International Journal of Geo-Information

MDPI

Article

Deep Learning-Based Named Entity Recognition and Knowledge Graph Construction for Geological Hazards

Runyu Fan [1,2], Lizhe Wang [1,2,*], Jining Yan [1,2], Weijing Song [1,2], Yingqian Zhu [1,2] and Xiaodao Chen [1,2]

1 School of Computer Science, China University of Geosciences, Wuhan 430074, China; runyufan@cug.edu.cn (R.F.); yanjn@cug.edu.cn (J.Y.); songwj@cug.edu.cn (W.S.); yingqianzhu@cug.edu.cn (Y.Z.); chenxd@cug.edu.cn (X.C.)

2 Hubei Key Laboratory of Intelligent Geo-Information Processing, China University of Geosciences, Wuhan 430074, China

* Correspondence: lzwang@cug.edu.cn

Received: 14 November 2019; Accepted: 23 December 2019 ; Published: 27 December 2019

Abstract: Constructing a knowledge graph of geological hazards literature can facilitate the reuse of geological hazards literature and provide a reference for geological hazard governance. Named entity recognition (NER), as a core technology for constructing a geological hazard knowledge graph, has to face the challenges that named entities in geological hazard literature are diverse in form, ambiguous in semantics, and uncertain in context. This can introduce difficulties in designing practical features during the NER classification. To address the above problem, this paper proposes a deep learning-based NER model; namely, the deep, multi-branch BiGRU-CRF model, which combines a multi-branch bidirectional gated recurrent unit (BiGRU) layer and a conditional random field (CRF) model. In an end-to-end and supervised process, the proposed model automatically learns and transforms features by a multi-branch bidirectional GRU layer and enhances the output with a CRF layer. Besides the deep, multi-branch BiGRU-CRF model, we also proposed a pattern-based corpus construction method to construct the corpus needed for the deep, multi-branch BiGRU-CRF model. Experimental results indicated the proposed deep, multi-branch BiGRU-CRF model outperformed state-of-the-art models. The proposed deep, multi-branch BiGRU-CRF model constructed a large-scale geological hazard literature knowledge graph containing 34,457 entities nodes and 84,561 relations.

Keywords: named entity recognition; knowledge graph; deep learning; geological hazards

1. Introduction

Knowledge graphs of geological hazards literature can facilitate the reuse of geological hazards literature and provide a reference for geological hazard mitigation. There is significant literature related to geological hazard research on the Wanfang academic platform (Wanfang database), and it is difficult for researchers to read all of these articles to find the information they need. Using machine learning methods to recognize the named entities from the geological hazard related literature and constructing a knowledge graph can greatly enhance the reuse of literature, and increase efficiency and convenience in the research and governance of geological hazards.

Named entity recognition (NER) is a technology to classify mentions of entities in unstructured text into pre-defined categories. Named entities in geological hazard literature are diverse in form, ambiguous in semantics, and uncertain in context. Named entities in geological hazard literature have diverse forms. For example, Los Angeles, the City of Los Angeles, and L.A., are different expressions of the same location name. Named entities in geological hazard literature have ambiguous semantics.

For example, Jordan is an Arab country named the Hashemite Kingdom of Jordan in Western Asia, but also refers to a famous basketball player named Michael Jordan, depending on the context. Besides, named entities in geological hazard literature have an uncertain context. The context of the same entity is not the same. For example, the phrase prior to "Los Angeles" can be the phrase "located at" or "near". Therefore, it is challenging to design features with complete accuracy, which makes the recognition of named entities difficult and potentially ineffective.

Focusing on the above problems, in this paper, we propose a deep learning-based method; namely, the deep, multi-branch BiGRU-CRF model, for NER of geological hazard literature named entities. The proposed deep, multi-branch BiGRU-CRF model combines a multi-branch BiGRU layer and a CRF model. Considering that named entities in geological hazard literature are diverse in form, we used the context information of the named entities in the whole sentence to help to predict on the named entities. Considering named entities in geological hazard literature are ambiguous in semantics, we propose a multi-branch structure to extract different levels of semantic information, and use the attention mechanism [1] and residual structure [2] to enhance the feature from each branch of different depths. Considering named entities in geological hazard literature are uncertain in context, we use BiGRU layers to extract the contextual features of the named entities in both the forward and reverse directions. However, because the tag sequences themselves are also constrained, the multi-branch BiGRU layer does not learn these dependencies very well. Therefore, we added a CRF layer on top of the multi-bidirectional GRU layer. The CRF model is used to further constrain the tags with context information in different time steps and ultimately to output the optimized tags of the currently observed Chinese characters.

Besides the deep, multi-branch BiGRU-CRF model, we proposed a pattern-based corpus construction method to construct the corpus needed for the deep, multi-branch BiGRU-CRF model. In the pattern-based corpus construction method, we first obtained a large number of seeds automatically by some manually designed patterns, and then backed up the seeds in a large amount of geological hazard research literature to construct a large-scale geological hazard NER corpus using a maximum forward matching (MFM) method.

The proposed NER model achieved an average precision of 0.9413, an average recall rate of 0.9425, and an average F1 score of 94.19. The proposed deep, multi-branch BiGRU-CRF model constructed a large-scale geological hazard literature knowledge graph containing 34,457 entities nodes and 84,561 relations.

The main contributions of the proposed method are as follows:

- To the best of our knowledge, this is the first work to apply the NER technique to extract named entities and build a knowledge graph for geological hazards literature.
- This paper proposed a deep learning-based NER model that combines a multi-branch BiGRU layer and a CRF model for geological hazard NER. The model uses a multi-branch structure; each branch contains a BiGRU layer of different depths to extract different levels of features, and then further enhances the preliminary features using the attention mechanism and the residual structure.
- This paper proposed a pattern-based method to build a large-scale geological hazard literature NER corpus with little manual costs.

The rest of this paper is structured as follows. Section 2 shows related work. Section 3 shows preliminaries. Section 4 introduces our approach, and Section 5 presents the implementation. Section 6 summarizes experimental results. Section 7 discusses the paper and Section 8 concludes the paper.

2. Related Work

With the development of statistical machine learning methods and natural language processing technology, in recent years, many scholars and institutions have begun to study how to use natural language processing (NLP) [3] technology to extract knowledge and construct knowledge graph from geoscience-related literature.

Zhu et al. [4] conducted knowledge extraction on a large number of geological hazard literature and linked open data (LOD) [5], and constructed a knowledge graph. Specifically, the TextRank [6] algorithm was first used to extract the literature keywords, and the geological domain entities were obtained by combining the entries of the open link data (such as Baidu Encyclopedia, Interactive Encyclopedia, and Wikipedia) and the extracted keywords. On this basis, the key rule algorithm was used to obtain the relationship and build the geological knowledge map. This method of using a LOD (Baidu Encyclopedia, Interactive Encyclopedia, and Wikipedia) entry catalog to acquire relevant geological domain entities was groundbreaking. However, this method can only get entities that are already included in the encyclopedic knowledge base and LOD. The coverage of geological knowledge contained in current general encyclopedias (Baidu Encyclopedia, Interactive Encyclopedia, and Wikipedia) is small. Therefore, the scale and depth of the knowledge graph constructed using this method are relatively small.

In order to better extract knowledge from the unstructured geoscience literature, Wang et al. [7] designed a workflow for knowledge extraction and construction of knowledge graph for geoscience literature. First, a corpus containing domain corpus and general domain corpus were constructed for word segmentation. Secondly, based on this corpus, a word segmentation model was trained using the conditional random field (CRF) [8]. Then, they used this model to segment the literature. Finally, the TF-IDF [9,10] method was used to extract the keywords of the literature, and the keywords with relatively large co-occurrence relations were connected to form a knowledge graph. Shi et al. [11] also used TF-IDF to extract keywords to construct a knowledge graph. However, unlike Wang et al. [7], Shi et al. [11] trained a CNN-based classifier that automatically divides the geoscience literature into four categories (geophysics, geology, remote sensing, and geochemistry) and then constructs the corresponding knowledge graph.

These methods have brought great inspiration to the extraction of knowledge and knowledge graph construction in the geoscience literature, but there are also some shortcomings worth improving. These methods use statistical analysis methods to extract keywords, high-frequency words, etc., rather than entities, as nodes in their knowledge graph. Nevertheless, more often, in order to better analyze and understand the geological disaster literature, we need to extract the entities in the literature that represent specific categories and meanings, such as methods, data, etc.

NER is the task of identifying a named entity in text and classifying it into a specified category [12]. NER was first proposed in the MUC [12] mission of the 1980s and has been a hot topic in natural language processing research.

Some studies start with text mining methods and build specific rules for NER. These methods adopt the strategy of bootstrapping to extract entities of the specified categories from the Web. Representative work includes the TextRunner system [13], the Snowball system [14], and the CasSys system [15]. The disadvantage of these methods is that the bootstrapping iteration introduces noise instances and noise templates, resulting in poor results.

Since the 1990s, statistical models have been the mainstream method for NER. There are a number of statistical methods [16,17] used to extract entities from text, such as as the maximum entropy model (ME) [18–20], support vector machines (SVM) [21–24], the hidden Markov model (HMM) [25–27], the CRF model [28–30], and so on. Statistical model-based methods typically formalize entity recognition tasks from the input text to predict specific target structures, use statistical models to model the association between input and output, and use machine learning methods to learn parameters of the model.

With the excellent performance of deep learning in different fields, more and more deep learning models have been proposed to solve the problem of NER. Currently, there are two typical deep learning architectures for NER. The first is the NN-CRF architecture [31–34], in which CNNs/RNNs are used to learn the vector representation at each word position. Based on the vector representation, the CRF layer decodes the best label at that location. The second adopts the idea of sliding window classification, uses neural networks to learn the representation of each n-gram in the sentence, and then predicts

whether the n-gram is a target entity [35–37]. Compared with the traditional statistical model, the main advantage of the deep learning method is that its training is an end-to-end process, without the need to manually design related features. Besides, deep learning facilitates learning a specific representation of the task. By learning the correlation of information between different modalities, different types, and language environments, better entity recognition performance can be achieved.

These NER methods provide a useful reference for NER tasks in geoscience. Sobhana et al. [38] first used the CRF model combined with some manually designed features (such as prefixes and suffixes for words) to extract 17 types of geoscience-related entities from geoscience texts. Considering named entities in geological hazard literature are diverse in form and complicated in context, it is challenging to design practical features, resulting in a poor performance by CRF models that rely on manually designed features.

Inspired by the above NN-CRF architecture [31–34], in this paper, we propose a deep learning-based method; namely, the deep, multi-branch BiGRU-CRF model, for NER of geological hazard literature named entities. The proposed deep, multi-branch BiGRU-CRF model combines a multi-branch BiGRU layer and a CRF model. The multi-branch structure combines the attention mechanism and the residual structure, which can learn different depths and levels of features. The BiGRU network can obtain the context information of named entities from both forward and reverse directions. The CRF model can further optimize the prediction results based on the dependencies between the tags.

3. Preliminaries

In the deep, multi-branch BiGRU-CRF model for geological hazard NER, we use two widely used models, GRU and CRF. They are introduced in the preliminary section.

3.1. GRU

Since a recurrent neural network (RNN) [39,40] does not handle long-range dependencies well, a long short-term memory network (LSTM) [41–43] is proposed. GRU [44], which can be seen in Figure 1, is a variant of LSTM. GRU maintains the effects of LSTM while making the structure simpler, and it has a wide range of applications in many tasks of natural language processing, sequence analysis, image processing, etc., [45–47].

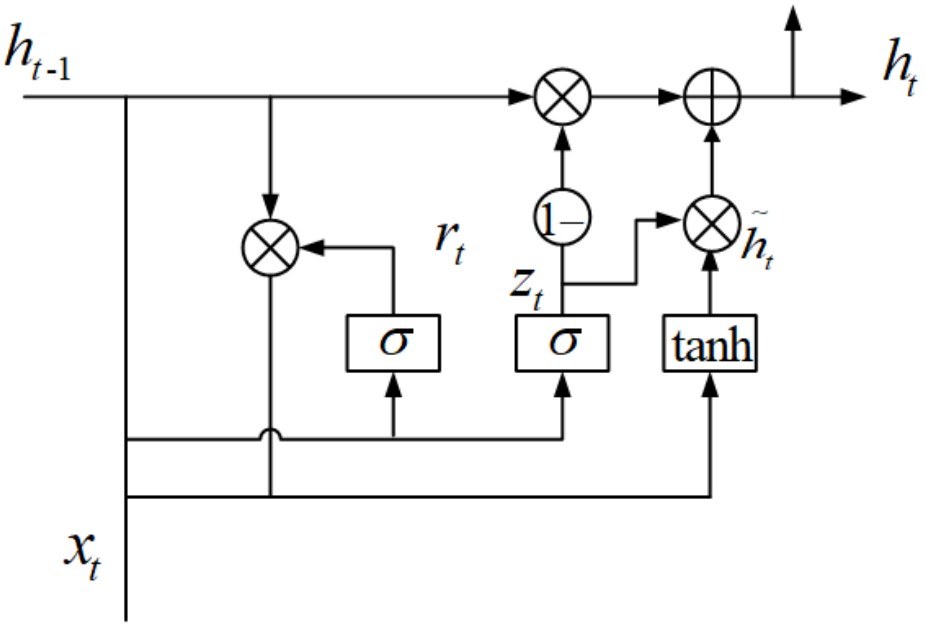

Figure 1. Gated recurrent unit.

The GRU model has only two gates, the update gate and the reset gate; namely, z_t and r_t in Figure 1. The update gate is used to control the degree to which the status information of the previous moment is brought into the current state. The larger the value of the update gate, the more the status information of the previous moment is brought in. The reset gate is used to control the degree to which status information from the previous moment is neglected or forgotten. The smaller the value of the reset gate, the more the information from the previous moment is neglected. The reset gate helps capture short-term dependencies in the time series data, while the update gate helps capture long-term dependencies in the time series data [42,45,48–50].

The reset gate r_t and update gate z_t are defined as follows:

$$r_t = \sigma\left(W_{xr}x_t + W_{hr}h_{t-1} + b_r\right) \tag{1}$$

$$z_t = \sigma\left(W_{xz}x_t + W_{hz}h_{t-1} + b_z\right), \tag{2}$$

where σ is the sigmoid activation function [51]. h_t represents the implied state and is defined as follows:

$$h_t = z_t \odot h_{t-1} + (1 - z_t) \odot \widetilde{h}_t \tag{3}$$

where $\odot$ is the element product operator of two vectors, and $\widetilde{h}_t$ represents the candidate implied state and is defined as follows:

$$\widetilde{h}_t = \tanh\left(W_{xh}x_t + W_{hh}\left(r_t \odot h_{t-1}\right) + b_h\right). \tag{4}$$

The candidate implied state $\widetilde{h}_t$ uses a reset gate r_t to control the inflow of the last implied state $\widetilde{h}_{t-1}$ containing past time information. If the value of the reset gate r_t converges to a value closed to 0, the last implicit state $\widetilde{h}_{t-1}$ will be discarded. Therefore, the reset gate r_t provides a mechanism to discard past implied states that are unrelated to the future; that is, the reset gate r_t determines how much information is left in the past. The implicit state h_t uses the update gate z_t to update the last implicit state h_{t-1} and the candidate implied state. Updating the gate can control the importance of the past implied state at the current moment. If the value of the update gate always converged to a value closed to 1, the past implied state will be saved over time and passed to the current time. This design can cope with the vanishing gradient problem [52,53] in the recurrent neural network and better capture the large interval dependencies in the time series data.

3.2. CRF

The CRF model is a discriminant probability, undirected graph learning model proposed by Lafferty [8] based on the maximum entropy model [54] and hidden Markov model [55]. CRF was first proposed for sequence data analysis and has been successfully applied in the fields of natural language processing (NLP), bioinformatics, machine vision, and network intelligence [56–59].

Let $G = (V, E)$ be an undirected graph, where V is the set of nodes and E is the set of edges, and let $Y = \{Y_v | v \in V\}$ be a set of random variables Y_v indexed by node v in V. Given a condition of X, if each random variable Y_v obeys the Markov property:

$$P(Y_v|X, Y_u, u \neq v) = P(Y_v|X, Y_u, u \sim v), \tag{5}$$

then (X, Y) constitutes a CRF, where X represents the observed sequence and $u \sim v$ represents all neighbor nodes of u connected by the node v in graph G.

Linear Chain-CRFs

Linear chain-CRFs [60], as shown in Figure 2 are a common form of CRF model. Let $x = \{x_1, x_2, \cdots, x_n\}$ denote the observation sequence and $y = \{y_1, y_2, \cdots, y_n\}$ be the set of finite states, according to the basic theory of the random field:

$$P(Y = y|x) = \frac{1}{Z(x)} \exp\left(\sum_k \lambda_k \sum_{i=1}^{n-1} t_k(y_{i+1}, y_i, x, i) + \sum_l \mu_l \sum_{i=1}^{n} s_l(y_i, x, i)\right) \quad (6)$$

$$Z(x) = \sum_y \exp\left(\sum_k \lambda_k \sum_{i=1}^{n-1} t_k(y_{i+1}, y_i, x, i) + \sum_l \mu_l \sum_{i=1}^{n} s_l(y_i, x, i)\right) \quad (7)$$

where the terms are defined as follows:

$t_k\left(y_{i+1}, y_i, x, i\right)$: transfer characteristic function between the marked positions i and $i+1$ of the observed sequence. It is used to characterize the correlation between adjacent finite states and the influence of observation sequences on them.

λ_k: weights of the transfer characteristic function $t_k\left(y_{i+1}, y_i, x, i\right)$.

$s_l\left(y_i, x, i\right)$: State feature function of the observed sequence at position i. It is used to characterize the effect of observation sequences on finite states.

μ_l: weights of the state feature function $s_l\left(y_i, x, i\right)$.

$Z(x)$: a normalization factor used to ensure that formula (6) is a correctly defined probability.

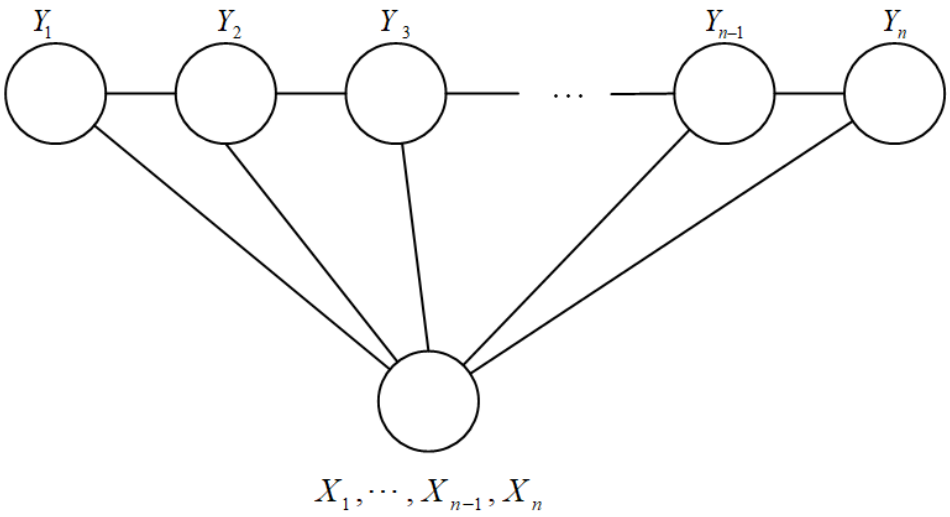

Figure 2. Linear chain-CRFs [8].

4. The Proposed Methods

In this section, the proposed method is introduced in detail. The proposed method aims to extract geological hazard named entities from the considerable body of geological hazard literature and build a geological hazard knowledge graph.

In this paper, we propose a geological hazard NER model based on the deep learning method; namely, the deep, multi-branch BiGRU-CRF model, to extract geological hazard named entities and construct a knowledge graph. Since the proposed model is a supervised model that requires an annotated corpus, we propose a pattern-based corpus construction method to provide a corpus for the deep, multi-branch BiGRU-CRF model. The proposed method is presented in two parts: pattern-based corpus construction and the deep, multi-branch BiGRU-CRF model for NER.

1. Pattern-based corpus construction. Given literature documents $F = \{f_1, f_2, ..., f_N\}$ where f_n $(n \in [1, N])$ is the n-th document and *Patterns* $P = \{p_m, p_l, p_d\}$ where p_m, p_l, p_d are patterns

for *methods*, *location*, and *data*, respectively. The pattern-based corpus construction method aims to construct a named entity corpus C.

2. The deep, multi-branch BiGRU-CRF model for NER. Given literature documents $F = \{f_1, f_2, ..., f_N\}$ where f_n $(n \in [1, N])$ is the nth document and the named entity corpus C, the proposed deep, multi-branch BiGRU-CRF model aims to extract *methods location*, and *data* entities from F and constructs a knowledge graph G.

4.1. Pattern-Based Corpus Construction

Pattern-based corpus construction can be divided into three steps. Firstly, we define the three named entities we want to extract. Then *Patterns* $P = \{p_m, p_l,$ and $p_d\}$ are used to get the named entity seeds from the geological hazard literature $F = \{f_1, f_2, ..., f_N\}$. Finally, the maximum forward matching method (MFM) is used to map the seeds, which are *MethodsSeeds* $M = \{m_1, m_2, ..., m_I\}$, *LocationSeeds* $L = \{l_1, l_2, ..., l_J\}$, *DataSeeds* $D = \{d_1, d_2, ..., d_K\}$ to the literature for labeling. By then a geological hazard named entity corpus C is constructed. We introduce the three parts in detail below.

4.1.1. Definition of Named Entities in Geological Hazard Literature

NER tasks are defined in MUC-7, in which named entities are defined as proper names and quantities of interest [12,17]. Named entities include person names, place names, and organization names, and times, dates, amounts, and percentages. Among them, the most commonly used named entities are person names, place names, and organization names [17,61].

For geological hazards literature research, the three named entities proposed in the literature are methods, data used, and descriptions of regions and locations. When reading geological hazards literature, researchers usually care about the study area targeted, the methods proposed, and the data used. Most articles generally have three Sections (methodology, data, and study area) that correspond to the three named entities mentioned above. These entities have the most important role in the understanding, research, and reuse of geological hazard literature. This article focuses on the extraction of the above three types of named entities: *methods*, *data*, and *location*. Table 1 shows the details of these entities.

Table 1. Three types of named entities defined.

Entity Type	Description	Tags
Location	descriptions of regions and location	B-LDS, I-LDS
Methods	methods, techniques and models	B-MED, I-MED
Data	data used	B-DAT, I-DAT
Not an entity	Not a entity	O

4.1.2. Pattern-Based Seed Acquisition

Given the three defined named entities above (*methods*, *location*, and *data*), we extract these entities in this section and build entity seed collections *MethodsSeeds* $M = \{m_1, m_2, ..., m_I\}$, *LocationSeeds* $L = \{l_1, l_2, ..., l_J\}$ and *DataSeeds* $D = \{d_1, d_2, ..., d_K\}$.

Considering there are often certain rules among named entities of geological hazards, discovering these rules and designing related patterns can help us extract these named entities. Therefore, we have designed a pattern-based seed acquisition method to obtain these named entity seeds. The manually defined *Patterns* $P = \{p_m, p_l,$ and $p_d\}$, where p_m, p_l, p_d are patterns for *methods*, *location*, and *data*, respectively, are shown in Table 2:

Table 2. Patterns (regular expressions) used.

Entity Type	Patterns (Regular Expressions)
Location	'.*(located in|located in|in|form|located in)(of){0,1} ([\S]+)(of){0,1}(area|region|mountain area|river basin|zone).*'
Methods	'.*(provide|apply|improve|utilize|using|put forward|design|invente|set up | construct|achieve |according to|take |base on|construct |produce|combine |adopt |adopt|by|construct)(of|of|of|corresponding|of){0,1} (and){0,1}([\S]+)(of) {0,1}(method|model).*'
Data	'.*(provide|apply|utilize|using|put forward|design|invente|set up |construct|according to |take|base on|construct|produce|combine|adopt|adopt |by|construct|collect) (of|of|of|corresponding|and){0,1} ([\S]+)(of){0,1}(data|material|data set).*'

Patterns of Table 2 in this work are in Chinese. Please refer to Appendix A for detailed translation.

We use these patterns (regular expressions) to match the sentences $S = \{s_1, s_2, ..., s_H\}$ in the literature F from papers in the Wanfang database (http://www.wanfangdata.com.cn). The words that match those patterns (regular expressions) P are the entity seeds we want to extract. After that, we randomly select 2000 entity seeds each and manually check the entity seeds to evaluate the accuracy can be calculated by the following equation:

$$Accuracy = \frac{n_c}{n}, \tag{8}$$

where n_c denotes the number of correct entity seeds and n denotes the total number of entity seeds. The results are shown in Table 3. After manually checking, all correct entities form the entity seed collections $M = \{m_1, m_2, ..., m_I\}$, $L = \{l_1, l_2, ..., l_J\}$, and $D = \{d_1, d_2, ..., d_K\}$.

Table 3. Statistics of seeds extracted by patterns

Entity Type	Correct	All	Accuracy
Location	1179	2000	0.5895
Methods	1467	2000	0.7335
Data	1305	2000	0.6525

4.1.3. MFM for Corpus Construction

Given the three types of entity seed collections above ($M = \{m_1, m_2, ..., m_I\}$, $L = \{l_1, l_2, ..., l_J\}$, and $D = \{d_1, d_2, ..., d_K\}$) and sentences $S = \{s_1, s_2, ..., s_H\}$, the MFM method shown in Algorithm 1 is used to automatically construct a geological hazards named entity corpus C in a character-based format named IOB format [31], where "B" indicates the starting character of an entity, "I" indicates the intermediate characters and the ending character of an entity, and "O" indicates that the character is not part of entity [62]. Table 4 shows a illustration of IOB format. We defined seven types of tags ("O", "B-MED", "I-MED", "B-DAT", "I-DAT", "B-LDS", and "I-LDS"); see Table 1.

Table 4. Illustration of IOB format.

Tags	O O B-MED I-MED I-MED
Translation	Build a numerical analysis model

The MFM method shown in Algorithm 1 contains six steps:

(1) Sort the elements in the entity seed collections $M = \{m_1, m_2, ..., m_I\}$, $L = \{l_1, l_2, ..., l_J\}$ and $D = \{d_1, d_2, ..., d_K\}$ separately in decreasing order according to length.
(2) Initialize the corpus set C to an empty set.
(3) For each sentence S_h ($h \in [1, H]$) in $S = \{s_1, s_2, ..., s_H\}$, look up seeds in the seed sets M, L, and D. If there is a seed that is contained by S_h and is unlabeled, then label the words containing the seed in S_h with the corresponding entity tags.
(4) After traversing all the seed sets M, L, and D, the remaining unlabeled words in S_h are labeled as "O".
(5) Add the label S_h to the corpus set C.
(6) When there is no unlabeled S_h in S, the program ends and returns the corpus set C.

Algorithm 1 MFM

Input: *Sentences* $S = \{s_1, s_2, ..., s_H\}$, *MethodsSeeds* $M = \{m_1, m_2, ..., m_I\}$, *LocationSeeds* $L = \{l_1, l_2, ..., l_J\}$, *DataSeeds* $D = \{d_1, d_2, ..., d_K\}$

Output: *CorpusSet* C

function MFM(S, M, L, D)
  $C \leftarrow \varnothing$
  Sort the elements in M, L, D in decreasing order according to length separately.
  for $h = 1$ to H **do**
    for $i = 1$ to I **do**
      if M_i in S_h, and the characters of M_i are unlabeled **then**
        label "B-MED" in the first character and "I-MED" in the remaining characters of M_i in S_h.
      end if
    end for
    for $j = 1$ to J **do**
      if L_j in S_h, and the characters of L_j are unlabeled **then**
        label "B-LDS" in the first character and "I-LDS" in the remaining characters of L_j in S_h.
      end if
    end for
    for $k = 1$ to K **do**
      if D_k in S_h, and the characters of D_k are unlabeled **then**
        label "B-DAT" in the first character and "I-DAT" in the remaining characters of D_k in S_h.
      end if
    end for
    label unlabeled characters in S_h as "O."
    $C \leftarrow S_h + C$
  end for
  return C
end function

4.2. The Deep Multi-Branch BiGRU-CRF Model

Given the corpus C constructed above, we proposed a deep learning-based model named the deep, multi-branch BiGRU-CRF model, which combines neural networks and CRF for geological hazard NER. The model is shown in Figure 3 consists of three components, which are the embedding layer, the multi-branch BiGRU layer, and the CRF layer. The embedding layer is the first layer of the model, which converts Chinese characters into dense vectors and passes them to the multi-branch BiGRU layer. The multi-branch BiGRU layer learns different levels of features through a multi-branch BiGRU layer and passes these features to the CRF layer. The CRF layer further enhances the mapping of characters to tags and the probability of transition between tags and outputs the optimized tags as the final output of the proposed model. We introduce these three layers in detail below.

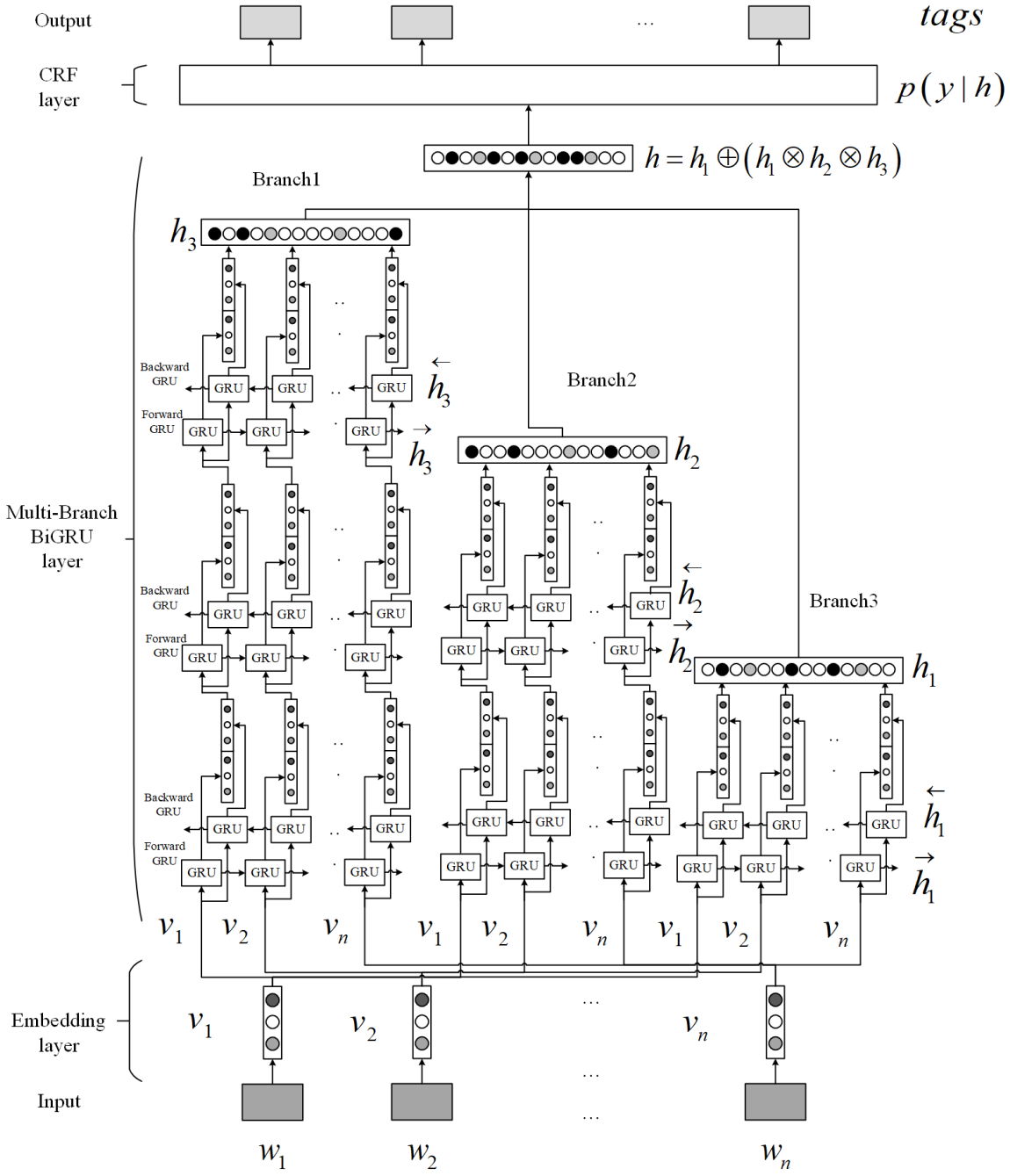

Figure 3. Deep, multi-branch BiGRU-CRF model.

4.2.1. Embedding Layer

Given Chinese characters $w_1, w_2, ..., w_n$ in sentence S_i as input, where $S_i \in S = \{s_1, s_2, ..., s_H\}$, the first step of deep neural networks is often to refer to discrete Chinese words in sentences as continuous vectors or a matrix. This step is called embedding. We use random 100-dimensional vectors $v_1, v_2, ..., v_n$ as the initialized representation of the character $w_1, w_2, ..., w_n$. $v_1, v_2, ..., v_n$ can be trained to get a better representation.

In this way, the input characters $w_1, w_2, ..., w_n$ are embedded as 100-dimensional vectors $v_1, v_2, ..., v_n$.

4.2.2. Multi-Branch BiGRU Layer

The output $v_1, v_2, ..., v_n$ of the embedding layer then passes through a multi-branch BiGRU layer. For every branch $t \in \{1, 2, \cdots, n\}$, n is the number of branches, the output $h_t = \left[\overrightarrow{h_t}; \overleftarrow{h_t}\right]$ is the concatenation of $\overrightarrow{h_t}$ and $\overleftarrow{h_t}$ where $\overrightarrow{h_t}$ and $\overleftarrow{h_t}$ represent the forward and inverse representation of $v_1, v_2, ..., v_n$ and can be calculated by Equation (3) from two different directions of GRU. Through this combination of the forward and reverse representations of $v_1, v_2, ..., v_n$, we can fully consider the context content of the characters, making the feature extraction more abundant. In the experiment, our multi-branch BiGRU layer consisted of three branches with depths of 1, 2, and 3, respectively. A large number of branches bring a lot of computational burden; too few branches cannot fully extract multiple levels of features. We use three branches with depths of 1, 2, and 3 to extract the low-level, middle-level, and high-level features, that is, h_1, h_2, and h_3. Then, we use the attention mechanism to weight the corresponding elements of h_1, h_2, h_3 to obtain the weighted feature matrix $h_{123} = h_1 \otimes h_2 \otimes h_3$, where $\otimes$ represents the multiplication of the corresponding elements of feature matrix. Then, the residual structure is used to add the weighted feature matrix h_{123} and the low-level features h_1, that is, $h_1 \oplus h_{123}$, to solve the problem of gradient disappearance and difficulty in training caused by increasing the number of layers. $\mathbf{h} = h_1 \oplus h_{123} = h_1 \oplus (h_1 \otimes h_2 \otimes h_3)$ is the output of the multi-branch BiGRU layer.

4.2.3. CRF Layer

The elements $\mathbf{h}_t$ in $\mathbf{h}$, where t represents the t-th element in $\mathbf{h}$, are not completely independent. For example, when $\mathbf{h}_t$ is "B-MED", the probability of $\mathbf{h}_{t+1}$ being "I-MED" is obviously much higher than the probability of being "B-DAT". Therefore, instead of treating $\mathbf{h}$ independently, we use a CRF layer to model the relationship between $\mathbf{h}$ and get the enhanced results. The CRF layer is added to calculate the conditional probability $p(\mathbf{y}|\mathbf{h})$ by Equation (9), where $\mathbf{y} = \{y_1, y_2, \cdots y_T\}$ represents the label sequences.

$$p(\mathbf{y}|\mathbf{h}; t, s) = \frac{\prod_{i=1}^{T} exp(\sum_{i=1}^{T} t(y_{i-1}, y_i, \mathbf{h}) + s(y_i, \mathbf{h}))}{\sum_{y' \in \gamma(\mathbf{h})} \prod_{i=1}^{T} exp(\sum_{i=1}^{T} t(y'_{i-1}, y'_i, \mathbf{h}) + s(y'_i, \mathbf{h}))} \tag{9}$$

where γ represents the sequences of all possible tags, t represents the transition probability for a given input sequence $\mathbf{h}$ from y_{i-1} to y_i, and s is the emission score of the transition from the output of BiGRU layer to y_i at time step i.

Finally, the model is trained by maximum conditional likelihood estimation [63] by Equation (10). The sequence that enables the conditional probability $p(\mathbf{y}|\mathbf{h}; t, s)$ to get the maximum value is the output of the model.

$$Loss\,(t, s) = \sum_{i}^{T} \log p(\mathbf{y}|\mathbf{h}; t, s). \tag{10}$$

5. Implementation

In this paper, the proposed multi-branch BiGRU-CRF model used the Python (version 3.6.3) programming language. The deep learning library used was TensorFlow-GPU (version 1.13.1). An NVIDIA Titan RTX GPU was used. We did not use any open APIs when obtaining the geological hazard research literature in the Wanfang database. We used web crawler technology to crawl the title and the abstract section of the theses related to geological disasters. The crawler used the Scrapy library and returned a text file, each line containing only the title and abstract of a paper. The knowledge graph was stored and visualized in the Neo4j database.

6. Experimental Results

This section shows the statistics of the corpus constructed by pattern-based methods, the parameter settings of training, the results of the proposed deep, multi-branch BiGRU-CRF model, and the knowledge graph constructed in the following four parts.

6.1. Corpus Constructed

The corpus was built automatically by the method mentioned in Section 4.1, containing 536,426 characters, 4548 sentences, and seven types of tags, for which the detailed statistics are shown in Table 5. We randomly split the data into a training set, a validation set, and a test set with a ratio of 8:1:1.

Table 5. Statistics of the tags in corpus.

Tags	"O"	"B-MED"	"I-MED"	"B-DAT"	"I-DAT"	"B-LDS"	"I-LDS"
The number of the tags	493,643	4230	13,682	1872	4550	6133	12,316

6.2. Training

For all the models mentioned, we update the parameters using the back-propagation algorithm and use stochastic gradient descent (SGD) to optimize our model. Our model uses three stacked BiGRU layers, each layer containing one forward GRU and one reverse GRU, and the number of neurons in each GRU is set to 100. We added a Dropout [64] between the BiGRU layer and the CRF layer to improve the model's effectiveness and prevent overfitting. The Dropout rate was set to 0.5, as higher rates negatively impacted our results, and lower rates led to longer training time.

6.3. Results

We used P (precision), R (recall rate), and F (F1 score), which are widely used evaluation criteria [31–34,65] in NER, to evaluate the three mentioned models. The larger the three evaluation criteria, the better the model's effect. P, R, and F can be calculated by the following three formulas:

$$P = \frac{n_p}{n_t} \tag{11}$$

$$R = \frac{n_p}{n_c} \tag{12}$$

$$F = \frac{2 * precision * recall}{precision + recall}, \tag{13}$$

where n_p denotes the number of true positive predictions; n_t denotes the total positive predictions, including both true and false; and n_c denotes the total number of predictions, including both positive and negative.

The result of our NER model is shown in Table 6.

1. The CRF model is the model proposed by Sobhana et al. [38], using CRF for NER in geosciences. We used the CRF method as our benchmark. As can be seen in the Table 6, the CRF model could initially identify these geological hazard named entities, achieving an average precision of 0.8210, a recall rate of 0.7765, and an F1 score of 79.81.
2. The BiLSTM-CRF model is the state-of-the-art model in current NER tasks[31]. It has one bidirectional LSTM layer and one CRF layer on top. As can be seen in the Table 6, the BiLSTM-CRF model has a significant lead on all indicators compared to the CRF model, with an average precision of 0.9205, an average recall rate of 0.9419, and an average F1 score of 93.10. It fully demonstrated that the BiLSTM-CRF model has more efficient feature extraction and more accurate discriminating ability after adding one bidirectional LSTM layer before the CRF layer.

3. The deep, multi-branch BiGRU-CRF model was the proposed model with a three-branch BiGRU layer, which consisted of three branches of stacked BiGRU layers with depths of 1, 2, and 3, respectively, and one CRF layer on top. As can be seen in the Table 6, the deep, multi-branch BiGRU-CRF model had a significant lead on almost all indicators (except the recall rate of methods) compared to the CRF model and BiLSTM-CRF model above, with an average precision of 0.9413, an average recall rate of 0.9425, and an average F1 score of 94.19. It fully demonstrated that the proposed model has more efficient feature extraction and more accurate discriminating ability after adding three branches of BiGRUs with depths of 1, 2, and 3, respectively.

Table 6. Result of proposed models. P, R, and F indicate our evaluation criteria precision, recall, and F1 score. DAT, LDS, and MED indicate the corresponding entity categories data, location and methods. "Avg." represents the overall weighted average score. The best performances are shown in bold.

Model	Evaluate	DAT	LDS	MED	Avg.
CRF	P	0.8697	0.8662	0.7259	0.8210
	R	0.7973	0.8486	0.6648	0.7765
	F	83.00	85.73	69.40	79.81
BiLSTM-CRF	P	0.9220	0.9450	0.8858	0.9205
	R	**0.9510**	0.9527	**0.9215**	0.9419
	F	93.63	94.89	90.33	93.10
Deep Multi-branch BiGRU-CRF model	P	**0.9645**	**0.9519**	**0.9135**	**0.9413**
	R	**0.9510**	**0.9622**	0.9100	**0.9425**
	F	**95.77**	**95.70**	**91.17**	**94.19**

6.4. Knowledge Graph Construction

We used the trained deep, multi-branch BiGRU-CRF model to perform NER on the geological hazard related papers in the Wanfang knowledge base, and obtain the three types of named entities (*location*, *methods*, and *data*) mentioned in this paper, and to construct a knowledge graph. Table 7 shows the named entities extracted from randomly selected papers. It can be seen that the proposed method correctly extracted the relevant location and area descriptions, the data used, and the models and methods used in these geological hazard research papers. This is very helpful for research, reuse, and reference on the geological hazard literature.

Table 7. Entities extracted from geological hazard research papers.

Paper Name	Location Entities	Methods Entities	Data Entities
Study on the Influence of Typical Human Activities on Slope Deformation and Stability	Southwestern China Slope area Goaf	Site survey Numerical simulation Numerical calculation Lagrangian difference method	Stratigraphic lithology Rainfall
Research on rainfall induced landslide prediction model	Chongqing area	Landslide prediction model Probabilistic prediction model Regression model Landslide probability model Landslide model Positive forecasting model	Rainfall Rainfall data Landslide data Rainfall
Research on Object-Oriented Methods for Extracting Geological Environment Information of Chengchao Iron Mine	Mountain	Fuzzy classification method Fuzzy classification Hierarchical network	Remote sensing data Remote sensing images Remote sensing Image data

Entities extracted from geological hazard research papers in Table 7 are in Chinese. Please refer to Appendix B for a detailed translation.

We used the proposed model to extract the three types of named entities from the 14,630 geological hazard-related research papers crawled on the Wanfang knowledge base, and constructed a knowledge graph containing 34,457 entities nodes and 84,561 relations. For the relationships ("in location," "use

methods," and "use data") that appear in the knowledge graph, in our article, we did not use any complicated relation extraction models. When constructing the knowledge graph, we simply think that if the paper contains an entity, it has a corresponding relationship. For example, if article A contains data B, we generate a triple (A-> "use data"-> B) and add it to the knowledge graph. Table 8 shows the detailed statistics of the entities of the geological hazards literature knowledge graph, and Table 9 shows the detailed statistics of the relations of the geological hazards literature knowledge graph. Figure 4 shows an overview of the geological hazard literature knowledge graph. For convenience, we only show 100 nodes in the knowledge graph and zoom in one of its parts in Figure 5. Obviously, the knowledge graph constructed can clearly reflect the relationship between literature and entities (methods, locations, and data).

Table 8. Statistics of the entities in geological hazards literature knowledge graph.

Entities Type	Methods	Location	Data	Paper
The number of the entities	8530	9123	2173	14,630

Table 9. Statistics of the relations in geological hazards literature knowledge graph.

Tags	In Location	Use Methods	Use Data	Paper Name
The number of the relations	24,934	25,364	19,633	14,630

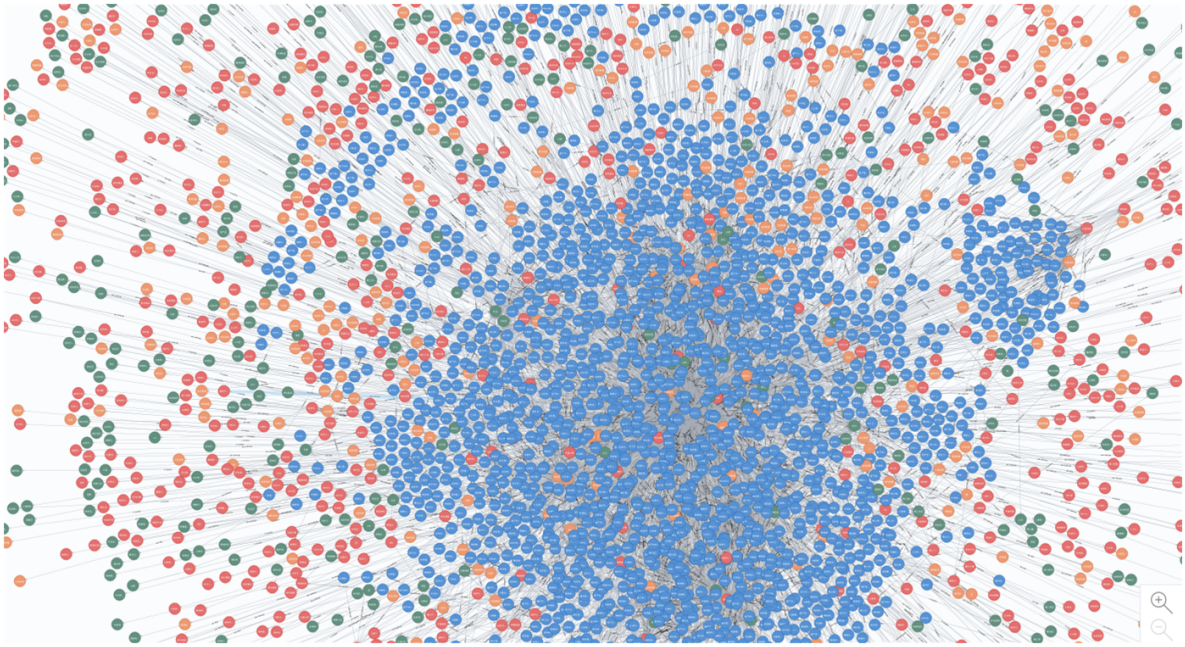

Figure 4. Geological hazard literature knowledge graph overview. The nodes of different colors in the figure represent different types of entities. (The blue nodes represent the name of paper, the red nodes represent the methods entities, the green nodes represent the location entities, and the orange nodes represent the data entities).

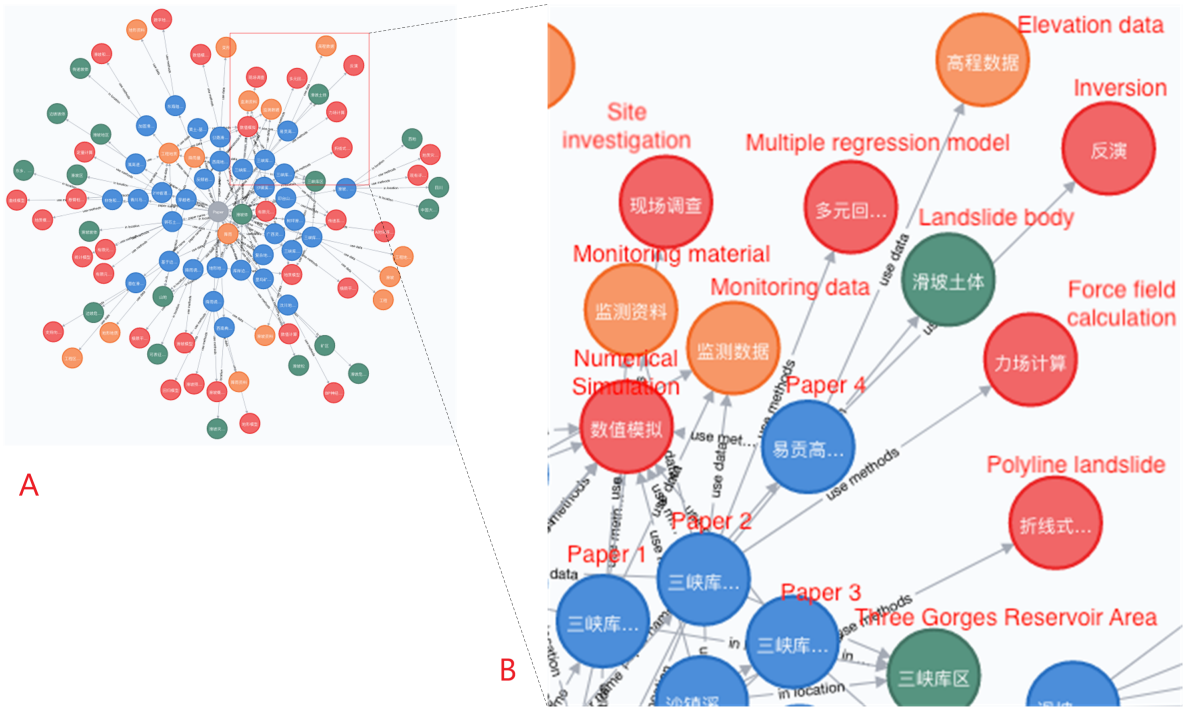

Figure 5. Partial zoom of geological hazard literature knowledge graph. (**A**) shows 100 nodes of the knowledge graph, and (**B**) is the zoom in version of the red part in (**A**).

At the same time, we counted the top-15 most frequently occurring entities of *methods*, *data*, and *location*, in the knowledge graph, which are shown in the Figures 6–8 with the corresponding English versions. It can be seen that in *methods* entities, the numerical simulation method is the most widely used research methods, with a frequency of 4542 times, and the number of other methods shows a smooth downward trend. In *data* entities, a similar phenomenon was also present. The rainfall data and vegetation data are the most widely used research data, respectively, with a frequency of 14,539 and 13,114 times. The other types of data are not very different, showing a trend of smooth decline. In *location* entities, mountainous areas, mining areas, and mountains are the most studied areas, with frequencies of 5172, 4354, and 4023, respectively, indicating that these three types of areas are the most significant areas of geological hazards.

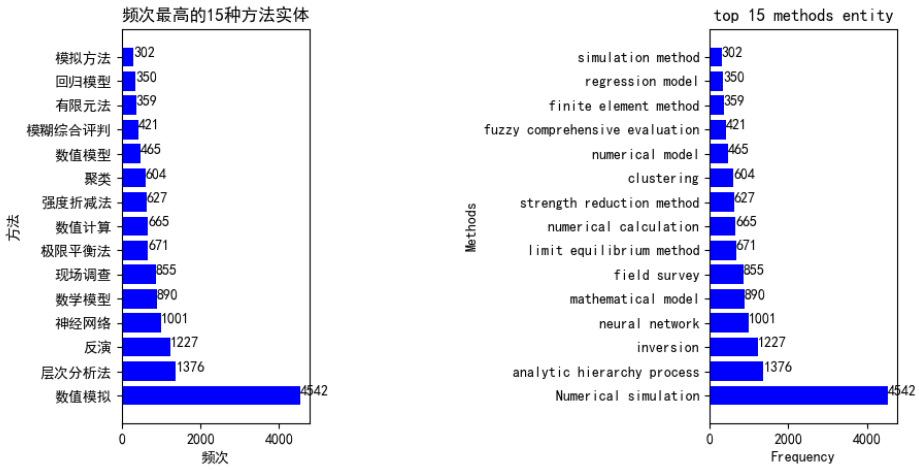

Figure 6. The top 15 methods entities in the knowledge graph.

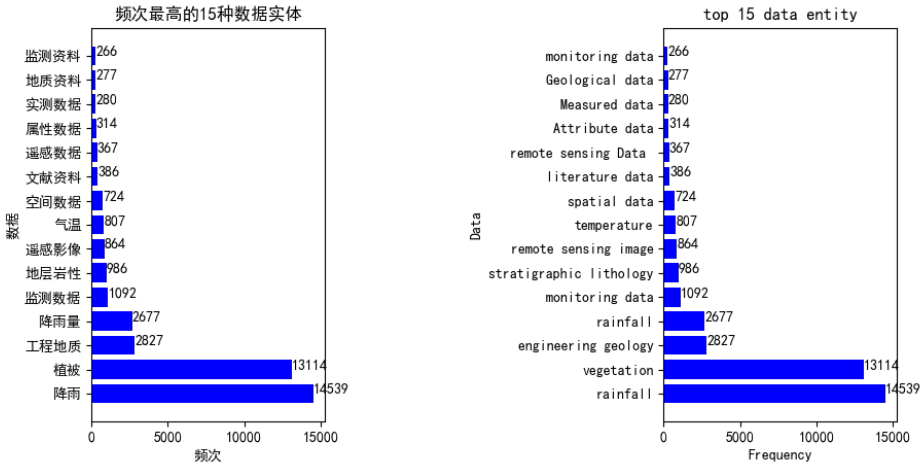

Figure 7. The top 15 data entities in the knowledge graph.

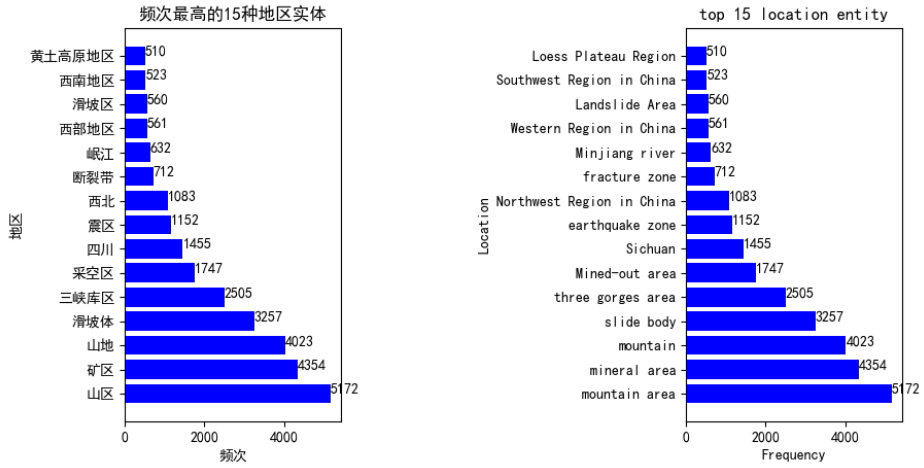

Figure 8. The top 15 location entities in the knowledge graph.

7. Discussion

7.1. Discussions of Generalizability

In this subsection, the outcomes generalizable to other contexts (e.g., use of papers written in English) are discussed in the following two aspects.

1. Paper structure. In our practice, we crawled the abstract parts of the articles named entity recognition. Therefore, our method has no special requirements for the structure of the article, so long as the article contains a complete summary section.
2. Paper language. In terms of language (in English, for example), the model needs to be adjusted as follows: Firstly, Chinese is based on characters, while English is based on words. Therefore, to extend our method to English papers, we need to rebuild the seed acquisition patterns (in Section 4.1.2) to build a training corpus for the model. Secondly, when doing NER tasks in Chinese, one character corresponds to one tag, but in English, one word corresponds to one tag. Therefore, to extend our method to English papers, we need to change the Chinese character vectors to the English word vectors in the embedding layer (in Section 4.2.1) of the deep, multi-branch BiGRU-CRF Model.

7.2. Discussions of Extensibility

In this subsection, the extensibility of the proposed methods is discussed in the following two aspects: the flexibility to accommodate new instances and the extensibility of the type of entities extracted from the paper.

1. The flexibility to accommodate new instances. When a new paper is added to the Wanfang database, the newly added papers can be processed into nodes and edges the following three steps. The first step: crawling the abstract part of the new papers from the Wanfang database through web crawler technology; the second step: using the deep, multi-branch BiGRU-CRF model to identify the method, data, and location entities; the third step: the entity acts as a node, and the connections between the entities and the papers act as edges to the knowledge graph.
2. The extensibility of the types of entity. At the same time, we also discussed what adjustments our methods need to be made if new entity types (e.g., theory) are added. If a new entity type is added, the deep, multi-branch BiGRU-CRF model needs to be adjusted as follows: First, we need to manually design the seed acquisition patterns (such as A) and build the training corpus using the methods mentioned in Section 4.1.2. Second, due to the addition of new entity types, the probability value of the softmax output of the last layer of our model needs to be changed from 7 ("O," "I-LDS," "I-MED," "B-LDS," "I-DAT," "B-MED," and "B-DAT") to 9 ("O," "I-LDS," "I-MED," "B-LDS," "I-DAT," "B-MED," "B-DAT," "B-THE," and "I-THE") in which "THE" represents the theory entity.

7.3. Discussions of Limitations and Future Work

This research, however, is subject to several limitations. In this subsection, some possible limitations are discussed in the following two aspects. The first limitation is that the proposed method involves some manual work. First of all, our approach needs to define some patterns to obtain the initial entity seed manually. And we also need to manually check the initial entity seed to get the correct entity seed collections in Section 4.1.2. The second limitation is that we only use the most straightforward method to obtain the relationships in our knowledge graph. That is, if article A contains data B, we generate a triple (A-> "use data"-> B) and add it to the knowledge graph.

Therefore, in future work, we believe that how to reduce manual costs is still an important research topic for the geological disaster knowledge graph construction. It would be a feasible method to reduce manual costs based on weak supervision and distant supervision strategies. At the same time, how to extract more accurate and diverse relationships, and even the joint extraction of entities and relationships are also important research topics.

8. Conclusions

Our work aims to extract geological hazard named entities from the considerable body of geological hazard literature and build a geological hazard knowledge graph. In this paper, a deep learning-based NER model, the deep, multi-branch BiGRU-CRF model, was proposed to extract the three types of entities (*location*, *methods*, and *data*) in geological literature and achieved the highest average precision of 0.9413, an average recall rate of 0.9425, and an average F1 score of 94.19. Besides, since the proposed model is a supervised model which requires a corpus, we proposed a pattern-based method to construct a large-scale geological hazard NER corpus. Finally, we used the proposed model to identify the entities in the 14,630 geological hazard-related research papers crawled on the Wanfang knowledge base and constructed a large-scale geological hazard literature knowledge graph containing 34,457 entities nodes and 84,561 relations. The following conclusions can be drawn: (1) The pattern-based method which uses some manually designed patterns combined with the MFM method can build an effective corpus with little manual costs. (2) The proposed deep learning-based NER model that combines a multi-branch BiGRU layer and a CRF model has the best results in the NER of geological hazards literature. (3) Knowledge graph technology can show the relationship

between papers and methods, locations, and data. It can facilitate the analysis and reuse of geological hazard literature.

Author Contributions: Conceptualization, Runyu Fan; data curation, Jining Yan, Weijing Song and Yingqian Zhu; formal analysis, Jining Yan, Weijing Song and Xiaodao Chen; methodology, Lizhe Wang; supervision, Lizhe Wang; validation, Yingqian Zhu; writing—original draft, Runyu Fan; writing—review and editing, Runyu Fan. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Natural Science Foundation of China (number U1711266).

Acknowledgments: The authors are grateful to the editors and reviewers for their valuable comments on the manuscript.

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

Appendix A

Since the data used in this paper are all from Chinese literature database, phrases in Table 2 in Section 4.1 and Table 7 in Section 6.4 are all in Chinese. For express convenience, we translate these phrases to English. In this appendices, we present the original phrases in Chinese associated with the Table 2 in Section 4.1, the Table 7 in Section 6.4 in this paper.

Table A1 shows the original patterns (regular expressions) in Chinese used in this work in Section 4.1. Since it is designed in Chinese, we translate it in English for better reading in Table A2.

Table A1. The original patterns (regular expressions) used.

Entity Type	Patterns (Regular Expressions)
Location	'.*(地处|位于|在|形成|处于)(了){0,1}([\S]+)(的){0,1}(地区|区域|山区|流域|区).*'
Methods	'.*(提供|使用|改进|利用|运用|提出|设计|发明|建立|构造|实现|根据|以 |基于|构建|结合|采取|采用|推广|通过) (了|的|于|对应的|出){0,1}(及){0,1}([\S]+)(的){0,1}(法|模型).*'
Data	'.*(提供|使用|利用|运用|提出|设计|发明|建立|构造 |根据|以|基于|构建|制作|结合|采取|采用|通过|构建|收集) (了|的|于|对|及){0,1}([\S]+)(的){0,1}(数据|资料|数据集).*'

Table A2. English translation of the pattern (regular expressions) used.

Entity Type	Patterns (Regular Expressions)
Location	'.*(located in|located in|in|form|located in)(of){0,1} ([\S]+)(of){0,1}(area|region|mountain area|river basin|zone).*'
Methods	'.*(provide|apply|improve|utilize|using|put forward|design|invente|set up | construct|achieve |according to|take |base on|construct |produce|combine |adopt |adopt|by|construct)(of|of|of|corresponding|of){0,1} (and){0,1}([\S]+)(of) {0,1}(method|model).*'
Data	'.*(provide|apply|utilize|using|put forward|design|invente|set up |construct|according to |take|base on|construct|produce|combine|adopt|adopt |by|construct|collect) (of|of|of|corresponding|and){0,1} ([\S]+)(of){0,1}(data|material|data set).*'

Appendix B

Table A3 shows the Chinese translation of the entities extracted from geological hazard research papers in Table A4 in Section 6.4.

Table A3. Entities extracted from geological hazard research papers in Chinese.

Paper Name	Location Entities	Methods Entities	Data Entities
典型人类活动对边坡变形及稳定性的影响研究	西南地区 边坡区 采空区	现场调查 数值模拟 数值计算 拉格朗日差分法	地层岩性 降雨
降雨诱发滑坡预测模型研究	重庆地区	滑坡预测模型 概率预测模型 回归模型 滑坡发生概率模型 滑坡模型 正预报模型	降雨 降雨资料 滑坡资料 降雨量
面向对象的程潮铁矿矿山地质环境信息提取方法研究	山地	模糊分类方法 模糊分类 层次网络	遥感数据 遥感影像 遥感 影像数据

Table A4. Entities extracted from geological hazard research papers.

Paper Name	Location Entities	Methods Entities	Data Entities
Study on the Influence of Typical Human Activities on Slope Deformation and Stability	Southwestern China Slope area Goaf	Site survey Numerical simulation Numerical calculation Lagrangian difference method	Stratigraphic lithology Rainfall
Research on rainfall induced landslide prediction model	Chongqing area	Landslide prediction model Probabilistic prediction model Regression model Landslide probability model Landslide model Positive forecasting model	Rainfall Rainfall data Landslide data Rainfall
Research on Object-Oriented Methods for Extracting Geological Environment Information of Chengchao Iron Mine	Mountain	Fuzzy classification method Fuzzy classification Hierarchical network	Remote sensing data Remote sensing images Remote sensing Image data

References

1. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, Ł.; 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; pp. 5998–6008.
2. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 26 June–1 July 2016; pp. 770–778.
3. Chowdhury, G.G. Natural language processing. *Annu. Rev. Inf. Sci. Technol.* **2003**, *37*, 51–89. [CrossRef]
4. Zhu, Y.; Zhou, W.; Xu, Y.; Liu, J.; Tan, Y. Intelligent learning for knowledge graph towards geological data. *Sci. Program.* **2017**, 2017; 5072427:1–5072427:13. [CrossRef]
5. Bauer, F.; Kaltenböck, M. *Linked Open Data: The Essentials*; Ed. Mono/Monochrom: Vienna, Austria, 2011.
6. Mihalcea, R.; Tarau, P. Textrank: Bringing order into text. In Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing, Barcelona, Spain, 25–26 July 2004.
7. Wang, C.; Ma, X.; Chen, J.; Chen, J. Information extraction and knowledge graph construction from geoscience literature. *Comput. Geosci.* **2018**, *112*, 112–120. [CrossRef]
8. Lafferty, J.; McCallum, A.; Pereira, F.C. Conditional random fields: Probabilistic models for segmenting and labeling sequence data. In Proceedings of the Eighteenth International Conference on Machine Learning (ICML 2001), Williamstown, MA, USA, 28 June–1 July 2001; pp. 282–289.
9. Powers, D.M. Applications and explanations of Zipf's law. In Proceedings of the Joint Conferences on New Methods in Language Processing and Computational Natural Language Learning, Sydney, Australia, 11–17 January 1998; Association for Computational Linguistics: Stroudsburg, PA, USA, 1998; pp. 151–160.

10. Ramos, J. Using tf-idf to determine word relevance in document queries. In Proceedings of the First Instructional Conference on Machine Learning, Piscataway, NJ, USA, 3–8 December 2003; Volume 242, pp. 133–142.
11. Shi, L.; Jianping, C.; Jie, X. Prospecting Information Extraction by Text Mining Based on Convolutional Neural Networks—A case study of the Lala Copper Deposit, China. *IEEE Access* **2018**, *6*, 52286–52297. [CrossRef]
12. Chinchor, N.; Robinson, P. MUC-7 named entity task definition. In Proceedings of the 7th Conference on Message Understanding, Frascati, Italy, 16 July 1997; Volume 29.
13. Yates, A.; Cafarella, M.; Banko, M.; Etzioni, O.; Broadhead, M.; Soderland, S. Textrunner: Open information extraction on the web. In Proceedings of the Human Language Technologies: The Annual Conference of the North American Chapter of the Association for Computational Linguistics: Demonstrations, New York, NY, USA, 23–25 April 2007; Association for Computational Linguistics: Stroudsburg, PA, USA, 2007; pp. 25–26.
14. Agichtein, E.; Gravano, L.; Pavel, J.; Sokolova, V.; Voskoboynik, A. Snowball: A prototype system for extracting relations from large text collections. In Proceedings of the International Conference on Digital Libraries, Kyoto, Japan, 13–16 November 2000.
15. Friburger, N.; Maurel, D. Finite-state transducer cascades to extract named entities in texts. *Theor. Comput. Sci.* **2004**, *313*, 93–104. [CrossRef]
16. Sundheim, B.M. Overview of results of the MUC-6 evaluation. In Proceedings of the 6th Conference on Message Understanding, Columbia, MD, USA, 6–8 November 1995; Association for Computational Linguistics: Stroudsburg, PA, USA, 1995; pp. 13–31.
17. Chinchor, N. Overview of MUC-7. In Proceedings of the Seventh Message Understanding Conference (MUC-7), Fairfax, VA, USA, 29 April–1 May 1998.
18. Chieu, H.L.; Ng, H.T. Named entity recognition: A maximum entropy approach using global information. In Proceedings of the 19th International Conference on Computational Linguistics, Taipei, Taiwan, 24 August–1 September 2002; Association for Computational Linguistics: Stroudsburg, PA, USA, 2002; Volume 1, pp. 1–7.
19. Borthwick, A.; Grishman, R. A Maximum Entropy Approach to Named Entity Recognition. Ph.D. Thesis, New York University, New York, NY, USA, 1999.
20. Curran, J.R.; Clark, S. Language independent NER using a maximum entropy tagger. In Proceedings of the Seventh Conference on Natural Language Learning at HLT-NAACL 2003, Edmonton, AB, Canada, 31 May–1 June 2003; Association for Computational Linguistics: Stroudsburg, PA, USA, 2003; Volume 4, pp. 164–167.
21. Hearst, M.A.; Dumais, S.T.; Osuna, E.; Platt, J.; Scholkopf, B. Support vector machines. *IEEE Intell. Syst. Their Appl.* **1998**, *13*, 18–28. [CrossRef]
22. Isozaki, H.; Kazawa, H. Efficient support vector classifiers for named entity recognition. In Proceedings of the 19th International Conference on Computational Linguistics, Taipei, Taiwan, 24 August–1 September 2002; Association for Computational Linguistics: Stroudsburg, PA, USA, 2002; Volume 1, pp. 1–7.
23. Kazama, J.; Makino, T.; Ohta, Y.; Tsujii, J. Tuning support vector machines for biomedical named entity recognition. In Proceedings of the ACL-02 Workshop on Natural Language Processing in the Biomedical Domain, Phildadelphia, PA, USA, 11 July 2002; Association for Computational Linguistics: Stroudsburg, PA, USA, 2002; Volume 3, pp. 1–8.
24. Ekbal, A.; Bandyopadhyay, S. Named entity recognition using support vector machine: A language independent approach. *Int. J. Electr. Comput. Syst. Eng.* **2010**, *4*, 155–170.
25. Zhou, G.; Su, J. Named entity recognition using an HMM-based chunk tagger. In Proceedings of the 40th Annual Meeting on Association for Computational Linguistics, Philadelphia, PA, USA, 7–12 July 2002; Association for Computational Linguistics: Stroudsburg, PA, USA, 2002; pp. 473–480.
26. Zhao, S. Named entity recognition in biomedical texts using an HMM model. In Proceedings of the International Joint Workshop on Natural Language Processing in Biomedicine and its Applications. Association for Computational Linguistics, Geneva, Switzerland, 28–29 August 2004; pp. 84–87.
27. Zhang, J.; Shen, D.; Zhou, G.; Su, J.; Tan, C.L. Enhancing HMM-based biomedical named entity recognition by studying special phenomena. *J. Biomed. Inform.* **2004**, *37*, 411–422. [CrossRef]

28. McCallum, A.; Li, W. Early results for named entity recognition with conditional random fields, feature induction and web-enhanced lexicons. In Proceedings of the Seventh Conference on Natural Language Learning at HLT-NAACL 2003, Edmonton, AB, Canada, 31 May–1 June 2003; Association for Computational Linguistics: Stroudsburg, PA, USA, 2003; Volume 4, pp. 188–191.
29. Settles, B. Biomedical named entity recognition using conditional random fields and rich feature sets. In Proceedings of the International Joint Workshop on Natural Language Processing in Biomedicine and Its Applications, Geneva, Switzerland, 28–29 August 2004; Association for Computational Linguistics: Stroudsburg, PA, USA, 2004; pp. 104–107.
30. Li, D.; Kipper-Schuler, K.; Savova, G. Conditional random fields and support vector machines for disorder named entity recognition in clinical texts. In Proceedings of the Workshop on Current Trends in Biomedical Natural Language Processing, Columbus, OH, USA, 19 June 2008; Association for Computational Linguistics: Stroudsburg, PA, USA, 2008; pp. 94–95.
31. Lample, G.; Ballesteros, M.; Subramanian, S.; Kawakami, K.; Dyer, C. Neural architectures for named entity recognition. *arXiv* **2016**, arXiv:1603.01360.
32. Chiu, J.P.; Nichols, E. Named entity recognition with bidirectional LSTM-CNNs. *arXiv* **2015**, arXiv:1511.08308.
33. Hammerton, J. Named entity recognition with long short-term memory. In Proceedings of the Seventh Conference on Natural Language Learning at HLT-NAACL 2003, Edmonton, AB, Canada, 27 May–1 June 2003; Association for Computational Linguistics: Stroudsburg, PA, USA, 2003; Volume 4, pp. 172–175.
34. Ma, X.; Hovy, E. End-to-end sequence labeling via bi-directional lstm-cnns-crf. *arXiv* **2016**, arXiv:1603.01354.
35. Xu, M.; Jiang, H.; Watcharawittayakul, S. A local detection approach for named entity recognition and mention detection. In Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), Vancouver, BC, Canada, 30 July–4 August 2017; Volume 1, pp. 1237–1247.
36. Zhao, D.; Huang, J.; Luo, Y.; Jia, Y. A Joint Decoding Algorithm for Named Entity Recognition. In Proceedings of the 2018 IEEE Third International Conference on Data Science in Cyberspace (DSC), Guangzhou, China, 18–21 June 2018; pp. 705–709.
37. Nguyen, T.V.T.; Moschitti, A.; Riccardi, G. Kernel-based reranking for named-entity extraction. In Proceedings of the 23rd International Conference on Computational Linguistics: Posters, Beijing, China, 23–27 August 2010; Association for Computational Linguistics: Stroudsburg, PA, USA, 2010; pp. 901–909.
38. Sobhana, N.; Mitra, P.; Ghosh, S. Conditional random field based named entity recognition in geological text. *Int. J. Comput. Appl.* **2010**, *1*, 143–147. [CrossRef]
39. Mikolov, T.; Karafiát, M.; Burget, L.; Černockỳ, J.; Khudanpur, S. Recurrent neural network based language model. In Proceedings of the Eleventh Annual Conference of the International Speech Communication Association, Makuhari, Chiba, Japan, 26–30 September 2010 .
40. Mikolov, T.; Kombrink, S.; Burget, L.; Černockỳ, J.; Khudanpur, S. Extensions of recurrent neural network language model. In Proceedings of the 2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Prague, Czech Republic, 22–27 May 2011; pp. 5528–5531.
41. Gers, F.A.; Schmidhuber, J.; Cummins, F. Learning to forget: Continual prediction with LSTM. In Proceedings of the 9th International Conference on Artificial Neural Networks: ICANN'99, Edinburgh, UK, 7–10 September 1999.
42. Sak, H.; Senior, A.; Beaufays, F. Long short-term memory recurrent neural network architectures for large scale acoustic modeling. In Proceedings of the Fifteenth Annual Conference of the International Speech Communication Association, Singapore, 14–18 September 2014.
43. Sundermeyer, M.; Schlüter, R.; Ney, H. LSTM neural networks for language modeling. In Proceedings of the Thirteenth Annual Conference of the International Speech Communication Association, Portland, OR, USA, 9–13 September 2012.
44. 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.
45. Chung, J.; Gulcehre, C.; Cho, K.; Bengio, Y. Empirical evaluation of gated recurrent neural networks on sequence modeling. *arXiv* **2014**, arXiv:1412.3555.
46. Chung, J.; Gulcehre, C.; Cho, K.; Bengio, Y. Gated feedback recurrent neural networks. In Proceedings of the International Conference on Machine Learning, Lille, France, 6–11 July 2015; pp. 2067–2075.

47. Dwibedi, D.; Sermanet, P.; Tompson, J.; Diba, A.; Fayyaz, M.; Sharma, V.; Hossein Karami, A.; Mahdi Arzani, M.; Yousefzadeh, R.; Van Gool, L.; et al. Temporal Reasoning in Videos using Convolutional Gated Recurrent Units. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Salt Lake City, UT, USA, 18–22 June 2018; pp. 1111–1116.
48. LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436. [CrossRef]
49. Schmidhuber, J. Deep learning in neural networks: An overview. *Neural Netw.* **2015**, *61*, 85–117. [CrossRef]
50. Graves, A.; Mohamed, A.R.; Hinton, G. Speech recognition with deep recurrent neural networks. In Proceedings of the 2013 IEEE International Conference on Acoustics, Speech and Signal Processing Vancouver, BC, Canada, 26–31 May 2013; pp. 6645–6649.
51. Hecht-Nielsen, R. Theory of the backpropagation neural network. In *Neural Networks for Perception*; Elsevier: Amsterdam, The Netherlands, 1992; pp. 65–93.
52. Bengio, Y.; Simard, P.; Frasconi, P. Learning long-term dependencies with gradient descent is difficult. *IEEE Trans. Neural Netw.* **1994**, *5*, 157–166. [CrossRef]
53. Pascanu, R.; Mikolov, T.; Bengio, Y. On the difficulty of training recurrent neural networks. In Proceedings of the International Conference on Machine Learning, Atlanta, GA, USA, 16–21 June 2013; pp. 1310–1318.
54. Ratnaparkhi, A. A maximum entropy model for part-of-speech tagging. In Proceedings of the Conference on Empirical Methods in Natural Language Processing, Philadelphia, PA, USA, 17–18 May 1996.
55. Baum, L.E.; Petrie, T. Statistical inference for probabilistic functions of finite state Markov chains. *Ann. Math. Stat.* **1966**, *37*, 1554–1563. [CrossRef]
56. Zheng, S.; Jayasumana, S.; Romera-Paredes, B.; Vineet, V.; Su, Z.; Du, D.; Huang, C.; Torr, P.H. Conditional random fields as recurrent neural networks. In Proceedings of the IEEE International Conference on Computer Vision, Santiago, Chile, 7–13 December 2015; pp. 1529–1537.
57. Christ, P.F.; Elshaer, M.E.A.; Ettlinger, F.; Tatavarty, S.; Bickel, M.; Bilic, P.; Rempfler, M.; Armbruster, M.; Hofmann, F.; D'Anastasi, M.; et al. Automatic liver and lesion segmentation in CT using cascaded fully convolutional neural networks and 3D conditional random fields. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Athens, Greece, 17–21 October2016; pp. 415–423.
58. Hoberg, T.; Rottensteiner, F.; Feitosa, R.Q.; Heipke, C. Conditional random fields for multitemporal and multiscale classification of optical satellite imagery. *IEEE Trans. Geosci. Remote. Sens.* **2015**, *53*, 659–673. [CrossRef]
59. Li, K.; Ai, W.; Tang, Z.; Zhang, F.; Jiang, L.; Li, K.; Hwang, K. Hadoop recognition of biomedical named entity using conditional random fields. *IEEE Trans. Parallel Distrib. Syst.* **2015**, *26*, 3040–3051. [CrossRef]
60. Sutton, C.; McCallum, A. An introduction to conditional random fields. *Found. Trends® Mach. Learn.* **2012**, *4*, 267–373. [CrossRef]
61. Marsh, E.; Perzanowski, D. MUC-7 evaluation of IE technology: Overview of results. In Proceedings of the Seventh Message Understanding Conference (MUC-7), Fairfax, Virginia, 29 April–1 May 1998.
62. Kudo, T. CRF++: Yet Another CRF Toolkit. Available online: http://crfpp.sourceforge.net/ (accessed on 22 December 2019).
63. Elkan, C. Log-linear models and conditional random fields. *Tutor. Notes CIKM* **2008**, *8*, 1–12.
64. Srivastava, N.; Hinton, G.; Krizhevsky, A.; Sutskever, I.; Salakhutdinov, R. Dropout: a simple way to prevent neural networks from overfitting. *J. Mach. Learn. Res.* **2014**, *15*, 1929–1958.
65. Nadeau, D.; Sekine, S. A survey of named entity recognition and classification. *Lingvisticae Investig.* **2007**, *30*, 3–26.