to be determined depending on the actual situation.

This was considered with the character graph convolution layer as the 1–3 layer, the Bi-LSTM layer as the 1–3 layer, the learning rate of 0.001, a bit loss rate of 0.25, and the optimization function as Adam. The initial settings for the parameters are listed in Table 4.

**Table 4.** Parameter settings of the named-entity recognition framework.

| Parameter Name               | Parameter Value |
|------------------------------|-----------------|
| Number of nodes              | 28, 12          |
| Number of convolution layers | 1–3             |
| Bi-LSTM Layers               | 1–3             |
| Learning rate                | 0.001           |
| Dropout                      | 0.25            |
| Optimization function        | Adam            |

To effectively alleviate the phenomenon of the gradient disappearance and the explosion in the process of training, this study used the Adam optimization function to optimize the algorithm and to conduct 200-iteration training. If the verification loss did not decrease after 10 iterations, the training process was stopped. For benchmark models that use pre-trained word embedding, the 150-dimensional Word2vec word embedding was used as the evaluation criteria.

The F-score value as the evaluation criteria was calculated as follows:

$$F - Score = \frac{2 \times P \times R}{P + R} \quad (13)$$

In Equation (13), the accuracy rate and the recall rate are respectively expressed.

#### 4.2.2. Impact of Features on Identification Framework Performance

Different features can reflect different attributes of the named entities in sports text from different perspectives. Therefore, to verify that the features proposed in this paper have a stronger representational ability, based on the named-entity recognition framework and the character graph convolutional network features, word vector features, and character convolution features were compared. The experimental results are summarized in Table 5.

**Table 5.** Experimental results.

| Dataset   | Features        | P (%)        | R (%)        | F-Score (%)  | Loss            | Time (s)     |
|-----------|-----------------|--------------|--------------|--------------|-----------------|--------------|
| SportsNER | Word vector     | 82.95        | 84.31        | 83.62        | 0.07241         | 15.36        |
|           | Char CNN        | 85.63        | 87.12        | 86.36        | 0.06815         | 17.49        |
|           | <b>Char GCN</b> | <b>90.73</b> | <b>94.36</b> | <b>92.51</b> | <b>0.04102</b>  | <b>9.33</b>  |
| Bakeoff-3 | Word vector     | 87.49        | 88.12        | 87.80        | 0.06134         | 27.12        |
|           | Char CNN        | 88.37        | 90.51        | 89.43        | 0.05944         | 29.54        |
|           | <b>Char GCN</b> | <b>91.09</b> | <b>92.74</b> | <b>91.91</b> | <b>0.03438</b>  | <b>12.97</b> |
| OntoNotes | Word vector     | 84.15        | 79.94        | 81.99        | 0.107413        | 19.36        |
|           | Char CNN        | 90.97        | 89.59        | 90.27        | 0.094297        | 12.77        |
|           | <b>Char GCN</b> | <b>94.89</b> | <b>93.08</b> | <b>93.98</b> | <b>0.085104</b> | <b>13.49</b> |
| ResumeNER | Word vector     | 93.55        | 93.12        | 93.34        | 0.064190        | 10.04        |
|           | Char CNN        | 94.77        | 94.30        | 94.53        | 0.057016        | 12.33        |
|           | <b>Char GCN</b> | <b>95.24</b> | <b>94.76</b> | <b>95.01</b> | <b>0.042131</b> | <b>14.09</b> |

In Table 5, CharCNN represents the character features extracted by the CNN network. Meanwhile, the character-level GCN represents the character features captured by the algorithm proposed in this paper. All of them have self-attention with Bi-LSTM and CRF. As shown in Table 5, the proposed method works well for four datasets. Under the same named-entity recognition framework, the recognition effect of the feature presented herein is the best. This is because when the named entities in a sports text passes through the character graph convolution neural network, they have effective character characteristics and a hierarchical relationship for the information. In addition, these also share and accumulate the graph node information in the graph convolution layer. In this manner, the correlation between the word vector feature and the named-entity is obtained. The convolutional character feature achieves good recognition results. However, when the CNN extracts character information, only local wandering is performed in the text, which causes the extracted character features to lose considerable detail. As a result, the effect of its recognition is lower. To be specific, the F-Score values of the character CNN is 6.15%, 2.48%, 3.71%, and 0.48% lower than the F-Score values of the feature proposed in this paper.

The representation ability of the word vector is the worst. Word vector is essentially a word clustering method. As a result, the words in the sports text are converted into the corresponding vector. Although a strong local correlation between the adjacent entities in the sports text occurs through this method, considerable global information is ignored.

As indicated, the 0.04102 loss value of the feature proposed in this paper is the lowest. This represents reductions of 0.03139 and 0.02713 in comparison to the other two. The proposed feature has the same effect on the bakeoff-3 public dataset. This demonstrates that the feature presented in this paper can prevent over-fitting.

Time in Table 5 denotes the period of one iteration, which is 9.33. This helps to reduce the time by almost 50% in comparison to the other two features.

#### 4.2.3. Comparison with Other Identification Frameworks

To verify the feasibility and accuracy of the proposed CGCN-SAN recognition framework, experiments were conducted based on the convolution feature of the text symbol graph. It was

compared with shallow machine learning and traditional named-entity recognition frameworks. The experimental results are listed in Table 6.

**Table 6.** Comparison with shallow machine learning and traditional named-entity recognition models.

| Framework  | P (%) | R (%) | F-Score (%) | Loss   | Time (s) |
|------------|-------|-------|-------------|--------|----------|
| SVM        | 80.42 | 82.73 | 81.56       | —      | 43.99    |
| CNN        | 84.17 | 86.33 | 85.24       | 0.4789 | 12.11    |
| Bi-LSTM    | 87.85 | 89.41 | 88.32       | 0.5412 | 35.78    |
| LSTM       | 86.19 | 89.02 | 87.58       | 0.6231 | 34.33    |
| Attention  | 85.62 | 87.09 | 86.34       | 0.4314 | 39.52    |
| BiLSTM_CRF | 89.32 | 91.57 | 90.43       | 0.1255 | 27.19    |
| CharRNN    | 89.82 | 91.87 | 90.83       | 0.1201 | 29.57    |
| CharCNN    | 90.10 | 92.74 | 91.41       | 0.1019 | 24.55    |
| Our Model  | 90.73 | 94.36 | 92.51       | 0.0410 | 9.33     |

The results in Table 6 (obtained for the sports dataset) demonstrate that under the condition of the convolution feature of the same character, the support vector machine (SVM) has the worst recognition ability with an F-Score value of 81.56%. Because SVM is a shallow machine learning model, it fails to capture deeper abstract features in feature learning.

Bi-LSTM-CRF is the best of the traditional named-entity recognition frameworks, with an F-Score of 90.43. In this model, bidirectional long short-term memory network solves the problem of the dependency between long-distance named entities in the text. In addition, it obtains the location information of the named entities text and the global semantic information in the sports text. Simultaneously, owing to the existence of the conditional random field, the sports text sequence has been decoded, which further enhances the recognition ability of the framework. However, CRF does not fully consider the contextual semantic information of sports text. Consequently, its recognition effect is 2.08% lower than the recognition framework proposed in this paper. Compared with the shallow SVM, the CNN captures the deep features of the sports text but is restricted to the local features; thus, failing to achieve the ideal recognition effect. Bi-LSTM and LSTM not only mines the deep features of the named entities in the sports text but also solves the problem of long-time dependence and obtains all of the semantic information. Although the attention mechanism has achieved a good recognition effect, the result is lower than the Bi-LSTM and the LSTM neural networks. This is because of its independence for each location when extracting deep information.

The named-entity recognition framework proposed in this paper not only effectively excavates the deep abstract features and the global semantic information of the sports text but also captures the key information of named entities in the text by introducing the self-attention mechanism. Furthermore, the relationship between the hierarchical structural information of the character graph convolution feature and the named-entity is demonstrated. Therefore, the recognition is better than the other traditional named-entity recognition framework, whose F-Score is 92.51%.

Table 6 indicates that the named-entity recognition framework proposed in this paper requires the least amount of iterations and minimizes the loss value. Although the LSTM, Bi-LSTM, and Bi-LSTM-CRF models have achieved good results in terms of recognition, the loss value and time efficiency are not ideal. The time series model can solve the problem of the long time series dependency; however, when the model parameter quantity increases, it is easy to overfit.

Table 6 further indicates that the F-Score values obtained by charCNN and charRNN are 91.41% and 90.83%, respectively; the loss values are 0.1019 and 0.1201, respectively. When charCNN extracts the character information, it uses the CNN to capture the character information in the text. Because the CNN considers the local characteristics of the text, which can also be understood as the spatial characteristic, the captured global semantic information of the text would be missing. Further, the charRNN focuses on the temporality of the text. In other words, charRNN captures the global semantic

information well, but ignores the spatial feature information of the text. Therefore, the two methods above fail to surpass the accuracy of the model proposed in this paper on our custom sports text.

#### 4.2.4. Comparison with Other Studies

To verify the feasibility of the proposed framework and the feature-extraction method, the experimental results were compared with those of other researchers. The comparison results are summarized in Table 7.

**Table 7.** Comparison of the proposed model with other models.

| Framework             | Features |   |   |    | SportNER     | Bakeoff-3    | OntoNotes    | ResumeNER    |
|-----------------------|----------|---|---|----|--------------|--------------|--------------|--------------|
|                       | W        | S | C | HS | F (%)        | F (%)        | F (%)        | F (%)        |
| Li et al. [26]        | √        | √ | – | –  | 90.94        | –            | –            | –            |
| Lample et al. [34]    | –        | √ | – | –  | 84.33        | 86.26        | –            | –            |
| Xie et al. [1]        | –        | – | √ | –  | 83.85        | –            | –            | –            |
| Yu et al. [35]        | √        | √ | – | –  | 85.07        | –            | –            | –            |
| Greenberg et al. [36] | √        | √ | √ | –  | 91.67        | –            | –            | –            |
| Wu et al. [37]        | √        | √ | – | –  | 87.15        | 89.42        | –            | –            |
| Bekouli et al. [38]   | √        | √ | – | –  | 89.49        | 80.79        | –            | –            |
| Yue Zhang [33]        | √        | √ | √ | –  | 92.07        | 90.57        | 93.18        | 94.46        |
| Our Model             | √        | √ | √ | √  | <b>92.51</b> | <b>91.91</b> | <b>93.98</b> | <b>95.01</b> |

In Table 7, “W” represents the word vector feature, “s” represents the sentence feature, “C” represents the character feature, and “HS” represents hierarchical semantic information.

To verify the generalization ability of the algorithm proposed in this paper, we reproduced the various methods proposed in the literature while keeping the relevant parameters consistent with those presented in the corresponding original paper. The methods were compared for both datasets. It can be seen from Table 7 that the method proposed by Xie et al. [1] only uses the character features obtained by the CNN and ignores many details. As a result, its recognition result is the worst with an F-Score value of 83.8%. The method proposed by Yu et al. [35] adopts a common representation of multiple features to reflect and describe the named entities in the text from multiple perspectives. This enables it to be more effective than the others with an F-Score value of 91.67%. However, in comparison to the proposed model, it ignores the hierarchical relationship between the named entities and the internal structural information of the entities in the text; thus, reducing the F-Score by 0.84%. Some of the methods, while effective for a conventional dataset, performed poorly on the sports dataset. The effects of the method proposed by Lample et al. [34] and Bekoulis et al. [38] are surprisingly opposite on the two datasets. Only one or two features are used to describe the named entities in the methods proposed by Li et al. [26], Lample et al. [34], and Yu et al. [35], which appear to be relatively simple. That is, they ignore many details and potentially deep semantic information. Therefore, they cannot comprehensively and effectively represent the content of the named entities and the correlation between the entities in the model. Table 7 indicates that the F-Score of our experimental method for the SportNER and Bakeoff-3 datasets was 92.51% and 91.91%, respectively, which are the optimal results.

The algorithm proposed in this paper has the best performance in two datasets: OntoNotes and ResumeNER, i.e., F-score is 93.98% and 95.01%, respectively. In comparison to the lattice LSTM Chinese named-entity recognition algorithm proposed by Yue Zhang et al. [33], our proposed algorithm improved by 0.7% and 0.55%, respectively. The character-level GCN feature-extraction method designed in this paper has several advantages. This includes an improvement in capturing the neighborhood information of the character nodes; emphasizing the relevance between the characters and fully capturing the dependency between the characters, characters, and words in the text. Finally, self-attention is used to capture the global semantic information of the text.

#### 4.2.5. The Effect of Layers on the Performance

In order to verify the effect of the GCN layers on the performance of our proposed model algorithm, the Bi-LSTM layers were fixed as one layer along with a change in the GCN layer. The specific experimental results are shown in Table 8.

**Table 8.** Effect of the different graph convolutional network (GCN) layers on the model performance.

| Layers-Datasets      | SportNER     | Bakeoff-3    | OntoNotes    | ResumeNER    |
|----------------------|--------------|--------------|--------------|--------------|
|                      | F (%)        | F (%)        | F (%)        | F (%)        |
| 1 Layers(GCN)        | 91.07        | 89.93        | 92.30        | 93.54        |
| <b>2 Layers(GCN)</b> | <b>92.51</b> | <b>91.91</b> | <b>93.98</b> | <b>95.01</b> |
| 3 Layers(GCN)        | 92.19        | 91.72        | 93.11        | 94.8         |

In Table 8, 1 Layers(GCN) indicates that the number of GCN layers in cl-GCN is one; 2 Layers(GCN) indicates that the number of GCN layers in the character-level GCN is two; 3 Layers(GCN) indicates that the number of GCN layers in the character-level GCN is three. From Table 8, we can see that with an increase in the number of GCN layers, the F-Score of the four reference data sets increases first and then decreases. When the GCN layers are two, the F-Score is the highest. This is because fewer layers cannot fully mine the character information of the named entities and the hierarchical information in sports text. However, the excessive number of layers may lead to the over-fitting phenomenon of the model, which makes the network fall into local optimality. Therefore, the F-Score on the four benchmark data sets declines.

In order to verify the influence of the Bi-LSTM layers on the performance of the proposed model algorithm, the GCN layers were fixed to two layers. By changing the Bi-LSTM layers, the specific results are provided in Table 9.

**Table 9.** The effect of different bidirectional long-short-term memory (Bi-LSTM) layers on the model performance.

| Layers-Datasets          | SportNER     | Bakeoff-3    | OntoNotes    | ResumeNER    |
|--------------------------|--------------|--------------|--------------|--------------|
|                          | F (%)        | F (%)        | F (%)        | F (%)        |
| <b>1 Layer (Bi-LSTM)</b> | <b>92.51</b> | <b>91.91</b> | <b>93.98</b> | <b>95.01</b> |
| 2 Layers (Bi-LSTM)       | 90.94        | 90.15        | 93.42        | 94.76        |
| 3 Layers (Bi-LSTM)       | 90.07        | 90.04        | 92.99        | 93.87        |

From Table 9, it can be seen that, in comparison to the case when the Bi-LSTM layer is one, the F-score of the Bi-LSTM layer is two. This is reduced by 0.56% and 0.25%, respectively, for OntoNotes and ResumeNER.

As the number of Bi-LSTM layers increases, the F-Score decreases. When the number of Bi-LSTM layer is one, the F-Score of the four benchmark data sets is 92.51%, 91.91%, 93.98%, and 95.01%, respectively, which is increased by 0.56% and 0.25%, respectively, in comparison with the case when there are two Bi-LSTM layers. The reason is that when the number of layers increases, the network model may fall into local optimality or overfit, so the recognition effect of the model decreases when the number of layers increases.

To show the recognition effect of the proposed algorithm for the various named entities, the confusion matrix of the ResumeNER data set is presented in Figure 4.

In Figure 4, the main diagonal is the F-Score of the various named entities. The sum of the main diagonal is 95.01%. COUNT is "Country;" EDU is "Education Institution;" LOC is "Location;" PER is "Personal Name;" ORG is "Organization;" PRO is "Profession;" ETHBACK is "Ethnicity background;" JOBT is "Job;" O is "non-entity."

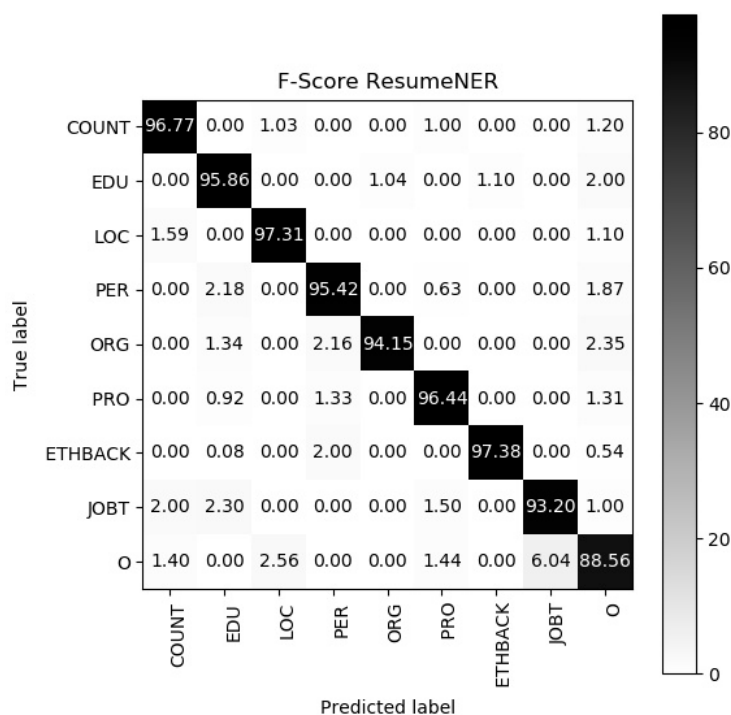


Figure 4. A classification result of the various named entities in ResumeNER.

### 5. Conclusions

To capture the hierarchy information and the correlation between the named entities in sports text, this paper proposes a named-entity recognition method based on the CGCN-SAN network. The experimental results show that the named-entity recognition framework proposed in this paper not only effectively excavates the deep abstract features and the global semantic information of the sports text but also captures the key information of named entities in the text by introducing the self-attention mechanism. Furthermore, the relationship between the hierarchical structural information of the character graph convolution feature and the named-entity is demonstrated. This method outperforms the traditional named-entity recognition methods in terms of P-, R-, and the F-Score. However, the proposed method is based on small-scale datasets for testing; consequently, it has considerable limitations. The next step is to expand the data scale. Further, after sorting the data, we plan to share this dataset. We also plan to perform experiments from the aspects of multi-feature fusion and sparse feature representation to capture the hierarchical structure.

**Author Contributions:** Conceptualization, X.S. and A.W.; methodology, X.S. and A.S.; validation, X.S. and A.W.; formal analysis, A.W.; investigation, X.S. and A.S.; resources, A.S. and T.Y.; data curation, L.W. and D.P.; writing—original draft preparation, X.S. and A.S.; writing—review and editing, A.S.; supervision, T.Y. and A.W. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the Opening Foundation of the Key Laboratory of the Xinjiang Uyghur Autonomous Region of China, grant number 2018D04019; the National Natural Science Foundation of China, grant numbers 61762084, 61662077, and 61462083; and the Scientific Research Program of the State Language Commission of China, grant number ZDI135-54.

**Acknowledgments:** The authors gratefully acknowledge all anonymous reviewers and editors for their constructive suggestions for the improvement of this paper.

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



## References

1. Xie, J.; Yang, Z.; Neubig, G.; Smith, N.A.; Carbonell, J. Neural cross-lingual named entity recognition with minimal resources. In Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, Brussels, Belgium, 31 October–4 November 2018; pp. 369–379.
2. Gu, C.; Song, X. Research on named entity recognition in sports events field. *J. Henan Norm. Univ. Nat. Sci. Ed.* **2018**, *43*, 163–167.
3. Feng, Y.; Yu, H.; Sun, G. Named entity recognition method based on BLSTM. *Comput. Sci.* **2018**, *20*, 872–884.
4. Habibi, M.; Weber, L.; Neves, M.; Wiegandt, D.L.; Leser, U. Deep learning with word embeddings improves biomedical named entity recognition. *Bioinformatics* **2017**, *33*, i37–i48. [[CrossRef](#)] [[PubMed](#)]
5. Pham, T.H.; Le-Hong, P. End-to-end recurrent neural network models for Vietnamese named entity recognition: Word-level vs. character-level. In *Communications in Computer and Information Science, Proceedings of the International Conference of the Pacific Association for Computational Linguistics, Yangon, Myanmar, 16–18 August 2017*; Springer: Singapore, 2017; pp. 219–232.
6. Augenstein, I.; Derczynski, L.; Bontcheva, K. Generalisation in named entity recognition: A quantitative analysis. *Comput. Speech Lang.* **2017**, *44*, 61–83. [[CrossRef](#)]
7. Unanue, I.J.; Borzeshi, E.Z.; Piccardi, M. Recurrent neural networks with specialized word embeddings for health-domain named-entity recognition. *J. Biomed. Inform.* **2017**, *76*, 102–109. [[CrossRef](#)]
8. Lee, J.Y.; Deroncourt, F.; Szolovits, P. Transfer learning for named-entity recognition with neural networks. *arXiv* **2017**, arXiv:1705.06273.
9. Wang, C.; Chen, W.; Xu, B. Named entity recognition with gated convolutional neural networks. In Proceedings of the International Symposium on Natural Language Processing Based on Naturally Annotated Big Data China National Conference on Chinese Computational Linguistics, Nanjing, China, 13–15 October 2017; pp. 100–121.
10. Luo, L.; Yang, Z.; Yang, P.; Zhang, Y.; Wang, L.; Lin, H.; Wang, J. An attention-based approach for chemical compound and drug named entity recognition. *J. Comput. Res. Dev.* **2018**, *34*, 1381–1388.
11. Cetoli, A.; Bragaglia, S.; O’Harney, A.D.; Sloan, M. Graph convolutional networks for named entity recognition. *arXiv* **2017**, arXiv:1709.10053.
12. Song, Y.; Shi, S.; Li, J. Joint learning embeddings for Chinese words and their components via ladder structured networks. In Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence, Stockholm, Sweden, 13–19 July 2018.
13. Huang, W.; Wang, J. Character-level convolutional network for text classification applied to Chinese corpus. *arXiv* **2016**, arXiv:1611.04358.
14. Chiu, J.P.; Nichols, E. Named entity recognition with bidirectional LSTM-CNNs. *Trans. Assoc. Comput. Linguist.* **2016**, *4*, 357–370. [[CrossRef](#)]
15. Xie, T.; Grossman, J.C. Crystal graph convolutional neural networks for an accurate and interpretable prediction of material properties. *Phys. Rev. Lett.* **2018**, *120*, 145301. [[CrossRef](#)] [[PubMed](#)]
16. Coley, C.W.; Jin, W.; Rogers, L.; Jamison, T.F.; Jaakkola, T.S.; Green, W.H.; Barzilay, R.; Jensen, K.F. A graph-convolutional neural network model for the prediction of chemical reactivity. *Chem. Sci.* **2019**, *10*, 370–377. [[CrossRef](#)] [[PubMed](#)]
17. Schwarzer, M.; Rogan, B.; Ruan, Y.; Song, Z.; Lee, D.Y.; Percus, A.G.; Srinivasan, G. Learning to fail: Predicting fracture evolution in brittle material models using recurrent graph convolutional neural networks. *Comput. Mater. Sci.* **2019**, *162*, 322–332. [[CrossRef](#)]
18. Abu-El-Haija, S.; Kapoor, A.; Perozzi, B.; Lee, J. N-GCN: Multi-scale graph convolution for semi-supervised node classification. *arXiv* **2018**, arXiv:1802.08888.
19. Abdelpakey, M.H.; Shehata, M.S.; Mohamed, M.M. Denssiam: End-to-end densely-Siamese network with self-attention model for object tracking. In Proceedings of the International Symposium on Visual Computing, Las Vegas, NV, USA, 19–21 November 2018; pp. 463–473.
20. Sun, F.; Li, W.; Guan, Y. Self-attention recurrent network for saliency detection. *Multimed. Tools Appl.* **2018**, *78*, 30793–30807. [[CrossRef](#)]
21. Cho, K.; Van Merriënboer, B.; Gulcehre, C.; Bahdanau, D.; Bougares, F.; Schwenk, H.; Bengio, Y. Learning phrase representations using RNN encoder-decoder for statistical machine translation. *arXiv* **2014**, arXiv:1406.1078.

22. Miao, Y.; Gowayyed, M.; Metze, F. EESSEN: End-to-end speech recognition using deep RNN models and WFST-based decoding. In Proceedings of the 2015 IEEE Workshop on Automatic Speech Recognition and Understanding (ASRU), Scottsdale, AZ, USA, 13–17 December 2015; pp. 167–174.
23. Sarker, M.M.K.; Rashwan, H.A.; Akram, F.; Talavera, E.; Banu, S.F.; Radeva, P.; Puig, D. Recognizing food places in egocentric photo-streams using multi-scale atrous convolutional networks and self-attention mechanism. *IEEE Access* **2019**, *7*, 39069–39082. [[CrossRef](#)]
24. Salazar, J.; Kirchoff, K.; Huang, Z. Self-attention networks for connectionist temporal classification in speech recognition. *arXiv* **2019**, arXiv:1901.10055.
25. Cross, J.; Huang, L. Incremental parsing with minimal features using bi-directional LSTM. *arXiv* **2016**, arXiv:1606.06406.
26. Li, S.; Yan, Z.; Wu, X.; Li, A.; Zhou, B. A method of emotional analysis of movie based on convolution neural network and bi-directional LSTM RNN. In Proceedings of the 2017 IEEE Second International Conference on Data Science in Cyberspace (DSC), Shenzhen, China, 26–29 June 2017; pp. 156–161.
27. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, L.; Polosukhin, I. Attention is all you need. In Proceedings of the Advances in Neural Information Processing Systems, Long Beach, CA, USA, 4–9 December 2017.
28. Verga, P.; Strubell, E.; McCallum, A. Simultaneously self-attending to all mentions for full-abstract biological relation extraction. *arXiv* **2018**, arXiv:1802.10569.
29. Chen, X.; Qiu, X.; Zhu, C.; Liu, P.; Huang, X. Long short-term memory neural networks for Chinese word segmentation. In Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing, Lisbon, Portugal, 17–21 September 2015; pp. 1197–1206.
30. Viterbi, A. Error bounds for convolutional codes and an asymptotically optimum decoding algorithm. *IEEE Trans. Inf. Theory* **1967**, *13*, 260–269. [[CrossRef](#)]
31. 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, Sydney, Australia, 22–23 July 2006; pp. 108–117.
32. Weischedel, R.; Pradhan, S.; Ramshaw, L.; Palmer, M.; Xue, N.; Marcus, M.; Taylor, A.; Greenberg, C.; Hovy, E.; Belvin, R.; et al. *Ontonotes Release 4.0*; LDC2011T03; Linguistic Data Consortium: Philadelphia, PA, USA, 2011.
33. Zhang, Y.; Yang, J. Chinese NER Using Lattice LSTM. *arXiv* **2018**, arXiv:1805.02023.
34. Lample, G.; Ballesteros, M.; Subramanian, S.; Kawakami, K.; Dyer, C. Neural architectures for named entity recognition. *arXiv* **2016**, arXiv:1603.01360.
35. Yu, X.; Mayhew, S.; Sammons, M.; Roth, D. On the strength of character language models for multilingual named entity recognition. *arXiv* **2018**, arXiv:1809.05157.
36. Greenberg, N.; Bansal, T.; Verga, P.; McCallum, A. Marginal likelihood training of BiLSTM-CRF for biomedical named entity recognition from disjoint label sets. In Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, Brussels, Belgium, 31 October–4 November 2018; pp. 2824–2829.
37. Wu, F.; Liu, J.; Wu, C.; Huang, Y.; Xie, X. Neural Chinese named entity recognition via CNN-LSTM-CRF and joint training with word segmentation. *arXiv* **2019**, arXiv:1905.01964.
38. Bekoulis, G.; Deleu, J.; Demeester, T.; Develder, C. Joint entity recognition and relation extraction as a multi-head selection problem. *Expert Syst. Appl.* **2018**, *114*, 34–45. [[CrossRef](#)]

