applied sciences

MDPI

Article

Intelligent Recognition of Key Earthquake Emergency Chinese Information Based on the Optimized BERT-BiLSTM-CRF Algorithm

Zhonghao Wang, Meng Huang *, Chenxi Li, Jilin Feng, Shuai Liu and Guang Yang

Institute of Disaster Prevention, Beijing 065201, China
* Correspondence: hm@cidp.edu.cn

Abstract: To address the problems of the text of earthquake emergency information keeps changing incrementally with the time of an earthquake's occurrence and there being more and more information categories, thus making it difficult to identify earthquake emergency key information, this paper proposes an intelligent recognition algorithm of earthquake emergency information based on the optimized BERT-BiLSTM-CRF algorithm. Based on the historical seismic emergency information dataset of the past 10 years, first, the BIO sequence labeling method is used to classify the seismic entities, and the BERT pretraining model is constructed to represent the seismic emergency text with sentence-level feature vectors. The BiLSTM algorithm is used to obtain the contextual information of the bidirectional seismic emergency text, and we introduce the attention mechanism to enhance the recognition effect of the seismic emergency key information in the statements. Finally, we use conditional randomization to enhance the recognition of earthquake emergency key information in the utterance. The conditional randomization algorithm is applied to extract the dependency relationship between adjacent vectors and improve the accuracy identification to realize the intelligent recognition of earthquake emergency information. The experimental results show that our model can extract earthquake emergency information from online media efficiently and accurately, with better performance than other baseline models.

Keywords: earthquake emergency information; named entity identification; BERT-BiLSTM; intelligent identification

check for updates

Citation: Wang, Z.; Huang, M.; Li, C.; Feng, J.; Liu, S.; Yang, G. Intelligent Recognition of Key Earthquake Emergency Chinese Information Based on the Optimized BERT-BiLSTM-CRF Algorithm. *Appl. Sci.* **2023**, *13*, 3024. https://doi.org/10.3390/app13053024

Academic Editor: José Salvador Sánchez Garreta

Received: 18 January 2023
Revised: 20 February 2023
Accepted: 24 February 2023
Published: 26 February 2023

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

1. Introduction

After a destructive earthquake, there is a large amount of earthquake news information in the network (special network information of the earthquake industry, news media websites, microblogs, forums, etc., as after a destructive earthquake, there can be tens of millions of relevant news pieces). At the same time, network news is also used as an important auxiliary information source for emergency decision making after an earthquake. The important Internet data application researched in this paper is the use of online auxiliary information for making emergency decisions after a major earthquake.

The volume and variety of emergency-related information that appears in text on the Internet after an earthquake are substantial. This information can clearly be helpful in organizing and directing emergency responses. Soon after an earthquake occurs, the official media will release the first characterization of the earthquake, giving three critical facts: time, location, and magnitude. This information usually appears in short texts. Over time, emergency earthquake information will proliferate, reporting casualties, economic loss, rescue information, and so on. Not only will the available texts appear with increasing frequency and from multiple sources, but the earthquake information texts themselves will increasingly transition from being short, minimal texts to being long texts, often containing information that is superfluous to emergency needs.

These later, longer texts will certainly contain key emergency information, such as casualty, rescue personnel, affected population, location, and urgency data, but it will be mixed with non-essential data about history, human interest, etc., causing difficulties in identifying critically urgent earthquake emergency information. Thus, these texts can fail to provide a reliable basis for on-time rescue decisions and other important relief decisions, significantly reducing the role and value of the texts in maximizing the timeliness and efficacy of earthquake emergency rescue and relief decisions.

In recent years, continual reform of natural language processing technology has achieved many substantial successes in various fields to the point where it is now possible to utilize natural language processing technology in the field of emergency response to major earthquakes. One of the key technologies in the advance of natural language processing has been named-entity recognition (NER), a fast way to obtain structured data, eliminate spurious data, and thus avoid information overload from the sheer volume of textual material as well as recognize the names of entities such as people, places, and time in the texts.

To serve the emergency needs of the earthquake field, any analyzing mechanism must recognize, in any earthquake-related text, key information such as the time of occurrence, location of occurrence and emergent critical needs, and casualties. NER algorithms accomplish this task by using various neural networks.

1.1. Convolutional Neural Network

In deep learning, a convolutional neural network (CNN, or ConvNet) is a class of artificial neural network (ANN) most commonly applied to analyze visual imagery [1]. The convolutional neural network (CNN) was first introduced to natural language processing by Kim in 2014, who demonstrated the effectiveness of CNNs in natural language processing classification tasks [2]. In 2015, Yao et al. proposed a CNN-based NER method suitable for medical text content that does not require building a lexicon yet ensures high accuracy [3]. In 2017, Strubell et al. proposed an iterated dilated CNN (IDCNN) NER method, similar to the most current expressive long short-term memory (LSTM) model mentioned below. The model requires only O(n) time complexity and can achieve an eightfold speedup while maintaining accuracy comparable to LSTM [4]. In 2018, Yang et al. used single-layer and multilayer CNNs at the word level and sentence level, respectively, to improve the accuracy of the model [5]. In 2021, Kong et al. proposed the use of GPU parallelism to improve the efficiency of the model [6]. In 2022, Junting Lin et al. proposed a CNN augmented by the bidirectional LSTM and conditional random field (CNN-BiLSTM-CRF) model based on a multi-headed self-attention mechanism to improve the accuracy of the model by fusing the learning features of multiple dimensions through a multi-headed attention mechanism [7]. In 2022, Sornlertlamvanich et al. added the CNN model to the BiLSTM-CRF model, fused the word vectors generated by BiLSTM with the word vectors generated by CNN models, and validated them in a THAI-NEST corpus, improving the model accuracy by 16.2% [8]. These studies demonstrated the effectiveness of convolutional neural networks in natural language processing and NER tasks and improved their network results to improve the training speed of the model. However, since CNNs cannot capture long-range text semantic features, they are not directly applicable to long-text earthquake emergency information.

1.2. Recurrent Neural Network

A recurrent neural network (RNN) is a type of recurrent neural network that takes sequence data as input, recurses in the evolution direction of the sequence, and connects all nodes in a chain [9]. The recurrent neural network (RNN) can also be used for NER, and a variant of the RNN, LSTM, has achieved remarkable success in NER. In 2015, Huang et al. applied the BiLSTM-CRF model to a natural language processing benchmark sequence-tagging dataset [10]. In 2018, Zhang et al. proposed a lattice LSTM model for Chinese NER. The best results were achieved by explicitly exploiting word sequence information compared with character-based approaches [11]. In 2019, Han et al. addressed the problem

that NER in professional domains usually faces a lack of in-domain annotation data by combining generative adversarial networks with long- and short-term memory network models, which significantly outperformed other models in all metrics [12]. In 2021, Eligüzel et al. used Glove word embedding combined with the RNN model to extract key information about the 2015 Nepal earthquake from Twitter postings, with good results. In 2021, Y. Zhao et al. constructed a character-level vector-based BiLSTM and CRF aviation safety event entity recognition model, using the BiLSTM model to obtain the contextual features of the text and to verify the output entity with CRF label consistency, and they improved the effectiveness of agency recognition in the aviation safety event domain. In 2022, Warto et al. used the RNN-BiLSTM-CRF model with the CoNLL2003 dataset to compare gradient descent (SGD), Adaptive Moment Estimation (Adam), and Adadelta. These three optimization algorithms were compared, and the results showed that the Adam algorithm outperformed the other two algorithms [13]. In 2022, Chen et al. proposed a novel binary tree model (BTPK) based on the TPK model combined with the Bi-RNN model, and the experimental results showed that the BTPK model outperformed the other Bi-RNN models in terms of self-focus [14]. These studies utilized the ability of recurrent neural networks to extract contextual semantic information, as well as word sequence information in both the positive and negative directions, to improve the accuracy of NER. For textual seismic emergency information, the limitation on the method's text length prevents recurrent neural networks from achieving good results.

1.3. The Transformer Model

The Transformer model was proposed in 2017. It no longer used the traditional neural network, using only the attention mechanism [15–17]. The BERT model was proposed in 2018 and has achieved impressive results in various fields of natural language processing [18]. The models based on recurrent neural networks, volume-based neural networks, and attention mechanism-based mechanisms have their advantages and disadvantages. Based on the strengths and weaknesses of the several approaches, many scholars have combined various models for NER research [19]. In 2019, Dai et al. used the network structure of BERT+BiLSTM+CRF for the application of Chinese electronic medical record form recognition and achieved good results [20]. Yoon et al. proposed a novel NER model consisting of multiple BiLSTM networks, where each network acts as a separate task to recognize a formulated entity type and multiple tasks transfer their learned knowledge to obtain more accurate predictions [21]. In 2020, Li et al. used a multilayer variant network structure for Chinese clinical NER and similarly achieved good recognition results [22]. In the literature, a pre-trained BERT model combined with BiLSTM was used to improve the accuracy of NER on the Weibo Chinese dataset [23]. Li et al. proposed FLAT to address the complexity of the existing lattice LSTM structure, improving both performance and efficiency [24]. In 2021, Cui et al. proposed a BERT and template NER method that treats the original sentence and the utterance template as the original sequence and the target sequence, respectively, for inferential classification, outperforming the BERT benchmark model by 10.88%, 15.34%, and 11.73% in the three datasets of MIT Movie, MIT Restaurant, and ATIS, respectively [25]. In 2022, An et al. proposed a bidirectional long- and short-term memory conditional random field model based on a multi-headed attention mechanism and evaluated it on two CCCS benchmark datasets, achieving the best performance yet compared with deep neural networks [26]. In 2022, Jeon et al. proposed an NER method for automatically extracting defect information from complex texts using a defect lexicon and migration learning, obtaining better performance than the benchmark model's results [27]. These studies show that in domain-specific problems, the combination of multiple models for their domain text information characteristics has better feature extraction results than using a single model. Therefore, it is necessary to analyze the characteristics of the earthquake domain text and design and combine models to extract the key information for earthquake emergencies.

1.4. Results Overview

Based on the above research, this paper proposes an intelligent recognition algorithm for earthquake emergency information based on a BERT pretrained language model to extract earthquake emergency information from text more efficiently and accurately. The algorithm is based on the BERT model for the feature representation of earthquake information in text and uses a bidirectional neural network combined with a self-attention mechanism to solve the problems of contextual semantic information in long texts, the varying length of input vectors, and the varying length of earthquake emergency texts.

2. Historical Earthquake Emergency Information Dataset

For the NER task, most of the currently existing public datasets are oriented to the general domain, such as the CLUENER2020 Chinese fine-grained NER dataset or ACE2005 English dataset. There are also some NER datasets in some specific domains, but there are fewer public data for the earthquake emergency domain. For the field of earthquake emergency response, we refer to the research of Bai Xianfu et al. [28] on earthquake emergency site information for identifying and annotating the key information of earthquake emergency response, and the specific key information is shown in Table 1.

Table 1. Earthquake information labeling table.

Earthquake Key Information	Label Name	Number of Labels
Seismic time	TIM	3531
Magnitude	MAG	4849
Earthquake location	PLA	5679
Seismic coordinates	COO	851
Depth of earthquake	DEP	1119
Intensity	INT	164
Number of injured	HUR	757
Number of deaths	DEA	794
Number of firefighters	FIR	230
News release time	NET	210
Number of materials	MAT	247
Geologic conditions	GEO	212
Number of rescued	RES	278
Description of secondary disasters	DIS	129
Economic loss	ECO	117
Number of volunteers	VOL	26
Number of medical staff	DOC	19
Earthquake population	PEO	45
Number of transferred persons	TRA	37
Total		19,294

In this paper, we establish the historical earthquake emergency dataset through two steps: data acquisition and processing and data annotation.

2.1. Data Acquisition and Processing

In order to establish the needed dataset, we used a Python web crawler combined with the Requests framework to access and crawl the earthquake-related news released by Xinhua, the China Earthquake Network, the CCTV news network, and microblogs, and then we used GeneralNewsExtractor, a text- and symbol density-based web body extraction library, to obtain the text data, such as headlines and body text, from news websites and build an earthquake history corpus. The procedures extracted 4719 texts for the emergency earthquake-related corpus. The text length distribution is shown in Figure 1.

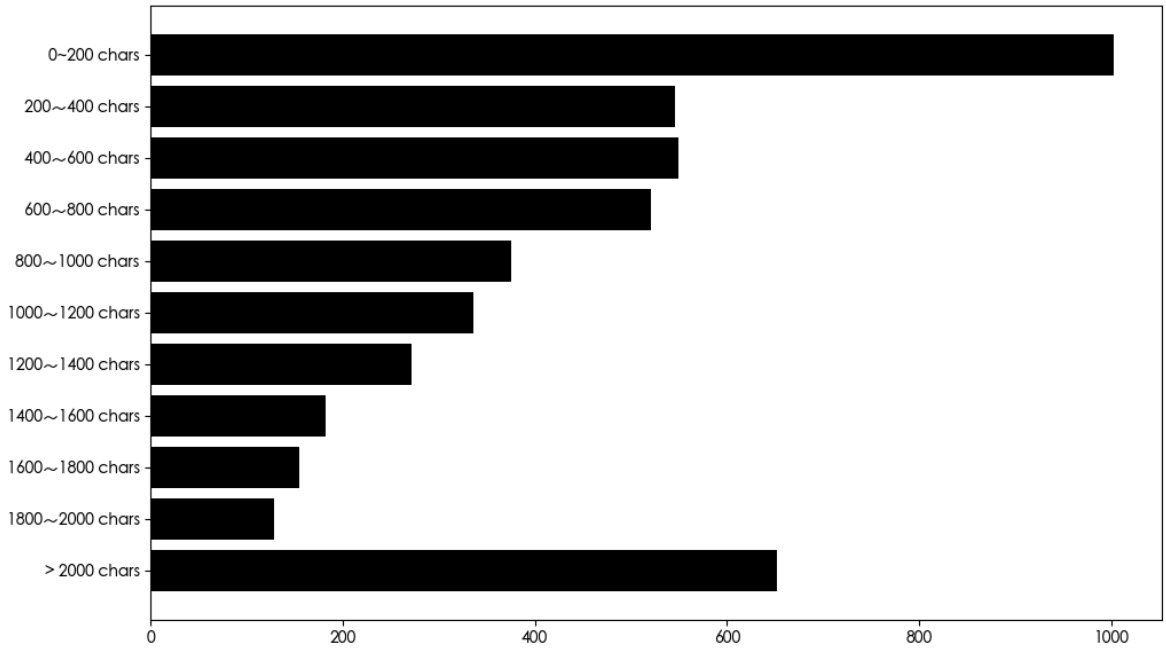

Figure 1. Earthquake-related emergency text length distribution.

2.2. Data Annotation

Seismic NER is a sequence classification problem which is usually annotated using the sequence annotation method of beginning–inside–outside (BIO) (coding to B-NP (the beginning of a noun phrase), I-NP (the middle of a noun phrase), and O (not a noun phrase)). The 4719 data texts in the earthquake historical emergency corpus were annotated according to the 19 classified categories of key information, and 19,294 key earthquake information samples were annotated, as shown in Table 1. After that, a corpus of earthquake news entities was constructed based on the annotated key earthquake information.

3. BERT-BiLSTM-Based Intelligent Recognition Algorithm for Earthquake Information

The texts of earthquake emergency information are different from other texts because the disaster information needs to be released for the first time after an earthquake. Usually, the text is relatively concise and contains a large amount of earthquake emergency information, such as the magnitude, time, location, and seismic fracture. Subsequent releases of information on casualties, economic losses, and rescue teams, involving a wide range of aspects and contents, are usually longer. Meanwhile, the texts contain numerous digital fragments, such as the magnitude, intensity, and source depth. These digital fragments often have the same structure but do not convey the same meaning. Traditional deep learning models focus more on character- and word-level features and ignore long-range semantic information, leading to an inability to extract semantic information and solve the problem of multiple meanings for a word [10]. To resolve these and other issues, we propose the BERT-BiLSTM four-layer model, as shown in Figure 2.

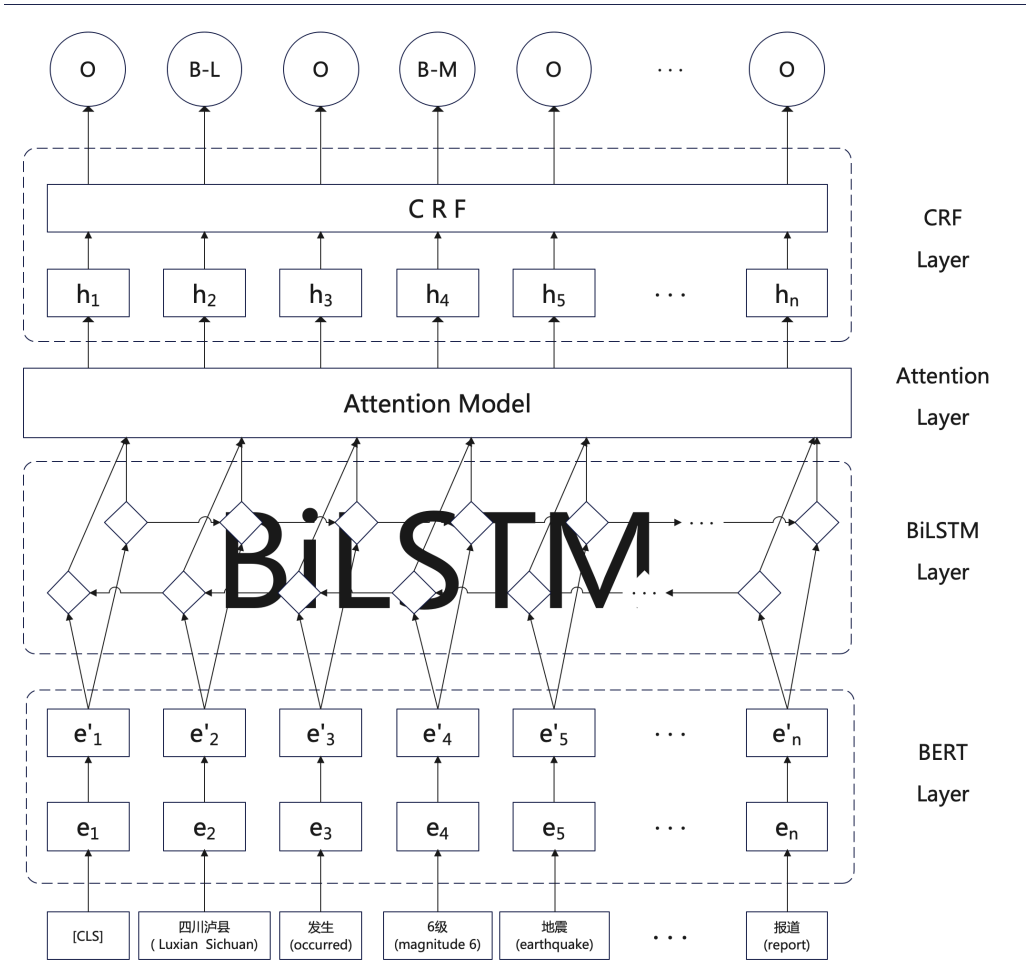

Figure 2. Model structure diagram.

The first layer of the model is the BERT layer, which converts each word in the input sentence into a vector. The second layer is the BiLSTM layer, which extracts the contextual semantic features and positional features of the vectors. The third layer is the self-attention layer, which extracts the relevant features of different vectors, solving the problem of an uncertain number of input vectors (i.e., the varying length of the earthquake emergency text). The final layer is the CRF layer, which considers the order relationship and dependency between vectors to make up for the lack of dependency relationships between adjacent vectors in recurrent neural networks, and it obtains the global optimal labeling we have developed to achieve intelligent recognition of earthquake emergency information.

3.1. Bert Layer

The BERT model is a masked language model based on a multi-headed attention mechanism that randomly masks tokens in the pretraining phase and predicts tokens based on the textual context of earthquake emergency information to better interpret the semantic meaning of words. The semantic information of the words is retained with a certain probability so that the information is not completely masked. Thus, the model makes a prediction based on the context of the rare words when they appear, solving the problem of rare words and semantic complexity. Unlike the traditional word2vec and fastText models, which generate fixed and unchanging word embeddings, BERT can be fine-tuned according to the downstream tasks accessed, thus achieving better results according to the tasks.

We used the historical earthquake emergency information dataset for training based on the Chinese BERT pretraining model, using word granularity for slicing to avoid the

overfitting problem and the out-of-vocabulary (OOV) problem caused by the sparse vector data because of word granularity.

3.2. Masked Language Model (MLM) and Next Sentence Prediction (NSP)

The MLM is performed by randomly selecting a certain percentage of words and replacing them with masked tokens (MASK), followed by the prediction of the masked words based on semantic features, as shown in Figure 3.

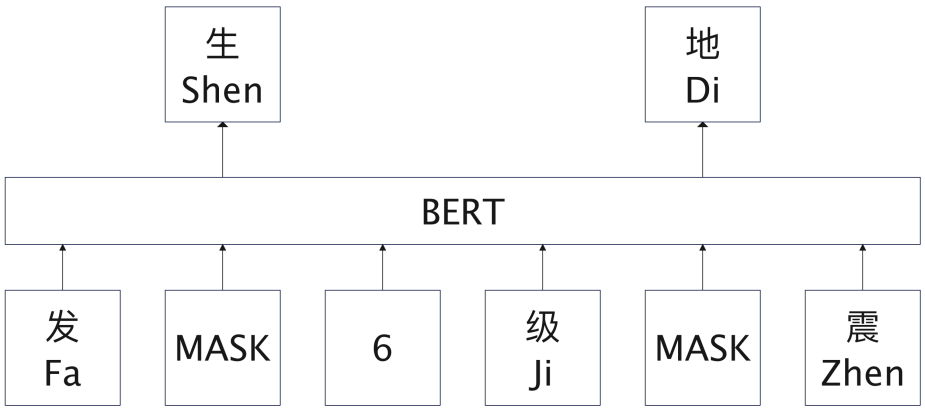

Figure 3. BERT model masking mechanism.

The earthquake news text is input into the BERT model, and the prediction is performed by the model to finally output an $n * d$ vector matrix S with elements $S_i = \{w_{i1}, w_{i2}, w_{i3}, \ldots, w_{ik}, \ldots, w_{in}\}$, where w_{ik} denotes the kth word in the ith sentence and n is the length of the sentence. The words in the sentence are converted into a d-dimensional vector that is made into a word embedding matrix [29]. The word embedding matrix will be used as an input to the BiLSTM model for contextual semantic information extraction.

3.3. BiLSTM Layer

BiLSTM, the bi-directional long short-term memory neural network, solves the problem of not being able to encode information from back to front based on long short-term memory (LSTM) by capturing longer-distance dependencies. LSTM's unit structure is shown in Figure 4.

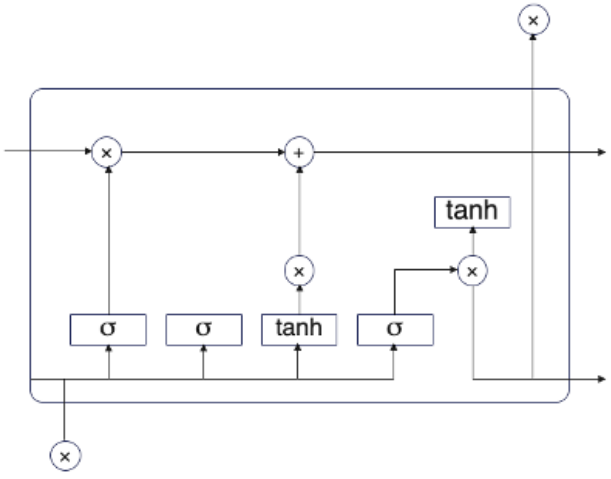

Figure 4. LSTM cell structure diagram.

The information to be forgotten and the information to be remembered are selected by calculating the forgetting gate and the memory gate, respectively. Their inputs are both the implicit state h_{t-1} at the previous moment and the input word X_t at the current moment.

The output value of the forgetting gate is the value f_t of the forgetting gate, and the output value of the memory gate is the value i_t of the memory gate and the temporary cell state i_t.

The formula for calculating the forgetting gate is as follows. W in the formula is a weight parameter, and b is a bias parameter:

$$f_t = \sigma(W_f \cdot [h_{t-1}, X_t] + b_f) \tag{1}$$

The formula for calculating the memory gate is as follows:

$$f_i = \sigma(W_f \cdot [h_{t-1}, X_t] + b_i) \tag{2}$$

$$\widetilde{C}_t = tanh(W_c \cdot [h_{t-1}, X_t] + b_c) \tag{3}$$

After that, the current state of the cell is calculated with the input of the value of the memory gate i_t and the values of the forgetting gate f_t, the temporary cell state $\widetilde{C}_t$, and the previous moment cell state C_{t-1}. The output is the cell state C_t at the current moment. The calculation formula is as follows:

$$C_t = f_t * C_{t-1} + i_t * \widetilde{C}_t \tag{4}$$

Finally, the output gate and the hidden state of the current moment are calculated, and the inputs are the hidden state h_{t-1} of the previous moment, the input word X_t of the current moment, and the cell state C_t of the current moment.

The calculation formula is as follows:

$$o_t = \sigma(W_o[h_{t-1}, X_t] + b_o) \tag{5}$$

$$h_t = o_t * tanh(C_t) \tag{6}$$

The overall structure of the BiLSTM model is shown in Figure 5.



Figure 5. LSTM cell structure diagram.

The hidden vector ht output by the BiLSTM model can capture the long-range contextual information in the earthquake emergency information text, but there will be a large number of repetitive words in the earthquake emergency information text, and the degree of the same words contributing to semantics in different sentences will be different, resulting in low recognition accuracy for the model. To solve this problem, in this paper, based on the BERT-BiLSTM-CRF model, an attention layer is added between the CRF layer and the BiLSTM layer to assign weights to the word vectors separately.

3.4. Attention Optimization Layer

The meaning of earthquake emergency information text not only relies on contextual information but also has a high correlation with the meaning of words in the text. However, not all words in an earthquake emergency information text contribute to the semantics of the whole sentence to the same extent. To solve this problem, a self-attention mechanism is used to extract these more important words of earthquake emergency information text and

give them higher weights [30]. Specifically, when the BiLSTM model is finished running, to generate the hidden vector h_t, h_t is input into the attention layer.

In the attention layer, h_t is first fed into a simple multilayer perceptron to obtain the new hidden vector u_t:

$$u_t = tanh(W_w h_t + b_w) \tag{7}$$

Then, the weight value of each word in the earthquake news text is computed by u_t and a word-level context vector u_w [30]. Here, u_w is considered a high-dimensional representation for determining the importance of different words in a sentence which is randomly initialized and learned jointly during the training process:

$$\vartheta_t = \frac{exp(u_t^T u_w)}{\sum_t exp(u_t^T u_w)} \tag{8}$$

Finally, the weight values of each word are weighted with the hidden vector h_t for a weighted average operation, which is calculated by the SoftMax function:

$$S = \sum_t \vartheta_t h_t \tag{9}$$

3.5. CRF Layer

In the CRF conditional random field algorithm, CRF will predict the named entity vector of text locations based on the incoming global feature vector and the named entity vector generated by the previous CRF unit by using the CRF representation of the dependency between the seismic named entity character labels and the global character labels with the following prediction equation:

$$S(X, y) = \sum_{i=0}^{n} A_{y_i, y_{i+1}} + \sum_{i=1}^{n} P_{i, y_i} \tag{10}$$

where $A_{y_i, y_{i+1}}$ is the transfer score between adjacent words in the earthquake news text and P_{i, y_i} is the ratio of the total score of correct annotation of characters in the earthquake news text to the total score for all possible annotations to obtain the probability of correct prediction of entity characters, and the formula for this is shown below:

$$p(y|X) = e^{S(X,y)} / \sum y \in Y_x^{e^{S(x,y)}} \tag{11}$$

where Y_x denotes all possible seismic named entity character annotations. During the training of the seismic NER model, the value of $p(y|X)$ is close to one, which indicates that the predicted entity labels and annotations are consistent, and the best result for seismic NER is achieved [12].

3.6. Summary of the Model

Based on the characteristics of the earthquake information text content summarized in the previous paper (i.e., texts of various lengths and more digital fragments), this paper proposes a BERT-BiLSTM-based intelligent recognition method for an earthquake information text to make it more applicable to the field of earthquake information text language. By capturing the long-range contextual information in the earthquake emergency information text through BiLSTM, the model's processing ability is enhanced when facing earthquake information texts of various lengths, and the attention optimization layer is introduced to enable the model to focus on the feature expression of word-to-sentence contributions, including numbers, which often have important significance to the earthquake information text, thus improving the accuracy of model recognition.

4. Experimental

4.1. Experimental Data and Preprocessing

In this experiment, the 4719 element corpus, with 19,294 earthquake key messages in the historical earthquake emergency information dataset described above, was divided into a training set and a validation set at a ratio of 7:3. The training set contained 13,505 key earthquake messages, and the validation set contained 5789 key earthquake messages. In addition, we crawled and annotated the texts of earthquake information from earthquakes of a magnitude of 6.1 in Lushan County, Ya'an, Sichuan on 1 June 2022, 6.0 in Markang City, Aba Prefecture, Sichuan on 10 June 2022, and 6.8 in Luding County, Ganzi Prefecture, Sichuan on 5 September 2022, obtaining 943 corpus elements annotated with 3989 key earthquake messages as a test set, which we included as a training set, and ensured that each model was trained using the same training set.

The training set was used to train the model, the validation set was used to validate the model, and the hyperparameters of the model were adjusted with reference to the validation results to make the model work best on the validation set. After that, the test set was used to evaluate the final effect of the model.

The data annotated in the historical earthquake emergency information dataset are shown in Table 2.

Table 2. BIO Chinese labeling method.

Sentences	2019年6月17日22时55分，在四川宜宾市长宁县发生6.0级地震，震源深度16千米。 (At 22:55 on 17 June 2019, a 6.0 magnitude earthquake occurred in Changning County, Yibin City, Sichuan, at a depth of 16 km.)
Labeling	[B-TIME] [I-TIME], 在[B-SITE] [I-SITE] 发生[B-MAG] [I-MAG]级地震，震源深度[B-DEP] [I-DEP] 。 (At [B-TIME][I-TIME], a [B-MAG][I-MAG] magnitude earthquake occurred in Changning County, Yibin City, Sichuan, at a depth of [B-DEP][I-DEP] km.)

4.2. Experimental Parameters

An experimental environment of Python 3.8 + PyTorch 1.9.1 was used to train and test the model. The experiments used the Engine-BERT-WWM model architecture, which contained 12 Transformer layers, 768 dimensional hidden layers, and a 12 head multi-head attention mechanism. In addition, other parameters such as access to the BiLSTM layer were set as shown in Table 3.

Table 3. Model parameter settings.

BiLSTM Hidden Layer Dimension	Number of BiLSTM Hidden Layers	Maximum Sequence Length	Learning Rate	Dropout
256	5	512	5×10^{-5}	0.2

4.3. Evaluation Criteria

In classification problems, the accuracy, precision, recall, and $F1$ value of the results (H-mean value) are usually used as evaluation metrics for classification results [31]. This study was performed as a dichotomous classification task, so the corresponding classes were C_1 and C_2. The class to be identified was taken as the positive class and vice versa for the negative class. The results predicted by the model on the dataset were classified as correct or incorrect. Thus, there were four cases, and the corresponding confusion matrix is shown in Table 4.

Table 4. Confusion matrix table.

Actual Value	Predicted to Be a Positive Example	Predicted to Be a Counter Example
Truth is a positive example	TP	FN
Truth is a counter example	FP	TN

Accuracy (Acc) is the ratio of the amount of data for which all predictions are correct to the amount of all data and is calculated as follows:

$$Acc = \frac{TP + TN}{TP + TN + FP + FN} \tag{12}$$

The accuracy rate (P) is the ratio of the number correctly predicted as category C_i to the number predicted as category C_i, which is calculated as follows:

$$P = \frac{TP}{TP + FP} \tag{13}$$

Recall (R) is the ratio of the number correctly predicted as category C_i to the total number of C_i in the dataset, which is calculated as follows:

$$R = \frac{TP}{TP + FN} \tag{14}$$

To evaluate the advantages and disadvantages of different algorithms, the concept of the $F1$ value is proposed to evaluate the overall accuracy and recall rates based on the accuracy and recall rates. The specific formula for calculation is as follows:

$$F1 = \frac{2PR}{P + R} = \frac{2TP}{2TP + FP + FN} \tag{15}$$

The structure of the formula allows concluding that when the precision and recall are higher, a higher $F1$ value will result, representing the better performance of the classification model.

4.4. Experimental Results

The recognition performance of the BERT-BiLSTM model proposed in this paper was compared with other legacy models. A total of four sets of experiments were set up to train on the CNN model, LSTM model, BERT-based model, and the BERT-BiLSTM model in this paper. The experimental environment was kept constant, and the $F1$ value described above was used as the evaluation index. The experimental results are shown in Table 5. The table shows the $F1$ values of 19 categories of labels under 4 different models, and the last row is the $F1$ value of model synthesis.

Combined with Figures 6 and 7, the horizontal axis lists the categories of key information for the recorded earthquakes. The vertical bars above each are the $F1$ values predicted for that category. The BERT-BiLSTM model had better recognition accuracy than the other models. In the test sets constructed for the 6.1 magnitude earthquake in Lushan County, Ya'an, Sichuan on 1 June 2022, the 6.0 magnitude earthquake in Markang City, Aba Prefecture, Sichuan on 10 June 2022, and the 6.8 magnitude earthquake in Luding County, Ganzi Prefecture, Sichuan on 5 September 2022, the BERT-BiLSTM model proposed in this paper, compared with the traditional CNN model, the LSTM model, and the BERT-based model, improves the results by 18.17%, 10.06%, and 5.5%, respectively. Among them, the BERT model had better performance in contextual semantic extraction and word polysemy resolution in the earthquake news text compared with the CNN model and LSTM model because of its mask setting based on a multi-head attention mechanism, and the BERT-BiLSTM model performed better in the earthquake news text after adding an LSTM layer and attention optimization layer.

Table 5. Experimental results.

Label Name	CNN	LSTM	BERT-Base	BERT-BiLSTM
TIM	71.06%	80.58%	85.82%	90.77%
MAG	70.55%	84.02%	85.71%	91.01%
PLA	75.39%	84.11%	85.63%	93.62%
COO	73.32%	82.51%	88.52%	92.43%
DEP	79.30%	82.55%	87.29%	93.95%
INT	77.53%	83.01%	85.99%	93.66%
HUR	76.29%	82.25%	85.50%	90.31%
DEA	70.73%	81.83%	86.83%	91.34%
FIR	70.81%	83.93%	88.24%	93.28%
NET	72.16%	80.66%	85.71%	90.53%
MAT	74.48%	81.17%	89.18%	94.09%
GEO	79.42%	82.12%	89.65%	92.31%
RES	71.76%	80.12%	89.32%	91.50%
DIS	79.02%	83.51%	86.45%	93.80%
ECO	74.15%	84.96%	86.30%	91.42%
VOL	76.10%	81.04%	86.40%	93.88%
DOC	71.86%	80.92%	88.31%	93.50%
PEO	72.90%	82.70%	86.71%	92.65%
TRA	75.31%	84.29%	85.43%	93.33%
	74.32%	82.43%	86.99%	92.49%

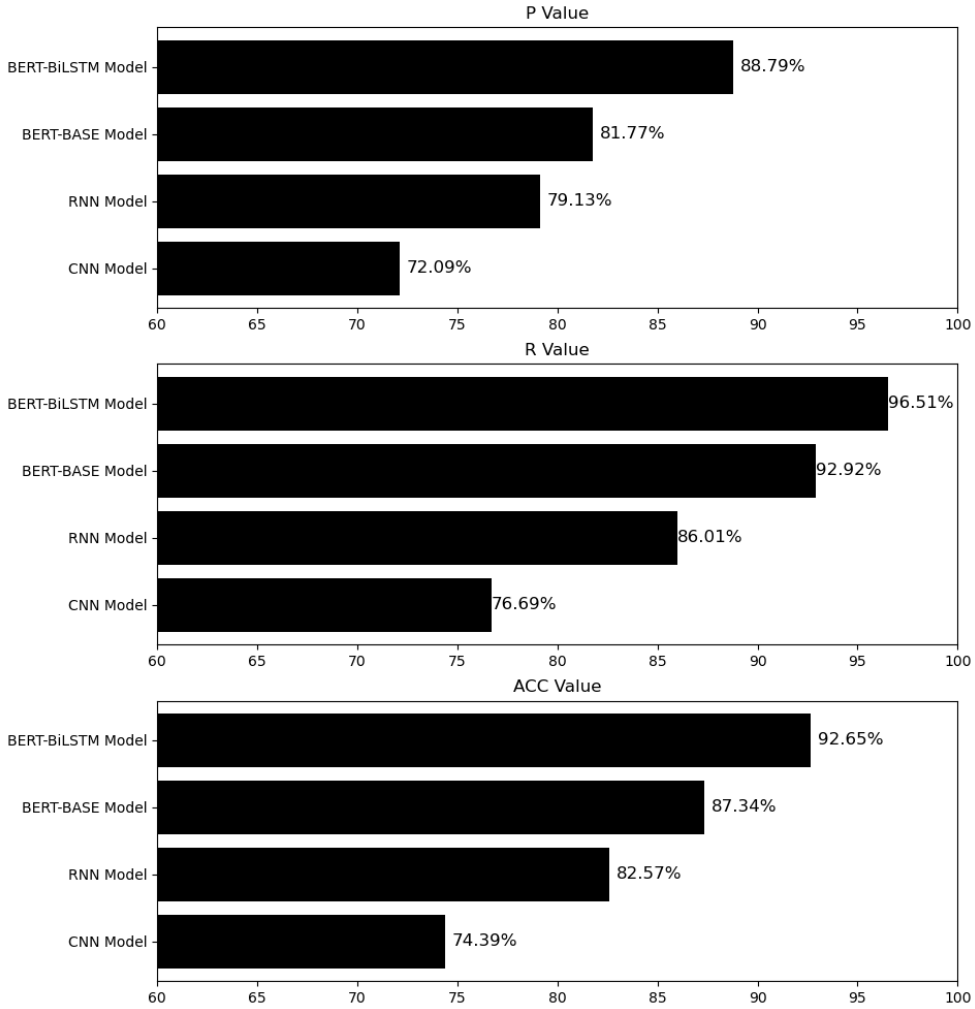

Figure 6. Experimental results for precision, recall, and accuracy.

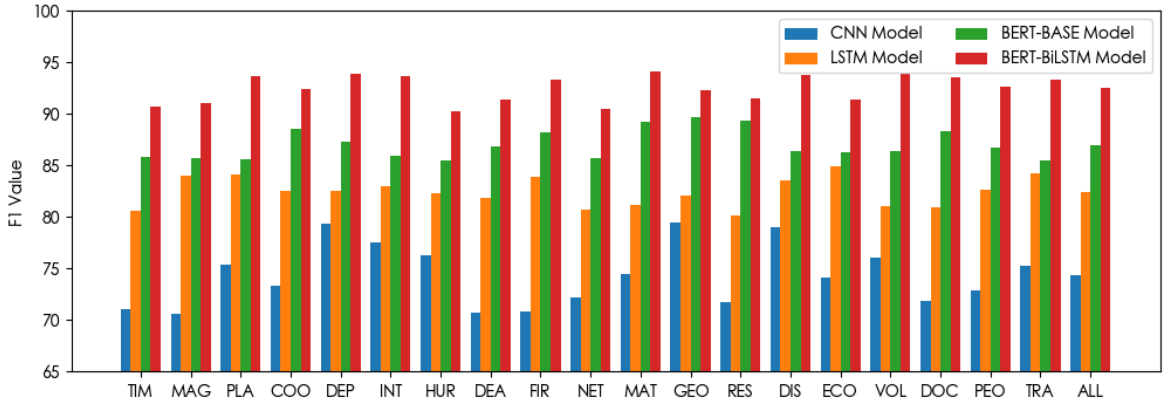

Figure 7. Experimental results for *F*1 values.

Finally, tests were conducted on the test sets constructed for the earthquakes listed above.

The slices were divided for 48 h after the earthquake, and the data were crawled for each time slice of earthquake news after the three earthquakes. Then, the earthquake attributes in the earthquake news were extracted.

Figure 8 shows the number of news items obtained within each time slice after the earthquakes occurred. As can be seen in the graph, the number of news items appearing in the timeline after the earthquakes increased incrementally.

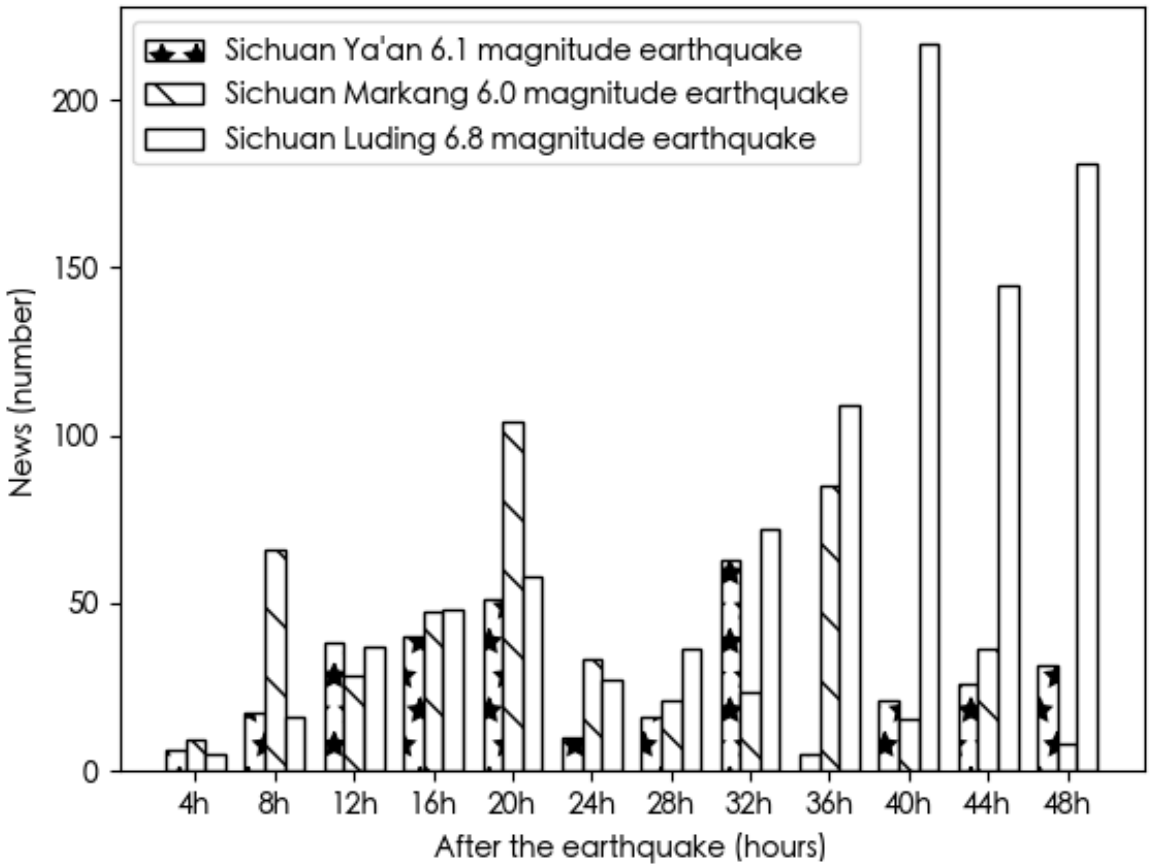

Figure 8. Number of news items obtained after each earthquake.

The intelligent recognition model of earthquake news text based on the BERT-BiLSTM algorithm in this paper was used for the intelligent recognition of seismic attributes in the incremental seismic emergency information text of three earthquakes. The results of

the recognition are shown in Figure 9. The proposed model was effective at extracting entities from the incremental seismic emergency information texts following the listed earthquakes, and it could effectively extract the entities in the incremental seismic emergency information text.

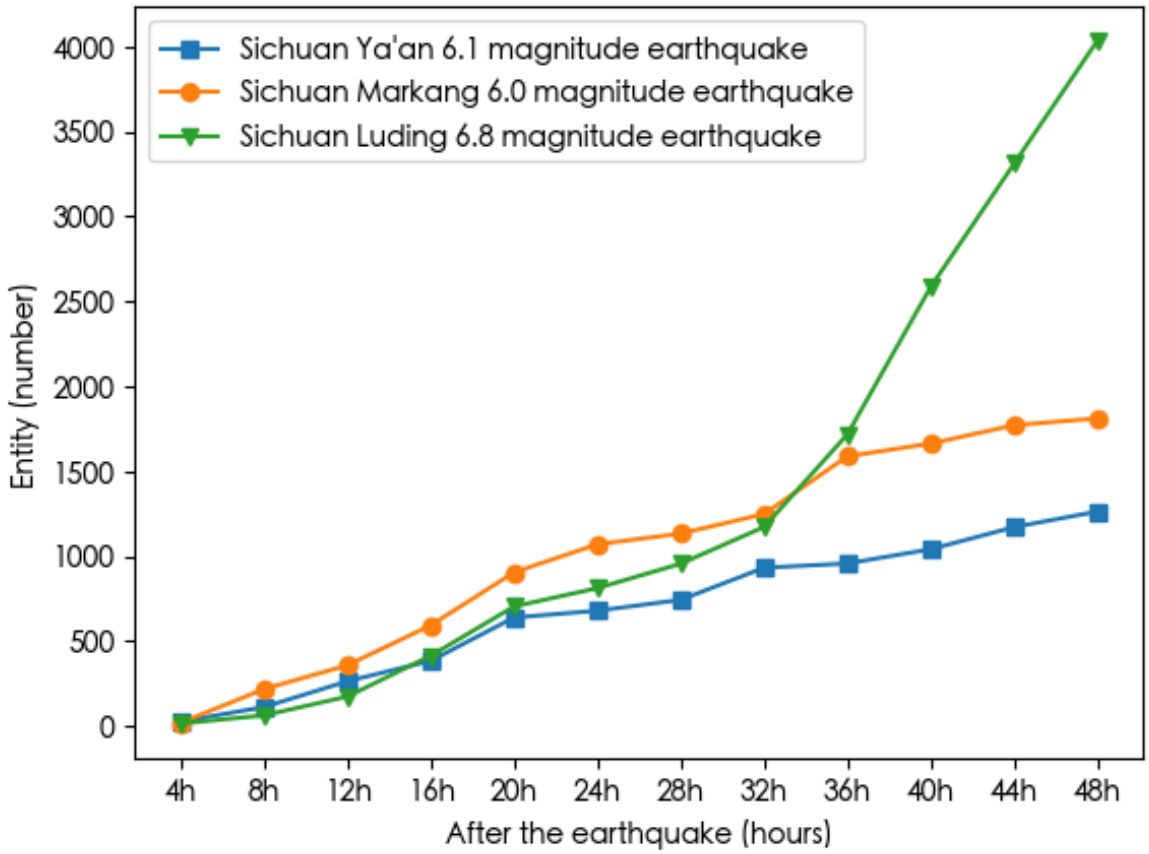

Figure 9. Number of entities extracted after each earthquake.

Taking the 6.8 magnitude earthquake in Luding County, Ganzi Prefecture, Sichuan Province on 5 September 2022 as an example, some of the obtained earthquake emergency information texts and the extracted earthquake key entities are shown in Table 6.

Table 6. Experimental results.

Release Time	Earthquake Emergency Information Text Content	Extracted Entities
5 September 2022, 1:03 p.m.	据中国地震台网正式测定：2022年 9月5日 12时 52分，在四川甘孜州泸定县（北纬 29.59度，东经 102.08度）发生 6.8级地震，震源深度 16公里。(According to the official measurement of the China Seismic Network: At 12:52 p.m. on 5 September 2022, an earthquake with a magnitude of 6.8 occurred in Luding County, Ganzi Prefecture, Sichuan Province (29.59 degrees north latitude, 102.08 degrees east longitude), with a focal depth of 16 km.)	TIM: 12:52 p.m. on 5 September 2022; PLA: Luding County, Ganzi Prefecture, Sichuan Province; COO: 29.59 degrees north latitude, 102.08 degrees east longitude; MAG: 6.8; DEP: 16

Table 6. *Cont.*

Release Time	Earthquake Emergency Information Text Content	Extracted Entities
5 September 2022, 1:28 p.m.	四川省应急救援队总队出动救援指挥车辆3辆，共14人赶赴震中甘孜州泸定县。(The Sichuan Provincial Emergency Rescue Team dispatched 3 rescue command vehicles, and a total of 14 people were rushed to Luding County, Ganzi Prefecture, the epicenter.)	VOL: 14
5 September 2022, 1:40 p.m.	甘孜、成都、德阳、乐山、雅安、眉山、资阳等7个救援队共530人地震救援力量立即赶赴震中展开救援。目前，武警消防、医疗救治、通讯电力、交通保畅等救援力量635人开展抢险工作。300名救援人员已到达震中，救援通道已抢通一条，已派出无人机侦查，部分房屋有受损情况。(Seven rescue teams from Ganzi, Chengdu, Deyang, Leshan, Ya'an, Meishan, and Ziyang, with a total of 530 earthquake rescue forces, rushed to the epicenter to start rescue operations. At present, 635 rescue forces, including armed police, firefighters, medical treatment, communication and electricity, and traffic security are carrying out rescue work. Three hundred rescuers have arrived at the epicenter, a rescue channel has been rushed through, drones have been sent to investigate, and some houses have been damaged.)	VOL: 530; DOC 635; FIR: 300
5 September 2022, 3:50 p.m.	9月5日12时52分，四川甘孜州泸定县磨西镇附近发生6.8级地震，震源深度16千米。地震已导致7人死亡。道路、通讯、房屋等受损情况正在核查中。(At 12:52 p.m. on September 5, an earthquake with a magnitude of 6.8 occurred near Moxi Town, Luding County, Ganzi Prefecture, Sichuan Province, with a focal depth of 16 km. The earthquake has killed seven people. Damage to roads, communications, houses, etc. is being checked.)	TIM: 12:52 on September 5; PLA: near Moxi Town, Luding County, Ganzi Prefecture, Sichuan Province; MAG: 6.8; DEA: 7

5. Conclusions

An intelligent recognition algorithm for discovery, extraction, and characterization of earthquake information in media and Internet texts that appear in the hours and days following significant seismic events was proposed and tested. It is based on a BERT-BiLSTM model and is a significant improvement over previous methods for examination of text data with specific content and incremental changes, such as earthquake information text. The effort required collecting earthquake-related data, building a historical earthquake emergency dataset, labeling it, and extracting sentence semantic information through a BERT pretraining model. In use, the BERT is combined with BiLSTM to obtain the contextual semantic environment of the sentence. An attention optimization layer is added to extract word weights, and ultimately, the earthquake news text is recognized by a CRF algorithm.

After collecting earthquake emergency texts for 48 h after three earthquakes (namely, the 6.1 magnitude earthquake in Lushan County, Ya'an, Sichuan on 1 June 2022, the 6.0 magnitude earthquake in Markang City, Aba Prefecture, Sichuan on 10 June 2022, and the 6.8 magnitude earthquake in Luding County, Ganzi Prefecture, Sichuan on 5 September 2022) and using the model proposed in this ontology for intelligent text recognition, the experimental results show that the method in this paper had good performance in the recognition of incrementally changing earthquake information text. It can achieve effective recognition of the named entities of earthquakes, extract key information such as rescue vehicles and personnel in their first occurrence in the mass of information on earthquakes, and improve the processing ability of massive earthquake news text information flow during an earthquake emergency. The trustworthy high accuracy and effective performance reduces the manual processing otherwise required to extract the text information, speeding up the application of this auxiliary information to earthquake emergency rescue and relief decision making and providing orderly followability.

This model can be used to discover the value of earthquake news. However, there are some shortcomings to this model. Since there are few destructive earthquakes of a magnitude of six or above in China, some attributes in the historical earthquake emergency dataset have relatively few exemplars, resulting in incomplete training data, which affects the accuracy of intelligent recognition of earthquake information texts. The model in this paper can only recognize the earthquake attributes in the earthquake information text. In the next phase of our research plan, we will further supplement the data in the historical earthquake emergency dataset and further analyze the potential relationships in the earthquake attributes to provide further data support for rescue and relief efforts during the post-earthquake emergency.

Author Contributions: Conceptualization, M.H.; Methodology, Z.W.; Software, Z.W.; Validation, S.L.; Resources, G.Y.; Data curation, C.L.; Writing—original draft, Z.W.; Writing—review & editing, M.H. and J.F. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported in part by the Science and Technology Innovation Program for Postgraduate students in IDP subsidized by Fundamental Research Funds for the Central Universities under Grant ZY20220326.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data used to support the findings of this study are available from the corresponding author upon request.

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

References

1. Valueva, M.; Nagornov, N.; Lyakhov, P.; Valuev, G.; Chervyakov, N. Application of the residue number system to reduce hardware costs of the convolutional neural network implementation. *Math. Comput. Simul.* **2020**, *177*, 232–243. [CrossRef]
2. Chen, Y. Convolutional Neural Network for Sentence Classification. Master's Thesis, University of Waterloo, Waterloo, ON, Canada, 2015.
3. Yao, L.; Liu, H.; Liu, Y.; Li, X.; Anwar, M.W. Biomedical named entity recognition based on deep neutral network. *Int. J. Hybrid Inf. Technol.* **2015**, *8*, 279–288. [CrossRef]
4. Strubell, E.; Verga, P.; Belanger, D.; McCallum, A. Fast and accurate entity recognition with iterated dilated convolutions. *arXiv* **2017**, arXiv:1702.02098.
5. Yang, J.; Liang, S.; Zhang, Y. Design challenges and misconceptions in neural sequence labeling. *arXiv* **2018**, arXiv:1806.04470.
6. Kong, J.; Zhang, L.; Jiang, M.; Liu, T. Incorporating multi-level CNN and attention mechanism for Chinese clinical named entity recognition. *J. Biomed. Inform.* **2021**, *116*, 103737. [CrossRef]
7. Lin, J.; Liu, E. Research on Named Entity Recognition Method of Metro On-Board Equipment Based on Multiheaded Self-Attention Mechanism and CNN-BiLSTM-CRF. *Comput. Intell. Neurosci.* **2022**, *2022*, 6374988. [CrossRef]
8. Sornlertlamvanich, V.; Yuenyong, S. Thai Named Entity Recognition using BiLSTM-CNN-CRF enhanced by TCC. *IEEE Access* **2022**, *10*, 53043–53052. [CrossRef]
9. Goodfellow, I.; Bengio, Y.; Courville, A. *Deep Learning*; MIT Press: Cambridge, MA, USA, 2016.

10. Huang, Z.; Xu, W.; Yu, K. Bidirectional LSTM-CRF models for sequence tagging. *arXiv* **2015**, arXiv:1508.01991.
11. Zhang, Y.; Yang, J. Chinese NER using lattice LSTM. *arXiv* **2018**, arXiv:1805.02023.
12. Zhang, H.; Guo, Y.; Li, T. Domain Named Entity Recognition Combining GAN and BiLSTM-Attention-CRF. *J. Comput. Res. Dev.* **2019**, *56*, 1851. [CrossRef]
13. Noersasongko, E. Capitalization Feature and Learning Rate for Improving NER Based on RNN BiLSTM-CRF. In Proceedings of the 2022 IEEE International Conference on Cybernetics and Computational Intelligence (CyberneticsCom), Malang, Indonesia, 16–18 June 2022; pp. 398–403.
14. Chen, Y.; Yao, Z.; Chi, H.; Gabbay, D.; Yuan, B.; Bentzen, B.; Liao, B. BTPK-based learning: An Interpretable Method for Named Entity Recognition. *arXiv* **2022**, arXiv:2201.09523.
15. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, Ł.; Polosukhin, I. Attention is all you need. *Adv. Neural Inf. Process. Syst.* **2017**, *30*. [CrossRef]
16. Zhao, H.; Jiang, L.; Jia, J.; Torr, P.H.; Koltun, V. Point transformer. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, BC, Canada, 11–17 October 2021; pp. 16259–16268.
17. Wolf, T.; Debut, L.; Sanh, V.; Chaumond, J.; Delangue, C.; Moi, A.; Cistac, P.; Rault, T.; Louf, R.; Funtowicz, M.; et al. Transformers: State-of-the-art natural language processing. In Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations, Online, 16–20 November 2020; pp. 38–45.
18. Devlin, J.; Chang, M.W.; Lee, K.; Toutanova, K. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv* **2018**, arXiv:1810.04805.
19. Hao, F.; Hao, H. News Title Classification Based on Contextual Features and BERT Word Embedding. *Inform. Sci.* **2022**, 40, 90–97.
20. Dai, Z.; Wang, X.; Ni, P.; Li, Y.; Li, G.; Bai, X. Named entity recognition using BERT BiLSTM CRF for Chinese electronic health records. In Proceedings of the 2019 12th International Congress on Image and Signal Processing, Biomedical Engineering and Informatics (CISP-BMEI), IEEE, Suzhou, China, 19–21 October 2019; pp. 1–5.
21. Yoon, W.; So, C.H.; Lee, J.; Kang, J. Collabonet: collaboration of deep neural networks for biomedical named entity recognition. *BMC Bioinform.* **2019**, *20*, 55–65. [CrossRef]
22. Li, X.; Zhang, H.; Zhou, X.H. Chinese clinical named entity recognition with variant neural structures based on BERT methods. *J. Biomed. Inform.* **2020**, *107*, 103422. [CrossRef]
23. Mingyi, M.; WU Chen, Z.Y.; Zhicheng, C. BERT named entity recognition model with self-attention mechanism. *CAAI Trans. Intell. Syst.* **2020**, *15*, 772–779.
24. Li, X.; Yan, H.; Qiu, X.; Huang, X. FLAT: Chinese NER using flat-lattice transformer. *arXiv* **2020**, arXiv:2004.11795.
25. Cui, L.; Wu, Y.; Liu, J.; Yang, S.; Zhang, Y. Template-based named entity recognition using BART. *arXiv* **2021**, arXiv:2106.01760.
26. An, Y.; Xia, X.; Chen, X.; Wu, F.X.; Wang, J. Chinese clinical named entity recognition via multi-head self-attention based BiLSTM-CRF. *Artif. Intell. Med.* **2022**, *127*, 102282. [CrossRef]
27. Jeon, K.; Lee, G.; Yang, S.; Jeong, H.D. Named entity recognition of building construction defect information from text with linguistic noise. *Autom. Constr.* **2022**, *143*, 104543. [CrossRef]
28. Bai, X.; Li, Y.; Chen, J.; Dai, Y.; Cao, K.; Cao, Y.; Zhao, H.; Gong, Q. Research on earthquake spot emergency response information classification. *J. Seismol. Res.* **2010**, *33*, 111–118.
29. Pennington, J.; Socher, R.; Manning, C.D. Glove: Global vectors for word representation. In Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), Doha, Qatar, 25–29 October 2014; pp. 1532–1543.
30. Xie, J.; Chen, B.; Gu, X.; Liang, F.; Xu, X. Self-attention-based BiLSTM model for short text fine-grained sentiment classification. *IEEE Access* **2019**, *7*, 180558–180570. [CrossRef]
31. Sasaki, Y. The truth of the F-measure. *Teach Tutor Mater* **2007**, *1*, 1–5.

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