Article

EmergEventMine: End-to-End Chinese Emergency Event Extraction Using a Deep Adversarial Network

Jianzhuo Yan [1], **Lihong Chen** [1], **Yongchuan Yu** [1], **Hongxia Xu** [1], **Qingcai Gao** [1], **Kunpeng Cao** [2]
and Jianhui Chen [1,3,4,*]

[1] Faculty of Information Technology, Beijing University of Technology, Beijing 100124, China;
     yanjianzhuo@bjut.edu.cn (J.Y.); chenlihong@emails.bjut.edu.cn (L.C.); yuyongchuan@bjut.edu.cn (Y.Y.);
     xhxccl@bjut.edu.cn (H.X.); gaoqingcai@emails.bjut.edu.cn (Q.G.)
[2] Beijing Research Institute of Smart Water, Beijing 100036, China; ckp_12345@sina.com
[3] Beijing International Collaboration Base on Brain Informatics and Wisdom Services, Beijing University of
     Technology, Beijing 100124, China
[4] Beijing Key Laboratory of MRI and Brain Informatics, Beijing University of Technology, Beijing 100124, China
[*] Correspondence: chenjianhui@bjut.edu.cn

Abstract: With the rapid development of the internet and social media, extracting emergency events from online news reports has become an urgent need for public safety. However, current studies on the text mining of emergency information mainly focus on text classification and event recognition, only obtaining a general and conceptual cognition about an emergency event, which cannot effectively support emergency risk warning, etc. Existing event extraction methods of other professional fields often depend on a domain-specific, well-designed syntactic dependency or external knowledge base, which can offer high accuracy in their professional fields, but their generalization ability is not good, and they are difficult to directly apply to the field of emergency. To address these problems, an end-to-end Chinese emergency event extraction model, called EmergEventMine, is proposed using a deep adversarial network. Considering the characteristics of Chinese emergency texts, including small-scale labelled corpora, relatively clearer syntactic structures, and concentrated argument distribution, this paper simplifies the event extraction with four subtasks as a two-stage task based on the goals of subtasks, and then develops a lightweight heterogeneous joint model based on deep neural networks for realizing end-to-end and few-shot Chinese emergency event extraction. Moreover, adversarial training is introduced into the joint model to alleviate the overfitting of the model on the small-scale labelled corpora. Experiments on the Chinese emergency corpus fully prove the effectiveness of the proposed model. Moreover, this model significantly outperforms other existing state-of-the-art event extraction models.

Keywords: text mining; event extraction; deep adversarial training

1. Introduction

Over the past few years, internet and mobile communication technologies have been radically transformed. Massive emergency information [1], especially online news, has been generated on various social media platforms. It is involved with both natural emergency events, such as fires and earthquakes, and man-made emergency events, such as traffic accidents, epidemics, and terrorist attacks, which often trigger heated discussions among netizens. The dissemination of emergency information affects public psychology and behavior, and then promotes the occurrence and evolution of emergency events. Therefore, collecting and analyzing emergency events on time is an important and urgent need for public safety, in which emergency event extraction is the key first step. However, the internet and the mobile internet are full of all kinds of disorderly emergency event news, which is intermingled with other news, and other news will hinder the clear cognition of users about an emergency event and relevant researchers' work in classification and

storage [2]. Moreover, the unstructured emergency information data are rich in content, diversified in form, multi-source, and heterogeneous [3]. Hence, efficient, accurate, and complete emergency event extraction is also a challenging text mining task.

At present, studies on emergency information mining mainly focus on text classification and event recognition. Although these studies have achieved satisfactory results, the limitations of mining tasks resulted in that they can only obtain the general and conceptual cognition of emergency events, such as category tag [4] or event type [5], which cannot meet the requirements of emergency risk warning based on event extraction results [6,7]. Different from text classification and event recognition, event extraction can obtain more detailed and structured content. An event extraction task usually includes four subtasks: trigger detection, trigger/event type identification, argument detection, and argument role classification. There are many state-of-the-art event extraction models in other professional fields, such as biomedical fields [8,9] and financial fields [10,11]. However, they cannot be applied directly to the emergency field. On the one hand, these models often depend on the domain-specific, well-designed syntactic dependency [12] or external knowledge base [8]. They have high accuracy in their professional fields, but the generalization ability is not good, and they are difficult to transfer to other fields. On the other hand, recent professional event extraction models have mainly focused on developing sophisticated networks to capture deep features in specialized language for improving the effect of event extraction, resulting in an increased demand for large-scale labelled corpora. However, existing Chinese emergency corpora, such as the Chinese Emergency Corpus (CEC, https://github.com/shijiebei2009/CEC-Corpus, accessed on 15 July 2021), are small-scale. There are often fewer than 200 instances of each event type. Therefore, it is impossible to realize emergency event extraction directly using these professional event extraction models. In fact, because of servicing the public, Chinese emergency information has a clearer syntactic structure compared with other specialized expressions in scientific articles or reports. Figure 1 gives a comparison between the genetics event mention and the emergency mention. Figure 1a shows a genetics event mention from the BioNLP Shared Task (BioNLP-ST) 2013 Cancer Genetics task. Green and orange denote entities and red denotes event triggers. A blue connection denotes the argument role. Event structures are constructed by an event trigger and one or more arguments. Although very brief, this genetics event mention has a nest event structure, containing one flat event and one nested event. The flat event, *transformed erythroid cells*, has only one entity as its argument. The nested event, *erbA/myb IRES transformed erythroid cells*, has the entity *erbA/myb IRES* and the flat event *transformed erythroid cells* as its arguments. Figure 1b is a seismic event in the CEC corpus which contains different categories of arguments, marked as orange, blue, green, and gray, respectively, and a trigger marked as red. The blue connecting lines denote the roles of the arguments. The whole event structure is simple and clear. Aiming at this characteristic of emergency texts, it is necessary to develop a new, lightweight model for emergency event extraction instead of directly using existing professional event extraction models.

Some open-domain event extraction methods [13,14] have been tested on multiple fields of datasets, often including emergency corpora, but their results often exhibit uncertainties and are relatively poor. The trade-off between the precision and recall rates is also a big challenge to open-domain event extraction. Therefore, these open-domain event extraction methods are also unsuitable for emergency event extraction, which has the clear event classification and definition, and a high requirement for the integrity, accuracy, and recall of the extraction results.
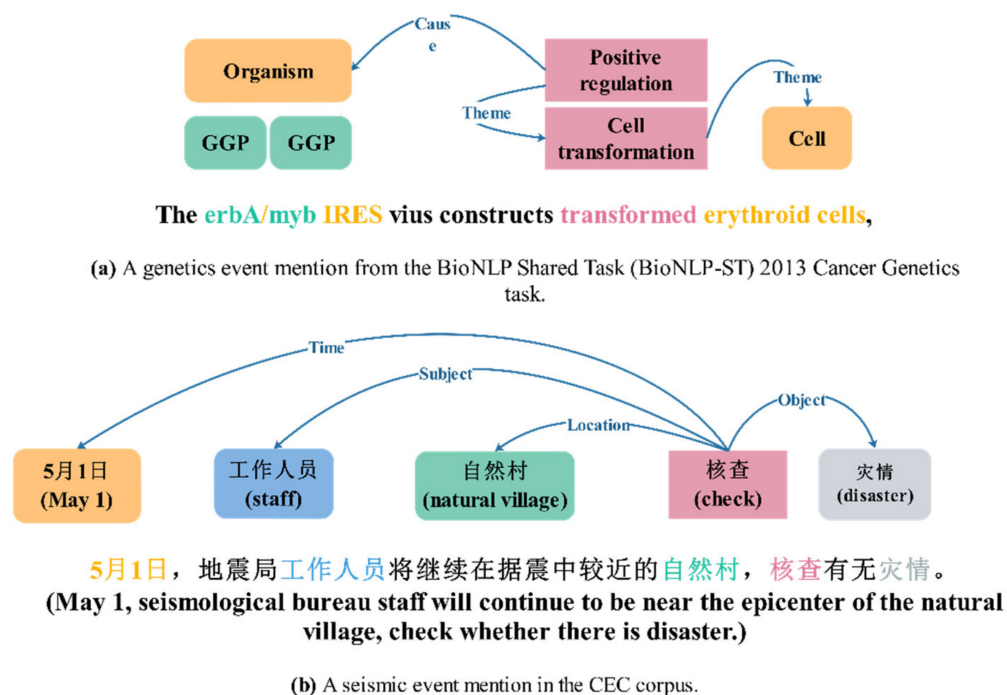
Figure 1. An example comparison between the genetics event mention and the Chinese emergency event mention.

Based on the above observations, this paper proposes an end-to-end Chinese emergency event extraction model, called EmergEventMine, using a deep adversarial network. The main contributions can be summarized as follows:

1. Firstly, this paper simplifies the event extraction with four subtasks as a two-stage task based on the goals of subtasks, and then develops a heterogeneous deep joint model for realizing the end-to-end Chinese emergency event extraction. The argument identification adopts the classic BiLSTM-CRF (Long Short-Term Memory-Conditional Random Fields) model, and the argument role classification is realized as a multi-head selection task. Compared with existing professional event extraction joint models, this model is a lightweight model, which does not depend on external syntactic analysis tools and has a simpler network structure for accelerating the convergence of the model on a small training data set.

2. Secondly, this paper integrates adversarial training, based on Free Adversarial Training (FreeAT) [15], into the joint model of event extraction. By adding small and persistent disturbances to the input of joint model, the overfitting of the model caused by the small training data set and simple network structure can be alleviated to improve the robustness and generalization of the model.

3. Thirdly, different from existing text mining studies based on the public CEC data set, XML format labels were removed from the standard CEC data set to fit the real application scenario. Experiments were performed on this restored CEC data set. Experimental results show that the proposed model can extract emergency events from online news texts more effectively and comprehensively compared with existing state-of-the-art event extraction models.

2. Related Work

2.1. Text Mining of Emergency Information

The text mining of emergency information has been a concern for a long time. Early research often considered topics from emergency information and then adopted shallow machine learning methods to classify emergency texts based on the obtained topics. Wang et al. [16] established a real-time classification model of emergency Weibo texts

combining the latent Dirichlet allocation (LDA) algorithm and support vector machine (SVM) algorithm. The LDA was used to identify topics from the original Weibo texts, and the identified topics were then used as the inputs of SVM for the Weibo classification. Han et al. [17] constructed a Weibo classification model based on the LDA algorithm and random forest (RF) algorithm. LDA was used to divide flood-related Weibo texts into six topics, and RF was used to classify Weibo texts based on the results of topic learning. With the development of deep learning, the effectiveness of deep neural networks in text mining has been widely recognized. Many emergency text classification methods based on deep learning have been developed. Yu et al. [18] integrated convolutional neural network (CNN), SVM, and logistic regression (LR) to classify tweet texts related to different hurricanes (Sandy, Harvey, Irma). Kumar et al. [19] used long short-term memory (LSTM) to classify tweet texts into different disasters, including Hurricane Harvey, the Mexico earthquake, and the Sri Lanka Floods. By using various deep neural networks, the hidden characteristics of emergency texts were effectively extracted to realize more effective text classification.

Emergency text classification can only obtain coarse-grained category tabs from emergency texts and cannot effectively recognize various natural or man-made emergency events. To address this problem, emergency event recognition has gradually gained attention. Zhang et al. [20] proposed the Chinese emergency event recognition model based on deep learning (CEERM), which obtained deep semantic features of words by training a feature vector set using a deep belief network, then identified triggers by means of a back propagation neural network. Li et al. [21] combined deep convolutional neural networks (DCNNs) and LSTM networks to develop the DCNNS-LSTM model for emergency event recognition on Uyghur texts. Naseem et al. [22] proposed a disaster event recognition method by using hierarchical clustering, in which the main events in Tweets were found by clustering, and then related sub-events were identified by clustering again according to spatio-temporal and semantic relationship information. Yin et al. [5] proposed a neural network model Conv-RDBiGRU (Convolution-Residual Deep Bidirectional Gated Unit), which extracted deep context semantic features through RDBiGRU and then activated them using the softmax function for identifying Chinese emergency events. These emergency event recognition methods can identify different types of emergency events from emergency texts. Therefore, emergency event recognition can often provide more detailed cognition about emergency information compared with emergency text classification.

However, both emergency text classification and emergency event recognition can only obtain a general and conceptual cognition of emergency events, i.e., category tag or event type. The in-depth emergency analysis and management, such as emergency risk warning [6,7], need to not only detect the occurrence of events, but also obtain the key elements of events, such as subject, time, and location. Therefore, it is necessary to develop an emergency event extraction model.

2.2. Event Extraction

Event extraction is an important research direction in the field of intelligent information processing, aiming at extracting specific types of events and elements from unstructured texts quickly [23]. It has been widely applied in data mining [24], information retrieval [25], question answering systems [26], and other fields.

Event extraction is a challenging work and consists of four subtasks [27]. These subtasks can be executed in a pipeline manner, where multiple independent classifiers are constructed separately and the output of one classifier can also serve as a part of input to its successive classifier. The classic pipeline-based event extraction model is DMCNN (dynamic multi-pool convolutional neural network) [28], in which the event extraction task is formulated as two stages. A DMCNN identifies the triggers, and then another similar DMCNN is used to identify arguments and their roles.

The pipeline manner for event extraction has some important shortcomings, including the error propagation [29], lack of subtask interaction [30], etc. Therefore, the joint manner,

where one classifier performs multi-task classification and directly outputs multiple sub-tasks' results, has become the recent research focus of event extraction. Many state-of-the-art joint models have been developed for event extraction in different professional fields.

Nguyen et al. [31] proposed a joint event extraction model based on RNNS which used a memory matrix to capture the dependency between trigger and arguments for improving the representations of event mentions. Sha et al. [32] proposed a joint event extraction model which captured dependency bridges among words and the potential interactions between arguments for enhancing the representations of event mentions. Li et al. [8] proposed a knowledge base-driven tree-structured long short-term memory networks (KB-driven Tree-LSTM) model, which incorporated dependency structures of contexts and entity properties from external ontologies to realize Genia event extraction. Liu et al. [12] proposed a hierarchical Chinese financial event extraction model (SSA-HEE) based on the entity structure dependency of trigger and arguments, and pre-defined eight types of entity dependency. By capturing syntax dependency information, the model effectively improved the ability to identify the relevance between trigger and arguments of financial events. However, these models depend on a domain-specific, well-designed syntactic dependency or external knowledge base, resulting in the low generalization ability of the model. They are also difficult to transfer to the field of emergency.

The small-scale labelled corpus set is another important challenge for Chinese emergency event extraction. At present, those state-of-the-art joint models for professional event extraction mainly focus on developing sophisticated networks to capture deep features in specialized language for improving event extraction. For example, the KB-driven Tree-LSTM model [8] designed the Tree-LSTM to learn hidden features from KB concept embedding. The BERT-BLMCC model [33] used the multi-CNN of different scale convolution kernels to extract local features from contexts. Zhao et al. [34] used a dependency-based graph convolutional network (GCN) to model local context, effectively capturing semantically related concepts in sentences through a syntactic dependency tree. The increasingly complex network structure brings an increased demand for large-scale labelled corpora. However, existing Chinese emergency corpora, such as the Chinese Emergency Corpus, are small-scale. There are often fewer than 200 instances of each event type. Experiments in [20] showed that, only for the event recognition subtask on CEC, the recognition performance deteriorated when the layer number of the restricted Boltzmann machine exceeded five. Yin et al. [5] pointed that "CEC has little data, which may lead to insufficient model training".

In order to effectively extract events from small-scale labelled corpora, existing studies have mainly adopted the event generation approaches to produce additional events for model training. According to different data labelled manners, these approaches can be divided into three types, including manual labelling [5], distant supervision-based methods [35,36], and bootstrapping-based methods [37,38]. Yin et al. [5] crawled 43,851 unstructured texts from we-media on the internet and manually labelled them to expand training corpora. Because manual labeling corpora is time-consuming and the reliability of the results is also difficult to evaluate objectively, distant supervision-based and bootstrapping-based automatic labeling methods are the research focuses. Chen et al. [36] proposed a soft distant supervision (SDS) method to automatically label the training examples of event extraction. SDS assigns key arguments and corresponding triggers to each event type through world knowledge (Freebase), and then uses the external knowledge base (FrameNet) to filter the triggers with noise for realizing large-scale event data labeling. Yang et al. [13] proposed a bootstrapping-based event generation method to automatically obtain new event sentences by using argument replacement and adjunct token rewrite. Ferguson et al. [35] proposed a bootstrapping-based event generation method to automatically identify new event instances by using paraphrase clustering. Yang et.al [37] proposed a bootstrapping-based method to automatically generate labelled event samples by editing prototypes and screening out generated samples by ranking the quality.

However, distant supervision-based methods need high-quality external resources, which is inconsistent with the technical route of this study, and they can easily bring noise. The bootstrapping process faces the difficulties of the similarity threshold setting and noise. Therefore, these event generation approaches are not suitable for few-shot emergency event extraction. It is necessary to develop a new few-shot emergency event extraction method following the approach of model optimization.

Besides the above close-domain methods, there are some open-domain methods for event extraction. Without predefined event schemas, open-domain event extraction aims at detecting events from texts and, in most cases, also clustering similar events via extracted event keywords [27]. Some open-domain event extraction methods have been tested on emergency corpora. Kunneman et al. [13] clustered tweets with overlapping mentions based on obtained time expressions and entity mentions and took the result of clustering as the event description. Liu et al. [14] extracted events from news reports by using three complex unsupervised generation models, which introduced latent variables generated by neural networks. However, extracted events of open-domain methods often exhibit uncertainties and incompleteness. Their precision and recall rates are also relatively poor. Moreover, because open-domain event extraction methods often need to set the thresholds for ranking or clustering, the trade-off between the precision and recall rates is also a big challenge. Therefore, these open-domain event extraction methods are also unsuitable for emergency event extraction, which has the clear event classification and definition, and a high requirement for the integrity, accuracy, and recall of extraction results.

3. Methods

3.1. Emergency Event Definition

According to content of CEC, this paper defines the emergency event as follows:

An emergency event is a natural or man-made disaster activity that takes place in a specific time and region, is participated in by one or more roles, and shows a series of action characteristics.

It can be defined as a five-tuple:

$$EmerEvent = (Subject, Object, Time, Location, Trigger) \tag{1}$$

where both *Subject* and *Object* refer to participants of events, *Time* refers to the time or period when an event occurs, *Location* refers to the place where the event occurred, and *Trigger* refers to a key action, which can indicate specific changes and features in the process of event occurrence. Figure 2 gives an example of an emergency event. Emergency event extraction involves finding the event, identifying the type of event, and comprehensively extracting the above five-tuple event content from unstructured emergency texts.
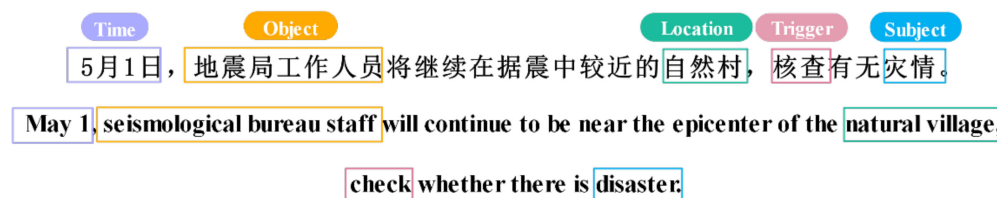

**Figure 2.** The five tuples of an emergency event.

3.2. Data and Pre-Processing

This paper adopts the CEC established by the semantic intelligence laboratory of Shanghai university. Corpus data were crawled from news reports of emergency events on the internet and we-media data. The corpus contains five types of emergency events: earthquake, fire, traffic accident, terrorist attack, and food poisoning. Table 1 gives the basic information of CEC.

Table 1. Basic information of CEC.

Event Category	Articles	Event Sentences	Events	Event Triggers	Event Arguments
earthquake	62	401	1002	1002	1290
fire	75	433	1216	1216	1609
traffic accident	85	514	1802	1802	2175
terrorist attack	49	323	823	823	1534
food poisoning	61	392	1111	1111	1100
total	332	2064	5954	5954	7708

Because CEC is in the XML language format, the text features are still obvious and easier to extract and train after preprocessing [5]. In order to better fit the actual event extraction scene, this study firstly restored corpora to the original news texts, as shown in Figure 3. After conducting word segmentation processing by using the jieba word segmentation tool, event elements and roles were annotated automatically according to the XML tags of CEC.
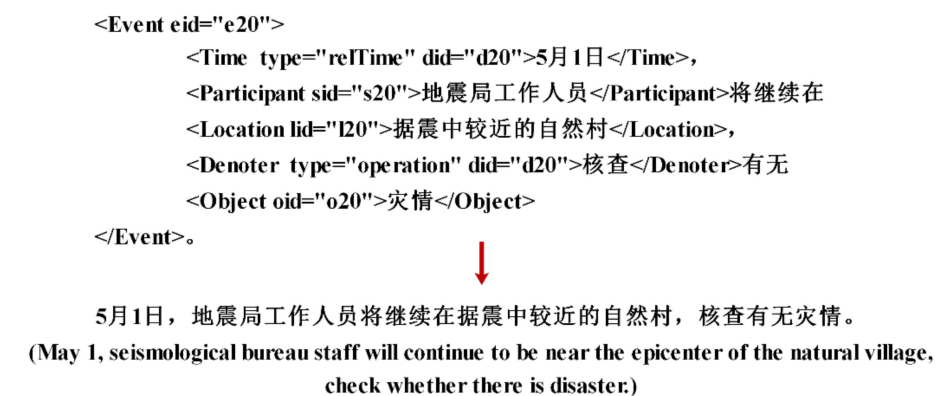
Figure 3. An example of CEC corpus in the XML language format.

Triggers and arguments were annotated by using the mode of "B/I/O Category abbreviation". The "B/I/O" is the BIO annotation mode [39], in which "B" indicates the beginning of entity, "I" stands for the middle or end of entity, and "O" stands for other. Category abbreviations are involved with argument categories and trigger categories. As shown in Table 2, "May 1" represents the argument of time type, "May" is identified as "B-Tim", and "1" is identified as "I-Tim". "check" belongs to the trigger of operation type and is identified as "B- operation".

Table 2. Event Element Annotation.

Annotation Example					
word	5月 (May)	1日 (1)	，	地震局 (seismological bureau)	工作人员 (staff)	将 (will)
annotation	B-Tim	I-Tim	O	B-Sub	I-Sub	O
word	继续 (continue to)	在 (be)	据	震中 (epicenter)	较近的 (near)	自然村 (natural village)
annotation	O	O	O	O	O	B-Loc
word	，	核查 (check)	有无 (whether)	灾情 (disaster)	。	
annotation	O	B-operation	O	B-Obj	O	

The annotation of event roles and attributes is based on triggers. As shown in Table 3, the first column is the position of the word and the second column is the corresponding word. If the word is a trigger, the third column is labelled as "Trigger" and the fourth column is labelled as "[$Argument position_1$, $Argument position_2$,..., $Argument position_i$,...]", where $Argument position_i$ is the end position of the corresponding argument in the role$_i$. If the word is an argument, its third column is labelled with the corresponding role type and its fourth column is empty.

Table 3. Event Role Annotation.

Position	Word	Event Role	Role Position
1	5月 (May)		
2	1日 (1)	Time	
3	，		
4	地震局 (seismological bureau)		
5	工作人员 (staff)	Subject	
6	将 (will)		
7	继续 (continue to)		
8	在 (be)		
9	据		
10	震中 (epicenter)		
11	较近的 (near)		
12	自然村 (natural village)	Location	
13	，		
14	核查 (check)	Trigger	[2,5,12,16]
15	有无 (whether)		
16	灾情 (disaster)	Object	
17	。		

3.3. EmergEventMine

This paper proposes an end-to-end Chinese emergency event extraction model, called EmergEventMine. The whole process of event extraction is simplified as a two-stage task, including argument identification and argument role classification, based on the goals of subtasks, and then performed synchronously by a heterogeneous deep joint model. Furthermore, FreeAT-based adversarial training is integrated into the joint model to alleviate overfitting of model caused by the small training data set. Figure 4 shows the overall structure of the model, which consists of four compounds, including a text vectorization layer, argument identification layer, argument role classification layer, and adversarial training mechanism.

3.3.1. Text Vectorization Layer

The text vectorization layer encodes input emergency texts as sentence vectors. In order to develop a thorough end-to-end model and avoid the dependency on external syntactic analysis tools, this paper only adopts word embedding based on word2vec [40] to construct sentence vectors. For a given sentence $S = [w_1, w_2, \ldots, w_t, \ldots, w_n]$, its sentence vector is as follows:

$$v_S = [v_1, v_2, \ldots, v_t, \ldots, v_n] \tag{2}$$
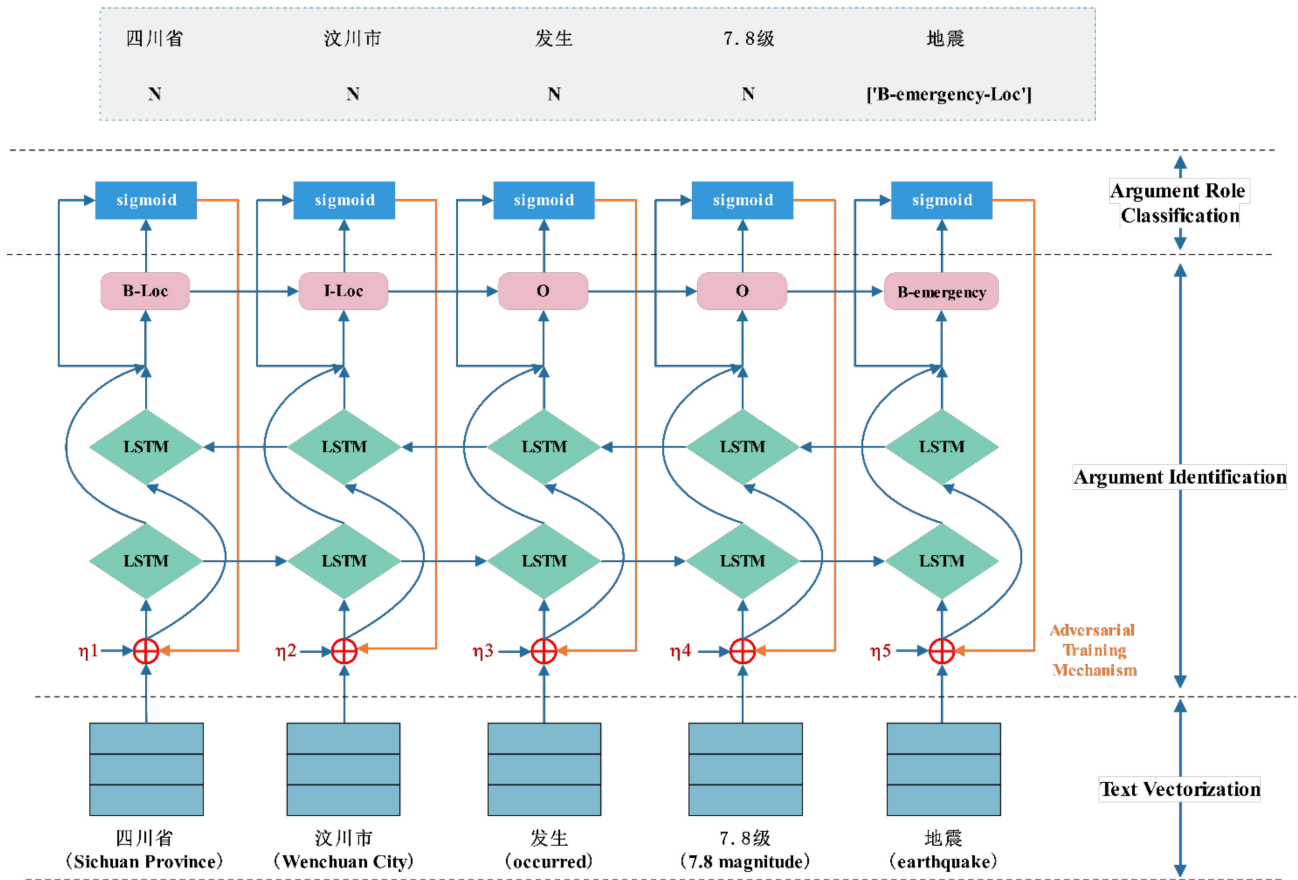
where v_t represents the word vector of w_t.

Figure 4. The model structure of EmergEventMine.

3.3.2. Argument Identification Layer

The argument identification layer predicts the potential event elements (argument and trigger) based on input sentence vectors. Argument identification is essentially a named entity recognition task, so we construct this layer by using the BiLSTM-CRF model, which is the most classic named entity recognition model in multiple text mining competitions [41].

As stated above, existing state-of-the-art joint models for professional event extraction mainly focused on developing sophisticated networks to capture deep features in texts, resulting in an increased demand for large-scale labelled corpora. However, Chinese emergency texts need a lightweight model because of their relatively clear syntactic structure and centralized parameter distribution, as well as the small-scale labelled corpora. Therefore, this paper only adopts the BiLSTM to capture global features of word tags. The process can be defined as follows:

$$h_t = [\overrightarrow{LSTM}(v_t), \overleftarrow{LSTM}(v_t)], \ t \in [1, n] \tag{3}$$

where v_t is combination vector of the word w_t, $\overrightarrow{LSTM}(v_t)$ is the forward hidden layer output, $\overleftarrow{LSTM}(v_t)$ is the backward hidden layer output, and h_t is a word representation of the combination vector of $\overrightarrow{LSTM}(v_t)$ and $\overleftarrow{LSTM}(v_t)$.

Based on the output of the BiLSTM, the CRF is used to perform the sequence labeling task for predicting event elements. In the CRF layer, the global optimal solution can be obtained by considering the relationship between adjacent tags. We calculate the score of each word w_t for each event element tag:

$$S_t(h_t) = V f(U h_t) , \ t \in [1, n] \tag{4}$$

where $f(\cdot)$ is an activation function, $V \in \mathbb{R}^{p \times l}$, $U \in \mathbb{R}^{2d \times l}$, p is the number of event element tag set, l is the width of LSTM layer, and d is the hidden size of LSTM layer. Therefore, given the sequence of score vectors S_1, S_2, ..., S_n and the vector of tag predictions y_1, y_2, ..., y_n, the linear-chain CRF score can be calculated as follows:

$$S(y_1, y_2, \ldots, y_n) = \sum_{t=0}^{n} s_{t,y_t} + \sum_{t=1}^{n-1} T_{y_t, y_{t+1}}, \ t \in [1, n] \tag{5}$$

where y_t is the tag of w_t, s_{t,y_t} is the predicted score of w_t given a tag y_t, and T is the transition matrix. Then, calculate the probability of the given tag sequence over all possible tag sequences in the input sentence s as follows:

$$\mathrm{Pr} = (y_1, y_2, \ldots, y_n | s) = \frac{e^{S(y_1, y_2, \ldots, y_n)}}{\sum \widetilde{y_1}, \widetilde{y_2}, \ldots, \widetilde{y_n}^{S(\widetilde{y_1}, \widetilde{y_2}, \ldots, \widetilde{y_n})}} \tag{6}$$

In the process of training, the BiLSTM-CRF layer minimizes the cross-entry loss $\mathcal{L}_{NER}$:

$$\mathcal{L}_{NER} = \sum_{t=0}^{n} -\log Pr(y_t | w_t; \theta) \tag{7}$$

where θ is a set of parameters.

3.3.3. Argument Role Classification Layer

The argument role classification layer recognizes the roles of arguments, i.e., the relations between the trigger and arguments, based on the predicted event elements. The trigger in events usually corresponds to multiple arguments, which is a one-to-many relation. Therefore, this paper models the argument role classification task as a multi-head selection problem [42]. For a given role tag r_k, the predication score between w_t and w_j can be calculated as follows:

$$Y(h_t, h_j, r_k) = Xf(Mh_t + Wh_j) \tag{8}$$

where $f(\cdot)$ is an activation function, h_i is the output of hidden state of LSTM corresponding to w_i, $X \in \mathbb{R}^l$, l is the hidden size of LSTM $M \in \mathbb{R}^{l*2d}$, $W \in \mathbb{R}^{l*2d}$, and d is the hidden size of LSTM layer. Then, the probability of w_j to be selected as the head of w_t with the role tag r_k between them can be calculated as follows:

$$Pr(head = w_j, \ tag = r_k | w_t, \theta) = \sigma(Y(h_t, h_j, r_k)) \tag{9}$$

where $\sigma(\cdot)$ is the sigmoid function.

In the process of training, this paper minimizes the cross-entry loss $\mathcal{L}_{rel}$:

$$\mathcal{L}_{rel} = \sum_{t=0}^{n} \sum_{j=0}^{m} -Pr(head = y_{t,j}, relation = r_k | w_t) \tag{10}$$

where y_t is the ground truth vector, $y_t \subseteq s$ and $r_k \subseteq R$ and m ($m < n$) is the number of associated heads of w_t. During decoding, the most probable heads and roles are selected using threshold-based prediction.

Finally, for the joint event extraction task, the joint model minimizes the final objective in the process of training:

$$\mathcal{L}_{joint}(W, \theta) = \mathcal{L}_{NER} + \mathcal{L}_{rel} \tag{11}$$

3.3.4. Adversarial Training Mechanism

The simpler network structure of the joint model accelerates the convergence of the model on the small training data set, but it can also easily lead to over fitting. Adversarial training adds perturbation into current model inputs for updating model parameters during model training. Although the output distribution of the model is basically consistent with

the original distribution, the generalization and robustness of the model are enhanced [38]. Therefore, this paper integrates the adversarial training based on FreeAT [15] into the joint model of event extraction. The following worst-case perturbation is added into the original embedding to maximize the loss function:

$$r_{adv\ m+1} = r_{adv\ m} + \epsilon \frac{\nabla_w \mathcal{L}(w; \hat{\theta})}{\|\nabla_w \mathcal{L}(w; \hat{\theta})\|} \tag{12}$$

where r_{adv} is the worst-case perturbation, m is the number of inner iterations, $\hat{\theta}$ is a copy of the current model parameter, and $\epsilon = \alpha\sqrt{D}$ is a perturbation hyperparameter, where α is the factor and D is the dimension of the embeddings x_i.

Finally, by combining the original inputs and the confrontation inputs produced by model adversarial training, the final loss of the joint model is as follows:

$$\mathcal{L}(w, \hat{\theta}) = \mathcal{L}_{joint}(x; \hat{\theta}) + \mathcal{L}_{joint}(x + r_{adv}; \hat{\theta}) \tag{13}$$

where $\hat{\theta}$ is the current value of the model parameter.

3.4. Evaluation

The precision rate P, the recall rate R, and the F_1 value [43] were adopted to evaluate the results of emergency event extraction. They can be calculated as follows:

$$Precision = \frac{|TP|}{|TP + FP|} \tag{14}$$

$$Recall = \frac{|TP|}{|TP + FN|} \tag{15}$$

$$F_1 = \frac{2 \times Precision \times Recall}{Precision + Recall} \tag{16}$$

where TP represents the number of events whose all arguments and roles are correctly recognized, |TP + FP| represents the number of recognized events, i.e., sentences in which an event argument is recognized, and |TP + FN| represents the number of events in the dataset.

4. Experiments and Results

4.1. Experimental Settings

The CEC data set was divided into the training set, the evaluation set, and the test set according to 8:1:1. Then, we obtained the mean P, R, and F_1.

In this experiment, the dimension of word vectors is set to 50, the number of iterations is 100, the hidden dimension is 64, and the layer of LSTM is 3. The learning rate is set to 1×10^{-3}, the activation function is tanh, and the adversarial disturbance parameter α is 0.01.

4.2. Result

4.2.1. Baseline Methods

This paper compared the proposed EmergEventMine model with the following state-of-the-art event extraction models:

- DMCNN [28]: It is a pipeline-based event extraction model with two-stage tasks, including trigger recognition and argument detection.
- JRNN [32]: It is a joint event extraction model based on RNNS, which enriches sentence representations by capturing the dependence between triggers and parameters.
- dbRNN [33]: It is a joint event extraction model which enriches sentence representations by capturing dependency bridges among words and the potential interactions between arguments.

- BERT-BLMCC [33]: It is a joint event extraction model with the sophisticated network, in which multi-CNN and BiLSTM are integrated to mine the depth features of specialized languages. This model adopts a joint feature vector consisting of the output of the BERT model, POS-feature, and entity-feature, and the output of the semantic coding layer.
- PLMEE [37]: It automatically generates labelled event samples by editing prototypes and screening out generated samples by ranking the quality, in order to solve the problem of insufficient training data.

4.2.2. Results Analysis

Experimental results are shown in Table 4 (The original source codes of DMCNN and PLMEE were downloaded from https://github.com/nlpcl-lab/event-extraction (20 November 2021) and https://github.com/boy56/PLMEE (16 December 2021). In this study, they are re-executed using default parameter settings. The experimental results of JRNN, dbRNN, and BERT-BLMCC are cited from reference [33]). The proposed EmergEventMine model achieves the best performance. Its precision, recall, and F_1 values are 92.73%, 89.91%, and 91.81%, respectively.

Table 4. Overall Performance.

Model	Precision	Recall	F_1
DMCNN (2015)	71.09%	76.23%	73.57%
JRNN (2016)	78.82%	81.44%	80.11%
dbRNN (2018)	82.15%	79.98%	81.05%
PLMEE (2019)	80.91%	83.01%	81.94%
BERT-BLMCC (2020)	88.83%	90.91%	89.86%
EmergEventMine-NoAT	90.51%	88.97%	89.73%
EmergEventMine	92.73%	89.91%	91.81%

The results in Table 4 show that all evaluation indicators of DMCNN are almost the lowest, which indicates that the pipeline model is really not suitable for the emergency event extraction task. Compared with DMCNN, the performance improvement of JRNN and dbRNN is very limited. This confirms our previous judgment, namely that domain-specific, well-designed syntactic dependency information cannot effectively improve the performance of event extraction models on the Chinese emergency information with a simple and clear syntactic structure. BERT-BLMCC achieves the best result in all baseline methods. This indicates that the joint feature vector can effectively enrich the feature representation of sentences for the training of sophisticated networks, which can reduce the dependence of the model on large-scale samples. However, the acquisition of features, such as the POS-feature, often depends on external syntactic analysis tools and makes it difficult to achieve end-to-end event extraction. In addition, BERT-BLMCC adopts convolution kernels with scales of 1, 3, and 5 to learn semantic information. This kind of complex network often means a high demand for computing resources, which seems unnecessary for Chinese emergency extraction with a simple and clear syntactic structure. PLMEE achieves the second-best result in all baseline methods. This shows that producing additional events for model training can make up for the lack of labelled samples to a certain extent, but it may bring disadvantages such as noise. Therefore, those event generation approaches are not suitable for few-shot emergency event extraction, as we mentioned earlier. Finally, the proposed EmergEventMine achieves the best performance on all evaluation indicators. Furthermore, EmergEventMine has a simpler structure which requires fewer labelled data and lower computing resources than the joint model with the complex structure, such as BERT-BLMCC.

The proposed EmergEventMine adopts a lightweight structure to accelerate the convergence of the model on small-scale training samples. However, this may lead to over

fitting of the model on the training set. Therefore, this study introduces FreeAT [15] based adversarial training into the joint model of event extraction. EmergEventMine-NoAT is an ablation study for verifying the effectiveness of adversarial training. As shown in Table 4, the precision rate of EmergEventMine is 2.22% higher than that of EmergEventMine-NoAT, which is EmergEventMine without the adversarial training module. Its recall rate is almost 1% higher than those of EmergEventMine-NoAT, and its F_1 value is 2.08% higher than that of EmergEventMine-NoAT. This indicates that adding adversarial training can effectively improve the generalization ability of the model.

4.3. Discussion

4.3.1. Analysis on the Performance of Subtasks

Event joint extraction is involved with two subtasks, "argument identification" and "argument role classification". This section will analyze their performances separately. Figure 5 shows the good performance of the proposed EmergEventMine on argument identification. The identification effect of trigger and time argument is the best. However, the identification effects of location and object arguments are poor. The main reason for this is that long arguments, such as the object argument "北京市119指挥中心" (Beijing 119 Command) and the location argument "尼加拉瓜西南部太平洋沿岸" (the Pacific coast in southwestern Nicaragua), make identification difficult.
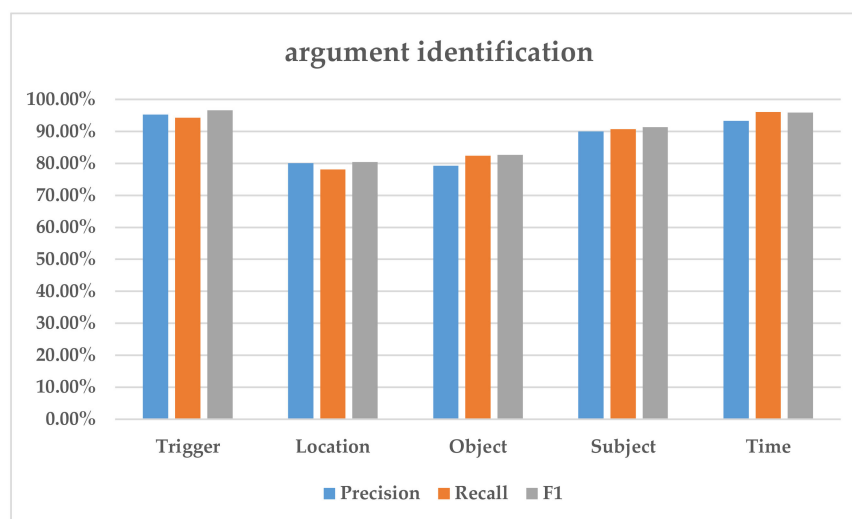

Figure 5. The results of argument identification.

Figure 6 shows the results of argument role classification. It can be seen that our model has achieved relatively good results. The classification results of the roles Trigger-Object and Trigger-Location are poor, which are also caused by the long object and location arguments.

4.3.2. Analysis on the Effectiveness of Emergency Event Extraction

As stated above, the existing emergency text mining mainly focuses on event recognition and text classification and can only obtain rough category labels. Figure 7 gives an example of the CEC corpus, which is a piece of news titled "a magnitude 6.6 earthquake occurred along the Pacific coast of Nicaragua". Different colored boxes represent different event types and different colored lines represent different argument types. From this piece of news, the method proposed by [44] can only obtain the category label "earthquake, casualties, loss or rescue".
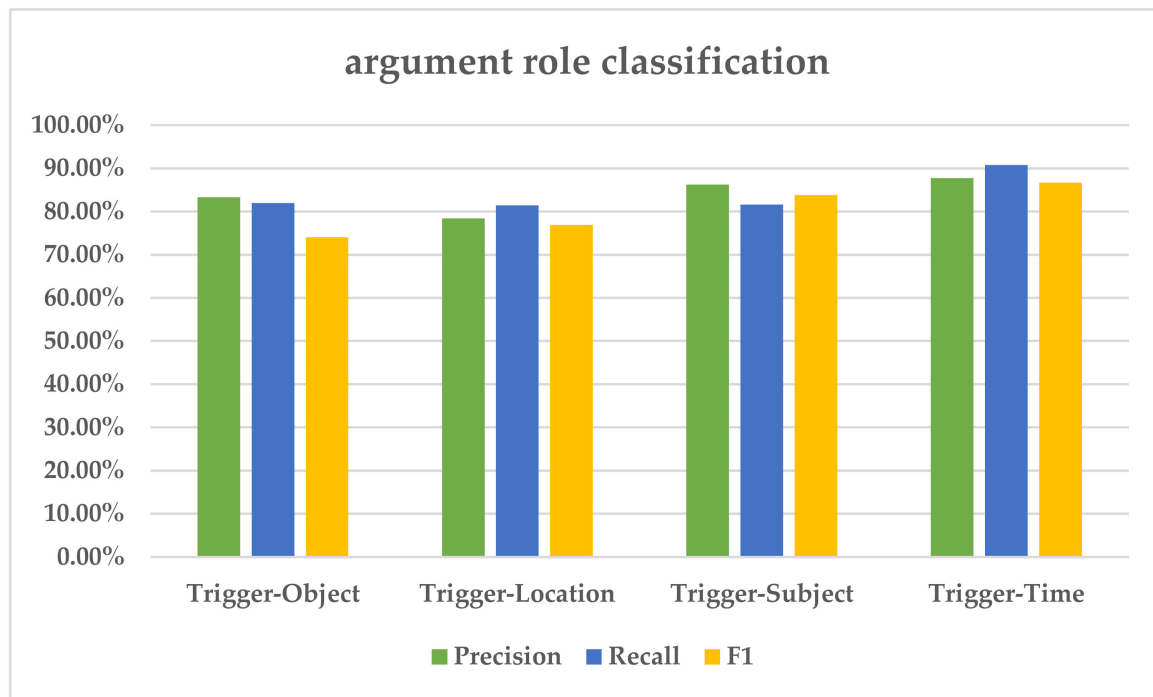
Figure 6. The results of argument role classification.

Figure 7. An example of CEC corpus.

Different from event recognition and text classification, EmergEventMine can extract eight emergency events, as shown in Figure 8. Circles represent triggers and boxes represent arguments. We can see that the trigger and the four types of arguments, including "Time", "Location", "Subject", and "Object" can be accurately recognized. Their relations can also be extracted to integrate the trigger and relevant arguments into different emergency events. Based on these emergency events, we can grasp the main content of news. Such a result is significantly better than the category label "earthquake, casualties, loss or rescue". More importantly, it can provide a more detailed and structured representation of this piece of news, for supporting further emergency information statistics and mining.
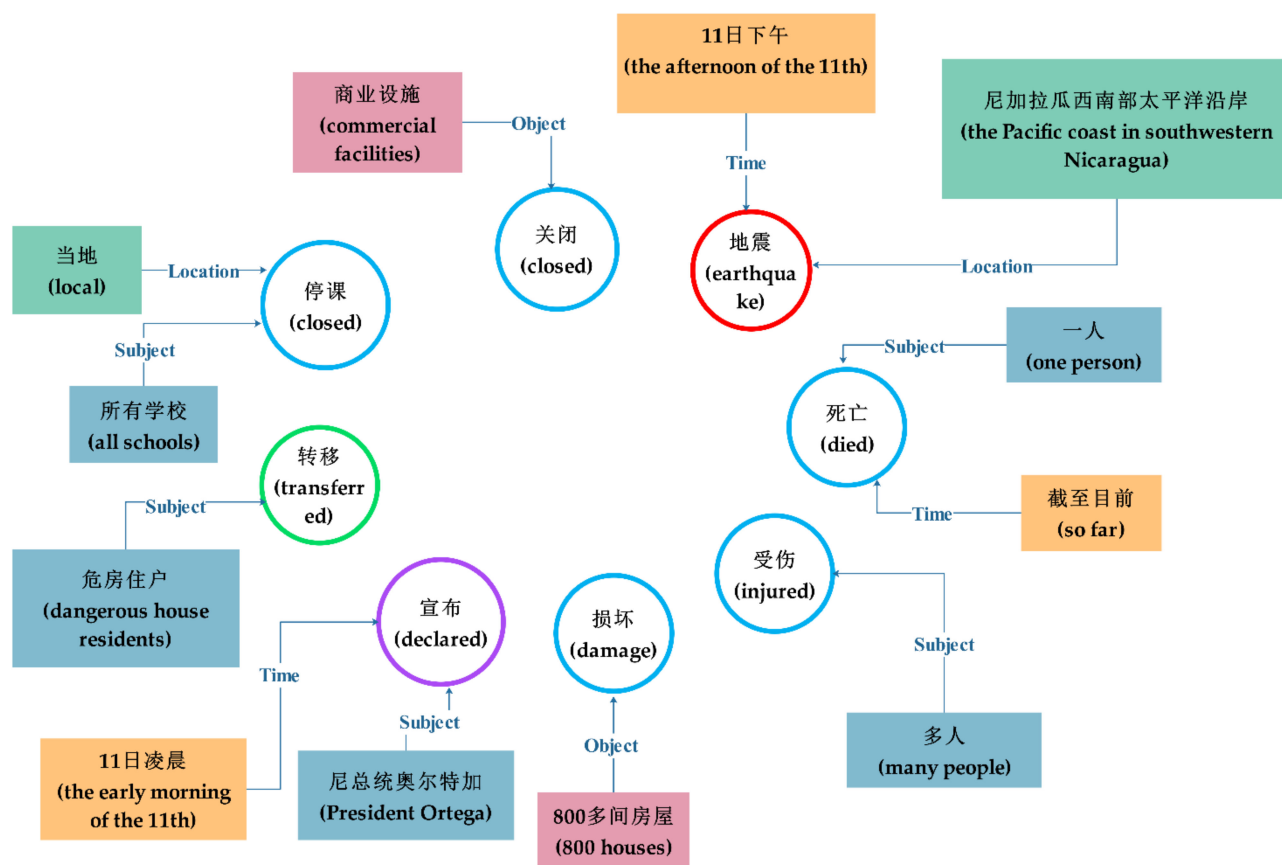
Figure 8. An example of event extraction results of EmergEventMine.

5. Conclusions

The purpose of this paper is to develop a practical method to extract emergency events from Chinese emergency texts. EmergEventMine is proposed to realize an end-to-end emergency event extraction, which does not require custom rules, feature engineering, and external syntactic analysis tools, and can automatically recognize and organize various event elements. Aiming to address the lack of large-scale label data and relatively simple emergency information, it adopts a lightweight deep joint model to accelerate the convergence of the model on a small training data set. The adversarial training is also integrated into the joint model to alleviate overfitting of the model caused by the small training data set and simple network structure for improving the robustness and generalization of the model. Experiments were performed on the restored CEC data set to fit the real application scenario. Experimental results show that EmergEventMine can obtain better event extraction results based on small-scale labelled data than existing state-of-the-art event extraction models.

In future work, we will focus on extracting the multi-dimensional relations between events to construct the event knowledge graph of an emergency. Based on this event knowledge graph, we will also study event-based emergency information mining, such as event tracing and causal analysis.

Author Contributions: Research conceptualization, Jianzhuo Yan, Jianhui Chen, Lihong Chen, Yongchuan Yu, and Hongxia Xu; methodology and software, Jianhui Chen and Lihong Chen; validation, Jianhui Chen, Lihong Chen, and Kunpeng Cao; writing—original draft preparation, Jianhui Chen and Lihong Chen; writing—review and editing, Jianhui Chen, Lihong Chen, and Qingcai Gao. All authors have read and agreed to the published version of the manuscript.

References

1. Shen, H.; Shi, J.; Zhang, Y. CrowdEIM: Crowdsourcing Emergency Information Management Tasks to Mobile Social Media Users. *Int. J. Disaster Risk Reduct.* **2021**, *54*, 102024. [CrossRef]
2. Yan, X. Research on Recognition of Sudden Events on Web Based on Combination of Rules and Statistical Method. *Data Anal. Knowl. Discov.* **2011**, *26*, 65–69.
3. Han, X.; Wang, J. Earthquake Information Extraction and Comparison from Different Sources Based on Web Text. *ISPRS Int. J. Geo-Inf.* **2019**, *8*, 252. [CrossRef]
4. Bai, H.; Yu, H.; Yu, G.; Huang, X. A novel emergency situation awareness machine learning approach to assess flood disaster risk based on Chinese Weibo. *Neural Comput. Applic.* **2022**, *34*, 8431–8446. [CrossRef]
5. Yin, H.; Cao, J.; Cao, L.; Wang, G. Chinese Emergency Event Recognition Using Conv-RDBiGRU Model. *Comput. Intell. Neurosci.* **2020**, *2020*, 7090918. [CrossRef]
6. Sheel, M.; Collins, J.; Kama, M.; Nand, D.; Faktaufon, D.; Samuela, J.; Biaukula, V.; Haskew, C.; Flint, J.; Roper, K.; et al. Evaluation of the early warning, alert and response system after Cyclone Winston, Fiji, 2016. *Bull. World Health Organ.* **2019**, *97*, 178–189C. [CrossRef]
7. Shoyama, K.; Cui, Q.; Hanashima, M.; Sano, H.; Usuda, Y. Emergency flood detection using multiple information sources: Integrated analysis of natural hazard monitoring and social media data. *Sci. Total Environ.* **2021**, *767*, 144371. [CrossRef]
8. Li, D.; Huang, L.; Heng, J.; Han, J. Biomedical Event Extraction based on Knowledge-driven Tree-LSTM. In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Minneapolis, MN, USA, 2–7 June 2019; pp. 1421–1430.
9. Yu, F. LSLSD: Fusion Long Short-Level Semantic Dependency of Chinese EMRs for Event Extraction. *Appl. Sci.* **2021**, *11*, 7237.
10. Zheng, S.; Cao, W.; Xu, W.; Bian, J. Doc2EDAG: An End-to-End Document-level Framework for Chinese Financial Event Extraction. In Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), Hong Kong, China, 3–7 November 2019; pp. 337–346.
11. Wang, P.; Deng, Z.; Cui, R. TDJEE: A Document-Level Joint Model for Financial Event Extraction. *Electronics* **2021**, *10*, 824. [CrossRef]
12. Liu, Z.; Xu, H.; Zhou, D.; Qi, G.; Sun, W.; Shen, S.; Zhao, J. Structural Dependency Self-attention Based Hierarchical Event Model for Chinese Financial Event Extraction. In *Knowledge Graph and Semantic Computing: Knowledge Graph Empowers New Infrastructure Construction*; Qin, B., Jin, Z., Wang, H., Pan, J., Liu, Y., An, B., Eds.; CCKS 2021; Communications in Computer and Information Science; Springer: Singapore, 2021; Volume 1466.
13. Kunneman, F.; Van Den Bosch, A. Open-domain extraction of future events from Twitter. *Nat. Lang. Eng.* **2016**, *22*, 655–686. [CrossRef]
14. Liu, X.; Huang, H.; Zhang, Y. Open Domain Event Extraction Using Neural Latent Variable Models. *arXiv* **2019**, arXiv:1906.06947.
15. Shafahi, A.; Najibi, M.; Ghiasi, A.; Xu, Z.; Dickerson, J.; Studer, C.; Davin, L.S.; Taylor, G.; Goldstein, T. Adversarial training for free! *arXiv* **2019**, arXiv:1904.12843.
16. Wang, Y.; Wang, T.; Ye, X.; Zhu, J.; Lee, J. Using Social Media for Emergency Response and Urban Sustainability: A Case Study of the 2012 Beijing Rainstorm. *Sustainability* **2016**, *8*, 25. [CrossRef]
17. Han, X.; Wang, J. Using Social Media to Mine and Analyze Public Sentiment during a Disaster: A Case Study of the 2018 Shouguang City Flood in China. *ISPRS Int. J. Geo-Inf.* **2019**, *8*, 185. [CrossRef]
18. Yu, M.; Huang, Q.; Qin, H.; Scheele, C.; Yang, C. Deep learning for real-time social media text classification for situation awareness using hurricanes sandy, Harvey, and Irma as case studies. *Int. J. Digit. Earth* **2019**, *12*, 1230–1247. [CrossRef]
19. Kumar, A.; Singh, J.P.; Dwivedi, Y.K.; Rana, N.P. A deep multi-modal neural network for informative twitter content classification during emergencies. *Ann. Oper. Res.* **2020**, 1–32. [CrossRef]
20. Zhang, Y.; Liu, Z.; Zhou, W. Event Recognition Based on Deep Learning in Chinese Texts. *PLoS ONE* **2016**, *11*, e0160147. [CrossRef]
21. Li, H.; Yu, L.; Tian, S.; Tuergen, I.; Zhao, J. Uighur Emergency Recognition Based on DCNNS-LSTM Model. *J. Chin. Inf. Sci. Technol.* **2018**, *6*, 52–61.
22. Anam, M.; Shafiq, B.; Shamail, S.; Chun, S.A.; Adam, N. Discovering Events from Social Media for Emergency Planning. In Proceedings of the 20th Annual International Conference on Digital Government Research, Dubai, United Arab Emirates, 18–20 June 2019; pp. 109–116.

23. Wang, Z.; Wang, X.; Han, X.; Lin, Y.; Hou, L.; Liu, Z.; Li, P.; Li, J.; Zhou, J. CLEVE: Contrastive Pre-training for Event Extraction. In Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing, Online. 1–6 August 2021; pp. 6283–6297.

24. Feng, R.; Yuan, J.; Zhang, C. Probing and Fine-tuning Reading Comprehension Models for Few-shot Event Extraction. *arXiv* **2020**, arXiv:abs/2010.11325.

25. Yu, W.; Yi, M.; Huang, X.; Yi, X.; Yuan, Q. Make it directly: Event Extraction Based on Tree-LSTM and Bi-GRU. *IEEE Access* **2020**, *8*, 14344–14354. [CrossRef]

26. Zhang, Y.; Xu, G.; Wang, Y.; Lin, D.; Huang, T. A Question Answering-Based Framework for One-Step Event Argument Extraction. *IEEE Access* **2020**, *8*, 65420–65431. [CrossRef]

27. Xiang, W.; Wang, B. A Survey of Event Extraction from Text. *IEEE Access* **2019**, *7*, 173111–173137. [CrossRef]

28. Chen, Y.; Xu, L.; Liu, K.; Zeng, D.; Zhao, J. Event extraction via dynamic multi-pooling convolutional neural networks. In Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing, Beijing, China, 26–31 July 2015; pp. 167–176.

29. Li, Q.; Ji, H.; Huang, L. Joint event extraction via structured prediction with global features. In Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics, Sofia, Bulgaria, 4–9 August 2013; pp. 73–82.

30. Nguyen, T.M.; Nguyen, T.H. One for All: Neural Joint Modeling of Entities and Events. *arXiv* **2018**, arXiv:1812.00195. [CrossRef]

31. Nguyen, T.H.; Cho, K.; Grishman, R. Joint event extraction via recur-rent neural networks. In Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, San Diego, CA, USA, 12–17 June 2016; pp. 300–309.

32. Sha, L.; Qian, F.; Chang, B.; Sui, Z. Jointly extracting event triggers and arguments by dependency-bridge RNN and tensor-based argument interaction. In Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence (AAAI-18), New Orleans, LA, USA, 2–7 February 2018; pp. 5916–5923.

33. Wang, S.; Rao, Y.; Fan, X.; Qi, J. Joint event extraction model based on multi-feature fusion. *Procedia Comput. Sci.* **2020**, *174*, 115–122.

34. Zhao, W.; Zhang, J.; Yang, J.; He, T.; Li, Z. A novel joint biomedical event extraction framework via two-level modeling of documents. *Inf. Sci.* **2021**, *550*, 27–40. [CrossRef]

35. Ferguson, J.; Lockard, C.; Weld, D.S.; Hajishirzi, H. Semi-Supervised Event Extraction with Paraphrase Clusters. In Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, New Orleans, LA, USA, 1–6 June 2018; pp. 359–364.

36. Chen, Y.; Liu, S.; Zhang, X.; Liu, K.; Zhao, J. Automatically labeled data generation for large scale event extraction. 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; pp. 409–419.

37. Yang, S.; Feng, D.; Qiao, L.; Kan, Z.; Li, D. Exploring Pre-trained Language Models for Event Extraction and Generation. In Proceedings of the 57th Conference of the Association for Computational Linguistics, Florence, Italy, 28 July–2 August 2019; pp. 5284–5294.

38. Zeng, Y.; Feng, Y.; Ma, R.; Wang, Z.; Yan, R.; Shi, C.; Zhao, D. Scale up event extraction learning via automatic training data generation. In Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence, New Orleans, LA, USA, 2–7 February 2018.

39. Sang, E.F.; Veenstra, J. Representing text chunks. *arXiv* **1999**, arXiv:cs/9907006.

40. Mikolov, T.; Chen, K.; Corrado, G.; Dean, J. Efficient estimation of word representations in vector space. *arXiv* **2013**, arXiv:1301.3781.

41. Li, X.; Wen, Q.; Lin, H.; Jiao, Z.; Zhang, J. Overview of CCKS 2020 Task 3: Named Entity Recognition and Event Extraction in Chinese Electronic Medical Records. *Data Intell.* **2021**, *3*, 376–388. [CrossRef]

42. Giannis, B.; Johannes, D.; Thomas, D.; Chris, D. Joint entity recognition and relation extraction as a multi-head selection problem. *Expert Syst. Appl.* **2018**, *114*, 34–45.

43. Goutte, C.; Gaussier, E. A probabilistic interpretation of precision, recall and F-score, with implication for evaluation. In Proceedings of the Advances in Information Retrieval, Proceedings of the European Conference on Information Retrieval, Santiago de Compostela, Spain, 21–23 March 2005; pp. 345–359.

44. Xing, Z.; Su, X.; Liu, J.; Su, W.; Zhang, X. Spatiotemporal Change Analysis of Earthquake Emergency Information Based on Microblog Data: A Case Study of the "8.8" Jiuzhaigou Earthquake. *ISPRS Int. J. Geo-Inf.* **2019**, *8*, 359. [CrossRef]