

Article

Research on the Construction of Typhoon Disaster Chain Based on Chinese Web Corpus

Hongliang Liu , Nianxue Luo * and Qiansheng Zhao

School of Geodesy and Geomatics, Wuhan University, Wuhan 430079, China; whulliu@whu.edu.cn (H.L.); qshzhao@sgg.whu.edu.cn (Q.Z.)

* Correspondence: nxluo@sgg.whu.edu.cn

Abstract: China is one of the countries most affected by typhoon disasters. It is of great significance to study the mechanism of typhoon disasters and construct a typhoon disaster chain for emergency management and disaster reduction. The evolution process of typhoon disaster based on expert knowledge and historical disaster data has been summarized in previous studies, which relied too much on artificial experience while less in-depth consideration was given to the disaster exposure, the social environment, as well as the spatio-temporal factors. Hence, problems, such as incomplete content and inconsistent expression of typhoon disaster knowledge, have arisen. With the development of computer technology, massive Web corpus with numerous Web news and various improvised content on the social media platform, and ontology that enables consistent expression new light has been shed on the knowledge discovery of typhoon disaster. With the Chinese Web corpus as its source, this research proposes a method to construct a typhoon disaster chain so as to obtain disaster information more efficiently, explore the spatio-temporal trends of disasters and their impact on human society, and then comprehensively comprehend the process of typhoon disaster. First, a quintuple structure (Concept, Property, Relationship, Rule and Instance) is used to design the Typhoon Disaster Chain Ontology Model (TDCOM) which contains the elements involved in a typhoon disaster. Then, the information extraction process, regarded as a sequence labeling task in the present study, is combined with the BERT model so as to extract typhoon event-elements from the customized corpus. Finally, taking Typhoon Mangkhut as an example, the typical typhoon disaster chain is constructed by data fusion and structured expression. The results show that the methods presented in this research can provide scientific support for analyzing the evolution process of typhoon disasters and their impact on human society.

Keywords: typhoon disaster chain; domain ontology; information extraction; BERT; sequence labeling



Citation: Liu, H.; Luo, N.; Zhao, Q. Research on the Construction of Typhoon Disaster Chain Based on Chinese Web Corpus. *J. Mar. Sci. Eng.* **2022**, *10*, 44. <https://doi.org/10.3390/jmse10010044>

Academic Editor: Han Soo Lee

Received: 6 December 2021

Accepted: 26 December 2021

Published: 1 January 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 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

Typhoon disasters pose fatal threats to life and cause serious losses to industry, agriculture, and transportation around the world every year, especially in the southeast coastal areas of China. Around the landing of a typhoon, factors, such as windstorms, rainstorms, huge waves, and storm surges, can cause catastrophes, which often interact with and influence each other to trigger secondary disasters. These series of secondary disasters constitute the typhoon disaster chain [1]. In order to guide typhoon disaster reduction more scientifically, it is necessary to understand the mechanism of typhoon disasters, explore the relationships between the secondary disasters, and build a typhoon disaster chain.

In previous studies, many scholars have discussed the evolution process of typhoon disasters. For example, based on the theory of natural disasters system and the Empirical Statistical Model, Ye analyzed the typhoon disaster cases in Fujian Province according to the historical statistics. Several typical elements were manually selected as the criteria to construct the typhoon disaster chain from the macro perspective [2]. This method can clearly reflect the regional characteristics and conduct disaster assessment rapidly.

Combined with the disaster dataset of Hebei Province from 1949 to 2018, Jing conducted research based on the theory of natural disasters system in order to build three disaster chains of Typhoon–Storm, Typhoon–Gale, and Typhoon–Surge. According to these different disaster chains, corresponding disaster reduction measures were proposed to provide support for the management of typhoon disasters [3]. Zhong took typhoon disaster as an example to discuss the mechanism, construction, and presentation of the incident chain. He proposed to perceive the disasters chain as an expression of priori knowledge which reflects the logic of the relationship between different types of secondary disasters in typhoon events. The probability of transformation between two incidents caused by typhoons can be calculated from plenty of historical cases through data mining, case reasoning, and the construction of a Bayesian network, etc. Such a theory can be applied to risk analysis and decision optimization [4]. Zheng selected a typhoon case and adapted the complex network model to construct the nature disaster network. Large numbers of disaster literature are explored with a data mining tool to find the relationship among disasters, and co-occurrence analysis is employed to build a disaster network. On this basis, a method to extract a proper disaster chain was proposed, which is greatly helpful for disaster mitigation and prevention [5]. It can be seen that previous studies mostly used natural disaster system theory, historical disaster statistics, and probability models to build the typhoon disaster chain, which requires much prior knowledge and expert experience. At the same time, less in-depth consideration is given to the distribution of disaster exposure, the constraints of the social environment, as well as the influence of spatio-temporal factors. In addition, due to the different focus of scholars and decision-makers, various disasters and emergency information are less connected with each other, causing problems, such as incomplete content, low reuse rate, and inconsistent expression of typhoon disaster knowledge.

With the rapid development of the internet, Web corpus, such as the online encyclopedia, internet news, and social media, contain massive unstructured or semi-structured information about typhoon disasters, which often erupt in a short time on the occurrence of typhoons. Compared with the traditional multi-source sensor observation methods, Web corpus is easier to obtain and more time-sensitive. Thus, it can be used as an important supplement for mining disaster processes and assessing disaster losses and has gradually become an implicit data source for obtaining typhoon event information [6,7]. In recent years, with the development of Natural Language Processing (NLP) technology, various information extraction methods have been used to carry out Named Entity Recognition, Relationship Extraction, and Knowledge Graph Construction. NLP has proved to be an effective method in the field of natural disasters study. For example, Kuzey distilled canonicalized events from news articles considering the textual contents, entity occurrences, and temporal ordering to organize them into fine-grained semantic classes. News reporting of the same disaster event was categorized, and a hierarchical structure among multiple events was established [8]. Trivedi extracted information from Twitter, constructed a real-time dynamic disaster event graph, and applied it to the tracking of disaster events [9]. To recap, extracting disaster event information from Web corpus, mining the correlation between typhoon secondary disasters, and building a typhoon disaster chain can explore the spatio-temporal trends of disasters and the impact of social activities, and then provide scientific guidance for emergency response and disaster reduction.

Taking these into consideration the present study proposes a method of constructing a typhoon disaster chain based on Web corpus. Adopting an appropriate structure, this research first constructs a Typhoon Disaster Chain Ontology Model (TDCOM) based on the theory of natural disaster systems. A formalized expression of the typhoon disaster chain is formed by establishing the concepts, properties, and relationships between elements of typhoon disasters. Then, NLP technology is adapted to extract disaster information from massive Web corpus, so as to construct the typhoon disaster chain through data fusion and structured expression. The results show that this method can be used to excavate the incident chain of typical typhoon disasters which can effectively sort out the spatio-

temporal variation of typhoon disasters as well as the relevant human social activities. The technique flowchart of this research is shown in Figure 1.

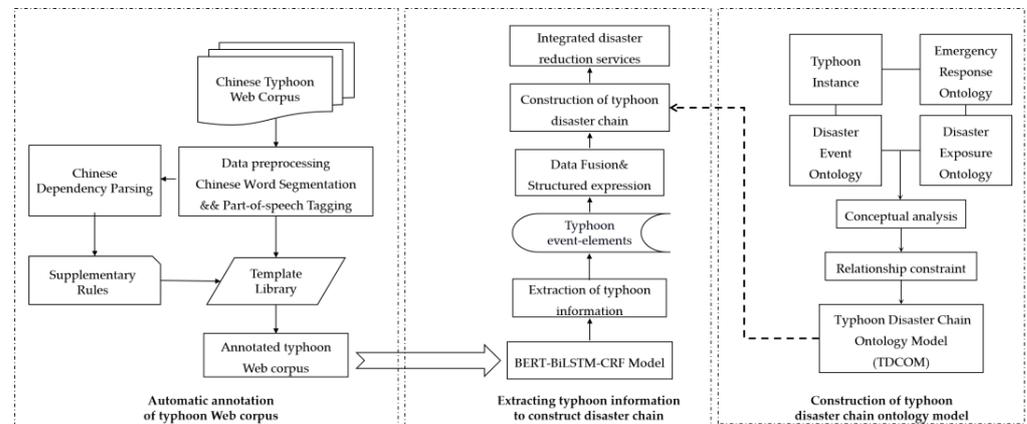


Figure 1. Construction of typhoon disaster chain based on Chinese Web corpus.

2. Methods

2.1. Construction of Typhoon Disaster Chain Ontology Model

Ontology, derived from Philosophy, is a systematic account of existence in terms of its essence and law. Ontology was defined as an explicit, formal specification of the terms and their relationships in a certain domain in the field of Information Science. The target of ontology in Information Science is to define a common vocabulary and describe words and their interrelationships on a formal and hierarchical level [10,11]. In recent years, ontology has provided a new method for the investigation of natural disasters. Applying the ontology to the field of Natural Disaster and Emergency Response can establish the relationship between the concepts, relations, properties, and constraints of disaster events, thereby providing a basis for emergency decisions that are of great significance to improve the comprehensive disaster reduction capabilities. For example, Yang constructed a formal and hierarchical typhoon disaster domain ontology by using Web Ontology Language (OWL) and manually identified several typical typhoon disaster chain patterns [12]. Web Ontology Language (OWL) is a semantic description language for ontology and can be used to explicitly represent the meaning of terms and the relationship between them [13]. Considering that a typhoon disaster is a typical geographical event that is essentially caused by the internal causal process and external accidental factors [14], the description and expression must take into account the temporal characteristics. Thus, the idea of event-oriented ontology modeling was adopted in this research. By using OWL and quintuple structure, based on the theory of natural disaster system, the logical structure of typhoon disaster chain was constructed from four aspects: Typhoon Instance, the disaster-inducing factors, mainly reflect the information of the typhoon event; Disaster Event Ontology Model is mainly used to reflect the disasters caused by typhoon and their constantly changing attributes over time and space; Disaster Exposure Ontology Model represents the main impacted object of the disaster; Emergency Response Ontology Model is used to reflect the human emergency response activities during the development of the disaster.

To construct an ontology model, it is necessary to clarify the logical relationship among the various elements of domain knowledge, and summarize them into a unified semantic framework, so as to establish the correlation between domain knowledge. The quintuple structure adopted in this research: Concepts, Properties, Relationships, Rules and Instances, provides a complete and unified knowledge description of typhoon disasters and the objects under their effects. To be specific, Concepts and Relationships form the basic framework of the Typhoon Disaster Chain Ontology Model (TDCOM), while Properties, Rules and

Instances enrich the content of the ontology model. The representation of the quintuple structure is shown in Formula (1).

$$Ontology = \{Con, Prop, Rel, Rule, Ins\}, \tag{1}$$

where *Con* represents a collection of concepts within the domain knowledge related to typhoon disaster which describes the disaster phenomena or events in the process of typhoon disasters; *Prop* represents the inherent properties of the elements in a typhoon disaster, such as Id, time, location, type of disaster, etc.; *Rel* represents as a collection of relations between two concepts in the process of typhoon disasters, such as spatial relationship, semantic relationship, etc., which is a mapping between two disaster elements; *Rule* represents the constraints among concepts, events, and data in the disaster field; *Ins* represent as a collection of specific object instances, which is represented as a specific event.

This study refers to the concepts and hierarchical relationships in relevant national standard documents issued by the Chinese government, such as Classification and codes for natural disasters (GB/T 28921-2012), Classification and coding for natural disaster exposure (GB/T 32572-2016), Emergency classification and coding (GB/T 35561-2017). The conceptual framework of typhoon disaster chain ontology is shown in Figure 2.

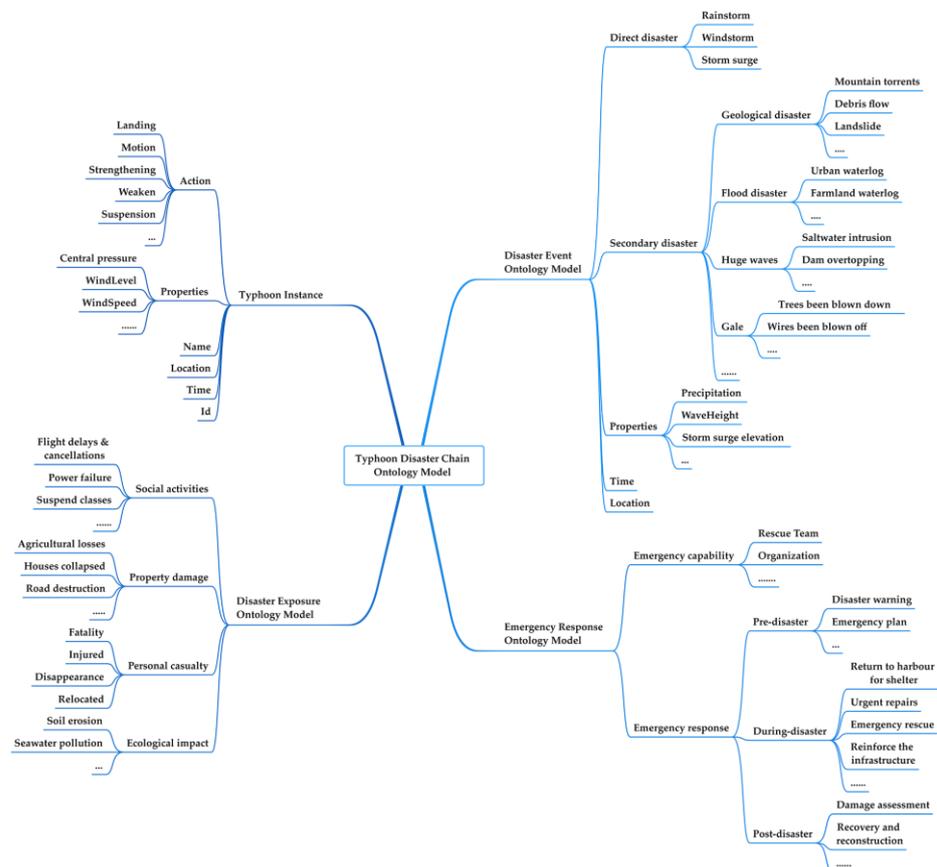


Figure 2. The conceptual framework of typhoon disaster chain ontology.

2.2. Extraction of Typhoon Disaster Information Based on Chinese Web Corpus

As a typical geographical event, a typhoon event contains temporal, spatial, and other attribute information. The information requiring attention in this study is defined as event-elements. The current Information Extraction (IE) methods for typhoon events mainly include pattern-matching algorithms based on grammar rules and dictionaries, machine learning methods based on statistics, and deep learning methods to solve the problem of sequence labeling. However, the drawbacks of these methods are illustrated as follows and improvements can be made for a better result on information extraction:

1. The pattern-matching algorithm requires manual construction of a knowledge base and statement expression so as to extract typhoon information. For example, Madhyastha presented a scheme for extracting semantic information from the syntactic structure given by the link grammar system and identifying instances of events. Information is derived by using a set of rules [15]. This method is simple with high recognition accuracy, yet it is time-consuming, less portable, and costly to maintain.
2. Machine learning methods are widely used, such as the Maximum Entropy model (ME), Hidden Markov Model (HMM), Support Vector Machine (SVM), Conditional Random Field (CRF), and, etc., which do not require complex matching rules. However, a large number of the manually annotated corpus is necessary and the quality of which influences the method significantly. For example, in order to detect location estimation for events, Sagcan proposed a hybrid system, in which regular expressions are used to define some of the toponym recognition patterns and conditional random fields (CRF) are conducted to extract toponyms from tweets [16].
3. The method based on deep learning can obtain the semantic features of the corpus. Especially in recent years, the sequence labeling model based on the neural network can better mine the contextual information and reduce the tedious artificial features, such as long short-term memory networks (LSTM), bidirectional long short-term memory networks (BiLSTM), and various neural network models, that are combined or improved on this basis [17–21]. For example, Xu proposed a deep neural network-based framework to jointly detect and extract events from Twitter by defining a joint loss function, a BiLSTM based common representation layer, and a control gate. A CRF layer is further employed to capture the strong dependencies among output labels [21]. These information extraction models with different network structures have strong generalization ability and perform well in tasks, such as Named Entity Recognition and Relationship Extraction. However, these methods cannot resolve the problem of polysemy. Furthermore, the amount of training corpus is relatively small to tackle the problem of specific domain tasks.

Devlin et al. proposed the BERT pre-trained language model [22], which uses large-scale unlabeled text to obtain rich semantic information and deepen the depth of the natural language processing model. With the better effect of tackling the problems of polysemy and small corpus size, the introduction of BERT into the information extraction model complements the corpus features in specific domains. In addition, it produces a better effect in obtaining the features at the character, word, and sentence levels, thereby improving the information extraction effect of the model. The information extraction method based on the pre-trained model has been applied in some specific domain information processing tasks. For example, Zhang et al. applied the BERT model to entity recognition in Chinese electronic medical records and attained better results [23]; Hou proposed the Chinese relation extraction algorithm for public security based on BERT, which can effectively mine security information [24].

At present, there are still few research on typhoon disaster information extraction based on the pre-trained language model. This study introduces the BERT pre-trained model on the basis of the BiLSTM-CRF model to realize the automatic extraction of typhoon event information from the Chinese Web corpus.

2.2.1. Automatic Annotation of Typhoon Web Corpus

Due to the lack of annotation datasets for typhoon disasters at the present, this study constructed a typhoon-related corpus based on data from websites, such as Weather China, Sohu News, Sina Weibo, etc., and then carried out automatic annotation to construct an experimental dataset. The specific steps are as follows:

1. Data preprocessing. Since the Web corpus contains various redundant information, it is necessary to use Chinese word segmentation and part-of-speech (POS) tagging technology to convert unstructured text data into a lexical combination, which is the basis for information extraction, information aggregation, and text categorization.

A large number of irrelevant words and unknown words in the text will introduce a lot of noise to the corpus dataset. Therefore, in this research, a user-defined stop words list, and a typhoon disaster domain dictionary were constructed based on the ontology model of the typhoon to filter phrases and to reduce the impact of noise.

2. Construction of template library based on POS tagging and regular expressions. Since similar rules are shared by event description, analyzing which can obtain the linguistic characteristics of event expression. Based on these characteristics, regular expressions and POS tagging are used to establish the rule templates for the temporal, spatial, and attribute information of typhoon events, respectively. For example, the text which translated into English is “central pressure/995/pha” can be matched with the POS combination of “stf + m + q” or the rule template of “N (Disaster attribute) + X (Attribute value) + Q (Unit noun)”.
3. Supplements templates with Chinese dependency parsing. The same meaning can be phrased by different expressions in typhoon events description, which imposes a limit on information extracting by using the single rule template. Dependency parser can analyze both the semantic associations and dependencies between the constituent units in a sentence, and it can supplement the rule templates effectively to improve the effect of information extraction [25]. For typhoon events, the syntactic relations that this research mainly focuses on are “Subject–Verb” (SVB), “Verb–Object” (VOB), and “Attribute” (ATT). For example, as shown in Figure 3, the sentence translated into English is “Typhoon Mangkhut landed in Taishan, Jiangmen, Guangdong Province, by September 17th, 1.053 million people in Guangdong province were evacuated and relocated”. To make it easier to understand, the Chinese syntactic structure has been transformed into the corresponding English expression. Among them, “Typhoon Mangkhut” is the Subject of “landed”, “Taishan, Jiangmen, Guangdong Province” is the Object of “landed”, “17th September” is the temporal adverbial, “people” is the Object of “evacuated and relocated”, “Guangdong Province” and “1.053 million” are the attributive complement of “people”. Through the traversal of the nodes and dependencies on the syntactic tree, event-elements can be obtained, and the rule templates can be supplemented effectively.
4. Corpus Annotation. Rule templates are used for customized annotation of the dataset. The BIO text annotation system is adopted in this research. The annotation specifications of event-elements, such as time, location, disaster, and emergency attributes, in the Web corpus are shown in Table 1. There are 8 types of event-elements that need to be obtained.

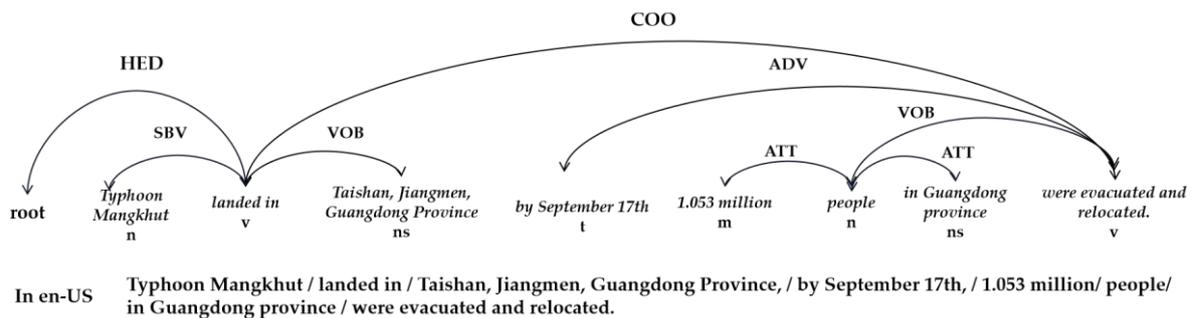


Figure 3. Example of Chinese dependency parsing.

Table 1. The annotation set of typhoon event-elements.

Annotation	Illustration
B	The beginning of the event-element
I	Other parts of the event-element
O	Parts that are not event-element;
name	The Name of typhoon (Name)
t	Time information (Time): Time of generation, landing, occurrence, suspension, stop coding, etc.
ns	Location information (LOC): location of generation, landing, occurrence, suspension, typhoon center, etc., as well as affected area, longitude and latitude.
nt	Organization information (ORG): government departments, schools, airports, etc.
stf	Disaster information (STF): Typhoon intensity, wind intensity, wind speed, central pressure, the elevation of storm surge, etc.; secondary disasters, such as storm surge, flood, landslide.
szq	Disaster exposure information (SZQ): casualties, property losses, agricultural losses, fisheries loss, etc.
syj	Emergency response information (SY): level of emergency response, emergency plan, urgent repairs, emergency rescue, etc.
vtf	Action of typhoon (VTF): generation, landing, approaching, strengthen, weaken, suspension, stop coding, etc.

2.2.2. Typhoon Information Extraction Model Based on BERT-BiLSTM-CRF

In this study, the extraction of typhoon event information was considered as a sequence labeling task. In addition, the BiLSTM-CRF model that is usually used for Named Entity Recognition was introduced, and a layer of BERT pre-trained language model was integrated on this model for word vector embedding. This integrated model structure was shown in Figure 4. At first, the character sequence was used as the input of BERT. After training, the word vectors integrated with the contextual semantic information were constructed and outputted to the BiLSTM layer for encoding. The forward LSTM layer learned the future features and the backward LSTM layer learned the historical features. The feature h_t at time t after BiLSTM network training was used as the output and decoded by CRF. Finally, the optimal labeling sequence was obtained.

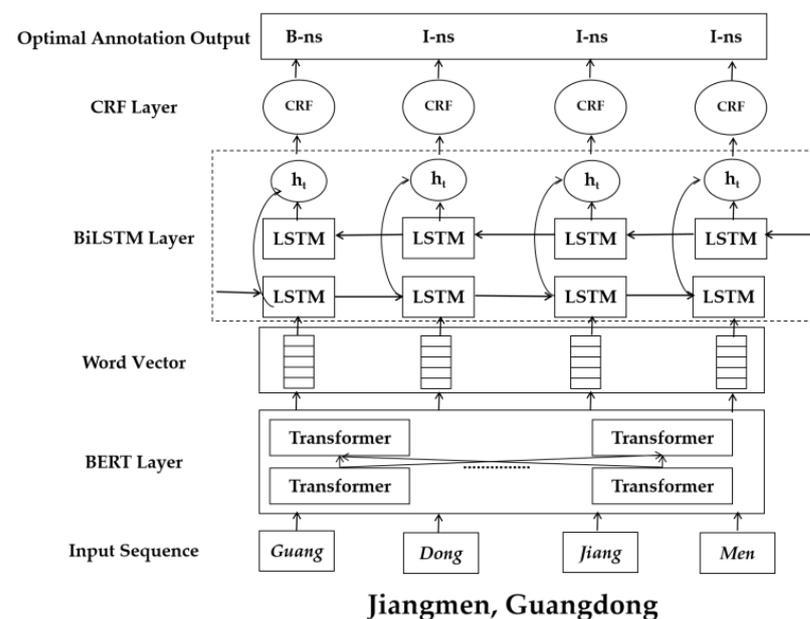


Figure 4. Structure of BERT-BiLSTM-CRF model.

BERT produces a better effect on processing long-distance dependencies in sentences. It can not only obtain feature information at the phrase level but also learn syntactic structure features by calculating the weight relations of contexts, so as to obtain the semantic information of the whole sentence. Therefore, the outputs of BERT can extract abundant word-level features, syntactic features, and semantic features. The word vectors processed by BERT demonstrate strong advantages in terms of semantic representation. There are two stages for the vectorization of typhoon corpus based on BERT:

1. Model training. Based on large-scale unlabeled corpus and deep Bidirectional Transformers, BERT performs unsupervised training to obtain corpus features. Meanwhile, a small-scale typhoon corpus was introduced to fine-tune the model.
2. Character vectorized representation. The typhoon text sequence is input into the BERT model to obtain the real-valued vector corresponding to the character sequence, and the results are spliced to obtain the typhoon corpus word vector matrix incorporating semantic features.

The BiLSTM model is composed of two-layer networks of forward LSTM and backward LSTM. The outputs of the two LSTM layers at the same time were combined to obtain the eigenvalues, which integrate the forward and backward information of the sequence. The unit structure of LSTM is shown in Figure 5 and specific calculation process can be expressed as follows:

$$i_t = \sigma(W_{xi}x_t + W_{hi}h_{t-1} + W_{ci}c_{t-1} + b_i), \tag{2}$$

$$z_t = \tanh(W_{xc}x_t + W_{hc}h_{t-1} + b_c), \tag{3}$$

$$f_t = \sigma(W_{xf}x_t + W_{hf}h_{t-1} + W_{cf}c_{t-1} + b_f), \tag{4}$$

$$c_t = f_t c_{t-1} + i_t z_{t-1}, \tag{5}$$

$$o_t = \tanh(W_{xo}x_t + W_{ho}h_{t-1} + W_{co}c_t + b_o), \tag{6}$$

$$h_t = o_t \tanh(c_t), \tag{7}$$

where σ is activation function, W are the weighted matrices, b are the biases of LSTM, z represents the candidate for cell state, c represents the cell state, i, f, o represents the value of input gate, forget gate, and output gate, respectively, and we concatenate the forward h_t and backward h_t as the output of BiLSTM at time t .

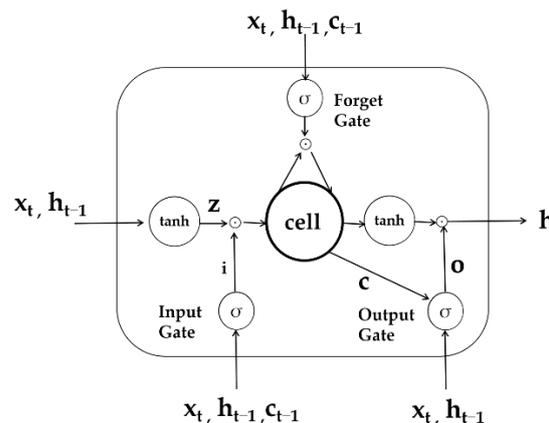


Figure 5. Structure of LSTM unit.

The CRF layer can consider the order of output tags and obtain the dependency between adjacent output tags. Based on the typhoon Web corpus, the optimal prediction labeling sequence results can be obtained. The maximum conditional likelihood estimation was used for the training of the CRF layer. For each input sequence $X = \{X_1, X_2, X_3, \dots, X_n\}$, the prediction score of the corresponding prediction sequence $Y = \{Y_1, Y_2, Y_3, \dots, Y_n\}$ was

defined as Formula (8) and the probability of the sequence Y for sequence X was calculated as Formula (9).

$$S(X, Y) = \sum_{i=0}^n A_{y_i, y_{i+1}} + \sum_{i=1}^n P_{i, y_i} \tag{8}$$

$$p(Y|X) = \frac{\exp(S(X, Y))}{\sum_{i=0}^n \exp(S(X, y_i))} \tag{9}$$

where A is the transfer matrix, A_{ij} represents the probability of moving from the i -th tag to the j -th tag, P_{ij} represents the probability that the tag at the i -th position is output as the j -th tag.

2.3. Data Fusion

The data source of this research was the Chinese Web corpus. Due to the diversity of Chinese descriptions, the same content may have multiple expressions. For instance, “Qiang jiang yu”, “Qiang jiang shui”, “Bao yu”, and “Te da bao yu” all represent rainstorm disasters in Chinese. This leads to a large amount of data duplication and redundancy in the results of typhoon event information extraction. Therefore, it was necessary to conduct data fusion on the event-elements obtained in the extraction stage.

In the present study, Word2Vec was used to calculate the semantic similarity of the event-elements, and a threshold was set based on the value of the similarity to achieve the fusion of disaster and emergency-related information. Word2Vec is a commonly used model for training word vectors, which can quantify the word into a dense real-valued vector in a low-dimensional vector space so as to realize the feature expression of text [26]. The information entities obtained in the extraction stage were mapped to word vectors to construct the bag-of-words of typhoon event-elements, and the transformation of information entities from semantic space to vector space was realized. The cosine value of the angle between different vectors was calculated. The larger the cosine value, the higher the semantic similarity of the two words. For two different information entities $m1$ and $m2$, the cosine similarity was calculated as Formula (10) [27].

$$S_{M(m1,m2)} = \frac{A \cdot B}{|A| \times |B|} = \frac{\sum_{i=1}^n (A_i \times B_i)}{\sqrt{\sum_{i=1}^n (A_i)^2} \times \sqrt{\sum_{i=1}^n (B_i)^2}} \tag{10}$$

where A and B are word vector representations of text $m1$ and $m2$, respectively.

3. Results

3.1. The Typhoon Disaster Chain Ontology Model

The construction of the Typhoon Disaster Chain Ontology Model mainly includes four elements: Typhoon Instance, Disaster Event Ontology Model, Disaster Exposure Ontology Model, and Emergency Response Ontology Model.

Typhoon Instance focuses on describing the state and properties of typhoon events at different times. It is the specific expression of disaster-inducing factors including unique attribute information, such as wind speed, wind force, rainfall, etc.

The Disaster Event Ontology Model describes the attribute information, such as time, location, type of natural disasters, caused by typhoon events as well as their relationships. Among them, rainstorms, windstorms, and storm surges are the main direct disasters caused by typhoons. There are correlations between various disasters, for example, a rainstorm can lead to geological disasters, such as debris flow, landslide, mountain torrents; storm surge can cause huge waves resulting in saltwater intrusion, which lead to urban waterlog in combination with a rainstorm.

The Disaster Exposure Ontology Model is the specific description of the situation of the main impacted objects of the disaster, which refers to the damage caused by the disaster to the natural and social environment of the affected area, including casualties, property losses, ecological impacts, impacts of social activities, etc.

The Emergency Response Ontology Model refers to the emergency response measures taken by emergency forces during the development of typhoon disasters, which can be divided into three stages according to the process of disaster: Pre-disaster, During-disaster, and post-disaster. Each stage has its own priority emergency tasks, such as pre-disaster early warning, urgent repair, and emergency rescue during the disaster, post-disaster recovery, and reconstruction, etc.

The partial descriptions of the concepts, properties, and relationships of each ontology model are shown in Tables 2–5, respectively, and the contents of Instance were translated into English. OWL language and Protégé tool were adopted for ontology modeling in this research. To be specific, the Protégé OWL API can be fully used for consistency maintenance to ensure that the description language of constructed ontology conforms to grammatical and logical rules and satisfies the customized domain rules. For example, based on the constraint of relationships, rainstorms induce floods, thus floods should occur after a rainstorm; based on conceptual constraints, missions, such as risk monitoring, disaster early warning, emergency plan, etc., should belong to pre-disaster emergency response. Part of the Typhoon Disaster Chain Ontology Model structure is shown in Figure 6.

Table 2. Part of Typhoon Instance description.

Properties of Typhoon Instance	Data Content	Instance (In en-US)
Location	...	hasLocation: Taishan, Jiangmen, Guangdong
Time	X (Year) X (Month) X (Date)	hasTime: 16 September 2018
Central Pressure	X hpa	hasCentralPressure: 945 hpa
Wind Intensity	Level X	hasWindIntensity: 14
Wind Speed	XX m/s	hasWindSpeed:45 m/s
...

Table 3. Part of Disaster Event description.

Type of Disaster	Properties of Disaster Event	Data Content	Relationships of Disaster Event	Instance (In en-US)
Storm surge	Location	...	(Storm surge)	hasLocation: Ningde, Fujian
	Time	X (Year) X (Month) X (Date)	-[r:Caused By]-> (Typhoon);	hasTime: 15 September 2018
	Storm Surge Elevation	X cm	(Storm surge) -[r:Induce]-> (Huge wave)	hasStormSurge: 40–110 cm

Huge wave	Location	...	(Huge wave)	hasLocation: Coastal areas of Guangdong
	Time	X (Year) X (Month) X (Date)	-[r:Caused By]-> (Storm Surge);	hasTime: 15 September 2018
	Wave Height	X m	(Huge wave) -[r:Induce]-> (Fishery losses)	hasWaveHeight: 4–7 m

Table 4. Part of Disaster Exposure description.

Type of Disaster Exposure	Properties of Disaster Exposure	Data Content	Relationships of Disaster Event	Instance (In en-US)
Personal casualty	Location	...	(Personal casualty) -[r:Caused By]-> (Typhoon)	hasLocation: Guangdong, Guangxi, Hainan
	Fatality	X (Number)		hasFatality: 4
	Relocation	X (Number)	hasRelocation: 1.419 million	
	
Property damage	Location	...	(Property damage) -[r:Caused By]-> (Typhoon);	hasLocation: Guangxi
	Direct economic loss	X (Yuan)		hasEconomicLoss: 113 million yuan
	Houses destroyed	X (Number)		hasHouseDestroyed: 100
	Agricultural losses	X (Hectares)	hasAgriculturalLosses: 2500 hectares	
	

Table 5. Part of Emergency Response description.

Emergency Phase	Aim of Emergency	Missions of Emergency
Pre-disaster	Disaster early warning, Disaster prevention	Risk monitoring, Disaster early warning, Emergency plan
During-disaster	Emergency response, Emergency disposal	Urgent repairs, Relocation and resettlement, Emergency rescue, Resource allocation
Post-disaster	Reconstruction and restoration, Loss assessment	Disaster assessment, Social activity restoration, Infrastructure reconstruction

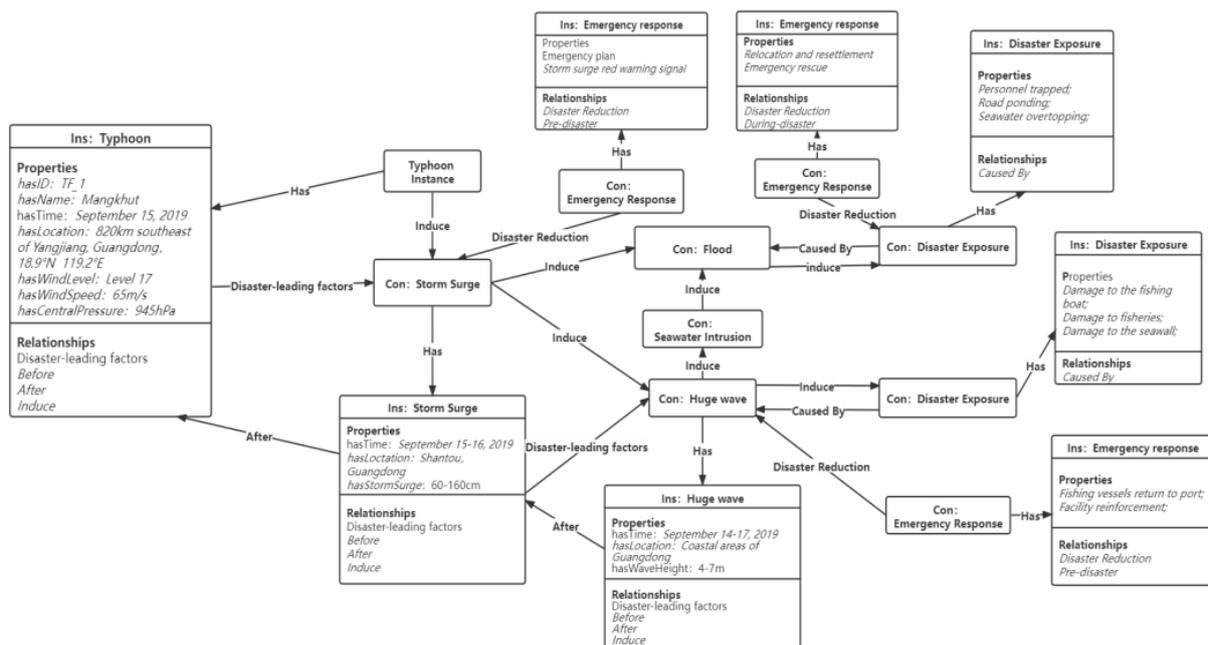


Figure 6. Part of the Typhoon Disaster Chain Ontology Model (TDCOM) structure.

3.2. Extraction Results of Typhoon Disaster Information

This research obtained 8361 typhoon news as corpus from websites, such as Weather China, Sohu News, and Sina Weibo. Jieba and LTP word segmentation tools were used to perform Chinese word segmentation, POS tagging, and dependency parsing, and then rule templates were used to match the corpus to achieve automatic annotation. Manual adjustments were made to correct minor errors of automatic annotation. In the experiment, 80% of the annotated corpus was used as the training set, 20% as the test set, and 25% of the corpus in the training set was used as the development set. The data set structure of the typhoon corpus is shown in Table 6.

Table 6. Dataset structure of typhoon corpus.

Dataset	Number of Characters	Number of Characters with Labels
Training Set	457,524	239,798
Dev. Set	154,001	84,627
Test Set	154,095	78,894

The model was trained and tested in the environment of Python 3.6 and Tensorflow 2.3 in the experiment. During the training, the model parameters were set as follows: Transformer contained 12 layers, the dimensions of the hidden layer were 768, Adam was used as the optimizer, the learning rate was 0.0001, batch_size was 30, the drop rate was 0.5, and epoch was set to 30. In the experiment, Precision(P), Recall (R), and F1-score were used to evaluate the performance of the typhoon information extraction model. The calculation of these three evaluation indicators are as follows:

$$P = \frac{N_{tp}}{N_p}, \quad (11)$$

$$R = \frac{N_{tp}}{N_t}, \quad (12)$$

$$F1 = \frac{2PR}{P + R}, \quad (13)$$

where N_{tp} is the number of event-elements correctly recognized, N_p is the sum of recognized event-elements which also contains the entities that do not match the annotation, and N_t is the total number of all annotated event-elements.

According to the indicators, the extraction effect of the BERT-BiLSTM-CRF model on typhoon event-elements in the test set could be obtained. In order to verify the superiority of the BERT-BiLSTM-CRF model, the CRF model, Bi-LSTM model, and Bi-LSTM-CRF model were also used on the same dataset for comparison. The results are shown in Table 7. It can be seen from Table 7 that for typhoon name, time, location, and disaster information, the BERT-BiLSTM-CRF model showed outstanding performance, with higher Precision, Recall, and F1-score. However, the recognition performance of the model for disaster exposure information was relatively poor, mainly because of the more complex expression of such event-elements. Specifically, disaster information exposure includes all kinds of loss information which increases the difficulty of information extraction.

Table 7. Effect of typhoon event-elements extraction.

Model	CRF			BiLSTM			BiLSTM-CRF			BERT-BiLSTM-CRF		
Typhoon Event-Element	P	R	F1	P	R	F1	P	R	F1	P	R	F1
NAME	0.9432	0.9708	0.9553	0.9503	0.9509	0.9506	0.9587	0.9726	0.9656	0.9827	0.9904	0.9863
TIME	0.8653	0.8217	0.8423	0.8854	0.8788	0.8821	0.9149	0.8995	0.9071	0.9476	0.9394	0.9432
LOC	0.6798	0.6353	0.6562	0.8028	0.8046	0.8037	0.8331	0.8494	0.8412	0.8950	0.9076	0.9011
ORG	0.6974	0.6270	0.6560	0.6352	0.7570	0.6907	0.8073	0.7848	0.7959	0.8612	0.8651	0.8598
STF	0.6743	0.5778	0.6216	0.8619	0.8935	0.8774	0.8929	0.9092	0.9010	0.9458	0.9446	0.9459
SZQ	0.6369	0.5755	0.6025	0.7366	0.7578	0.7471	0.8046	0.7760	0.7900	0.8535	0.8752	0.8637
SYJ	0.6977	0.5743	0.6046	0.7594	0.7275	0.748	0.7416	0.7767	0.7599	0.8655	0.8861	0.8802
VTF	0.9462	0.9728	0.9571	0.9676	0.9659	0.9668	0.9805	0.9910	0.9857	0.9799	0.9924	0.9854
Mean Value	0.7676	0.7194	0.7370	0.8249	0.8420	0.8333	0.8667	0.8699	0.8683	0.9164	0.9251	0.9207

A visual comparison of F1 scores on different models was conducted by using the radar diagram, and the results are shown in Figure 7. It can be seen from Figure 7 that the F1-scores of the BERT-BiLSTM-CRF model on the eight types of event-elements were all at a higher level, indicating that the performance of this model was better than others in the field of typhoon disaster. BERT can fully extract character-level, word-level, and sentence-level features, enhancing the generalization ability of the model. Thus, the BERT-BiLSTM-CRF model achieved the best information extraction effect, with the F1-score being 0.9207, which was 5.24% higher than the BiLSTM-CRF model.

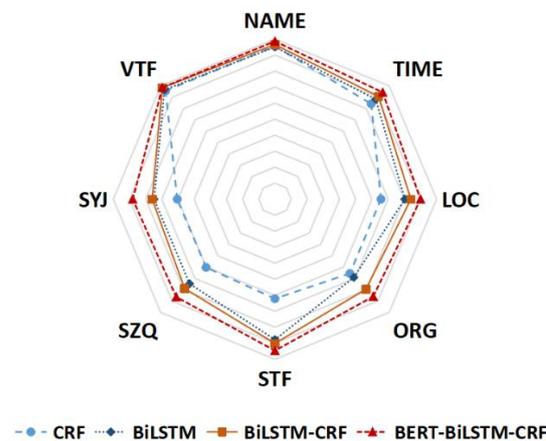


Figure 7. F1 scores of different event-elements on different models.

In addition, this study also compared the update of the F1-score in the first 20 epochs of each model. The results are shown in Figure 8. As can be seen, since CRF is a statistical-based sequence labeling model, the F1-score barely changed with the training. While the F1-score of the traditional neural network model was quite low at the initial stage of training, it rose and then stabilized at a high level after several learning iterations. After the introduction of BERT, the model achieved a higher F1-score at the initial stage of training, and gradually improved and stabilized at a high level. The BERT-BiLSTM-CRF information extraction model showed the best performance in the F1 value update.

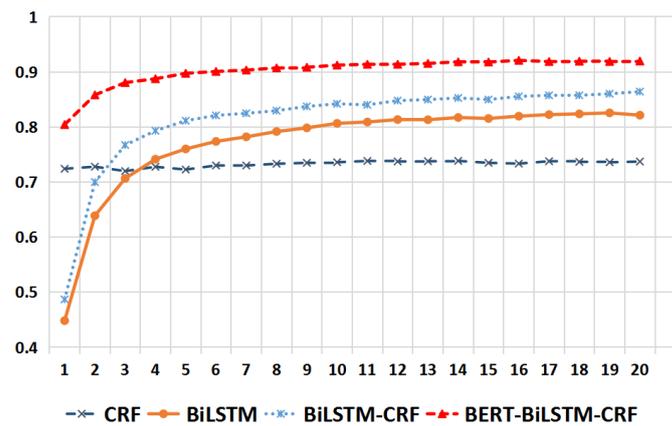


Figure 8. The comparison of each model’s updated F1 scores.

In this study, typhoon information was extracted by using the model designed in the experiment. The example of extraction results is shown in Figure 9, the Chinese text was translated into English. In the experiment, the obtained typhoon event-elements were regarded as nodes that were stored in the Neo4j graph database. Neo4j is a widely used open-source graph database, which stores connected data as a network structure of flexible graphs rather than traditional static tables [28]. Part of the results is shown in Figure 10 where the blue nodes represent typhoon attribute information, the green nodes represent emergency response information, the orange nodes represent disaster event information, and the red node represents disaster exposure information.

In en-US

Text: At 09:00 on the September 15th, the center of Typhoon Mangkhut was about 980 kilometers southeast of Yangjiang, Guangdong. The intensity has weakened to the Very Strong Typhoon level (50 m/s, wind intensity up to level 15). The National Marine Environmental Forecasting Center issued the Storm surge red warning signal.

Result:

```

<TF_ID = T_1>
  <NAME = "Typhoon Mangkhut">
  <TIME = "09:00 on the September 15th">
  <LOC = "980 kilometers southeast of Yangjiang, Guangdong.">
  <ORG = "The National Marine Environmental Forecasting Center">
  <STF = "Very Strong Typhoon">
  <STF = "50 m/s">
  <STF = "level 15">
  <SYJ = "Storm surge red warning signal">
  <VTF = "weakened" >
</TF_ID=T_1>

```

Figure 9. Results of typhoon event-elements extraction (In en-US).

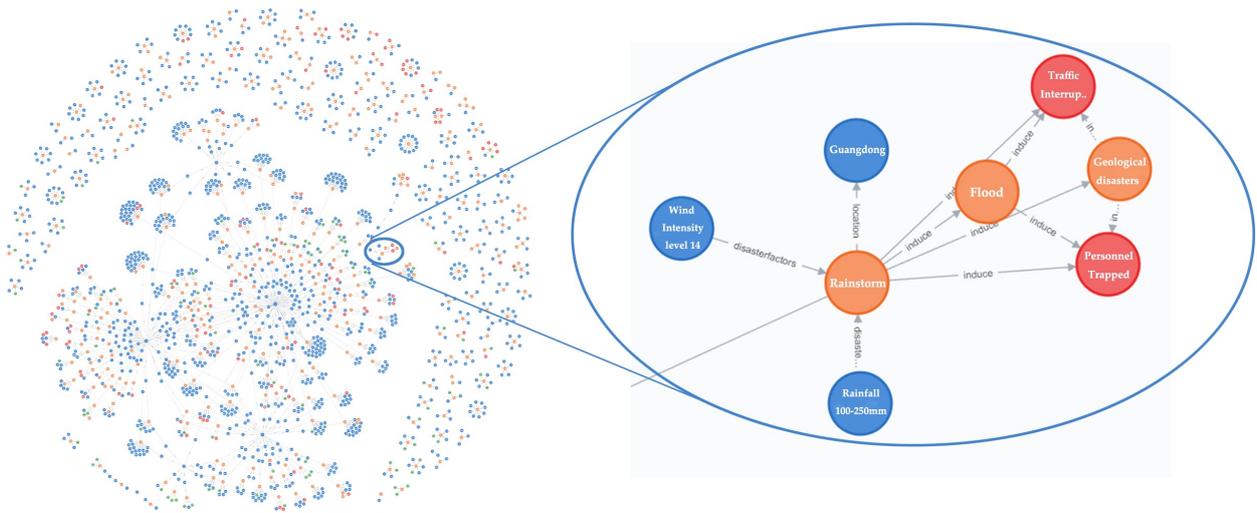


Figure 10. Part of the typhoon event graph in Neo4j database (In en-US).

3.3. Construction of the Disaster Chain for Typhoon Mangkhut

Based on the scheme mentioned above, this research obtained 2984 news and Weibo texts related to Typhoon Mangkhut. After preprocessing the text, the typhoon event-elements were obtained based on the model proposed in §2.2. According to the semantic similarity calculation method proposed in §2.3, the acquired typhoon information at the same time was fused. After several comparison experiments, the best fusion effect could be obtained when the threshold was 0.85, and the set of event-elements that were larger than the threshold was merged. At the end, based on the Typhoon Disaster Chain Ontology Model (TDCOM) constructed in §3.1, event-elements were expressed in structured form in chronological order. As shown in Figure 11, the disaster chain of Typhoon Mangkhut is constructed.



Figure 11. The disaster chain of Typhoon Mangkhut.

It can be seen from Figure 11 that Typhoon Mangkhut was generated in the Pacific Ocean on 7 September, passed through Luzon Island in the Philippines, and approached Guangdong, China on 15 September. It landed in Taishan City, Jiangmen, Guangdong at 17:00 on September 16 and the typhoon code was canceled at 20:00 on 17 September. During the period from 14 to 17 September, there were gales and huge waves in the South China Sea, the coastal areas of Guangdong, Fujian, Guangxi, and Hainan, and marine personnel and ships were evacuated. During 15 to 16 September, storm surges occurred along the coast of Fujian and Guangdong, and rainstorms began to occur in Guangdong, Guangxi, and Fujian. Among them, the Pearl River Estuary area was greatly affected, with trains, flights, and schools suspended. From 16 to 17 September, the impact of Typhoon Mangkhut expanded to southwest China, including Guizhou and Yunnan, and the impact of the disaster became more serious. In Guangdong and Guangxi, there were strong winds on land and a large number of trees fell causing damage to roads, wires, and vehicles. Geological disasters and floods caused by rainstorms occurred in many places, resulting in people being trapped, urban flooding, traffic control, water and electricity outages, etc. Local emergency departments took response procedures, such as rescue and urgent repairs. On 18 September, the influence of Typhoon Mangkhut was reduced, but there were still rainstorms in parts of Guangdong and southwest China.

Various regions began to carry out post-disaster recovery and reconstruction, such as the resumption of transportation, work and production.

4. Discussion

The typhoon disaster chain includes a series of secondary disasters, which are manifested as the extension of disasters in spatio-temporal dimensions and the continuous impact on human society. In order to form a comprehensive understanding of the process of typhoon disaster, this research designed a typical typhoon disaster chain construction scheme based on the Chinese Web corpus. OWL language and quintuple structure (Concept, Property, Relationship, Rule and Instance) were adopted at first, according to the idea of event-oriented ontology modeling. Then, based on the theory of natural disaster system, the Typhoon Disaster Chain Ontology Model (TDCOM) was constructed from four aspects: Typhoon Instance, Disaster Event Ontology Model, Disaster Exposure Ontology Model, and Emergency Response Ontology Model. Expressing the concepts and relationships in a specific domain, the ontology model enables departments and individuals to have a consistent understanding of typhoon disasters, and on this basis, realize the reuse and sharing of typhoon domain knowledge. TDCOM constructed in this research demonstrated its ability in generalization by fully describing the relationships between various elements of the disaster and comprehensively expressing the specific typhoon disaster chain.

For the demand of typhoon event information extraction, rule templates and Chinese dependency parsing were used to realize automatic annotation of the typhoon Web corpus in this research, which solved the problems of lacking in the model training dataset. On this basis, the BERT-BiLSTM-CRF typhoon information extraction model was constructed, and the typhoon event-elements were extracted efficiently from the typhoon Web corpus. The model achieved 91.64% precision, 92.51% recall, and 92.07% F1 score on the customized typhoon corpus dataset. Compared with traditional information extraction models, the performance of the proposed model in this study was greatly improved so that it can meet the application requirements.

In the end, Word2Vec was used to calculate the semantic similarity of extracted information, and the disaster chain of Typhoon Mangkhut was constructed through data fusion and structured expression of event-elements. The results showed that the TDCOM model and BERT-BiLSTM-CRF model can be used to build a typical typhoon disaster chain based on the disaster information hidden in the Web corpus, and then analyze the evolution process of typhoon disaster and its impact on social activities. This method addresses the limitations of relying too much on historical disaster data in previous studies, improves the

timeliness of data acquisition, and provides scientific support for disaster prevention and emergency rescue.

5. Conclusions

This research realized the construction of a typical typhoon disaster chain, which not only shed light on analyzing the evolution process of typhoon disasters but also provided a basis for exploring the impact of typhoon disasters on social activities. In addition, it provided scientific support for typhoon disaster prevention and emergency management. Hopefully, combined with practical applications to build an intelligent comprehensive disaster management and reduction system based on the emergency scenarios of the entire disaster process, this research can provide services, such as disaster information management, emergency information push, and assistant decision-making, etc. in the future.

However, the limitations of the research have to be underlined. First, the construction of the typhoon disaster chain was related to the comprehensive description of the disaster knowledge in the ontology. Nevertheless, the construction of the TDCOM in this research mainly refers to the existing national disaster classification standard documents. Some colloquial expressions in the Web corpus bring redundancy to the extraction of disaster information. In future studies, typhoon disaster knowledge involved on the Internet can be further summarized to enrich the ontology framework. Second, in this research the information extraction process was regarded as a sequence labeling task; hence, the Web corpus involved can be supplemented and the automatic annotation process can be further improved to increase the precision of information extraction.

Author Contributions: Conceptualization, methodology, software, validation, formal analysis, investigation, resources, data curation, writing—original draft preparation, H.L.; writing—review and editing, visualization, supervision, project administration, funding acquisition, Q.Z. and N.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Key R&D Program of China, grant number 2017YFC1405300.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The study did not report any data.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

1. Liu, W.; Xiao, S.; Sui, Y.; Zhou, J.; Gao, H. Analysis of natural disaster chain and chain-cutting disaster mitigation mode. *Chin. J. Rock Mech. Eng.* **2006**, *25*, 2675–2681.
2. Ye, J.Y.; Lin, G.F.; Zhang, M.F. Spatial Characteristics of Typhoon Disaster Chains in Fujian Province. *J. Fujian Norm. Univ. Nat. Sci. Ed.* **2014**, *30*, 99–106.
3. Hua, J.; Hu, H.F.; Chen, X.L.; Hou, Q.Q. Analysis of typhoon disaster chain and risk management in northward typhoon effecting Hebei. In Proceedings of the Fourth Symposium on Disaster Risk Analysis and Management in Chinese Littoral Regions (DRAMCLR 2019), Qingdao, China, 7 June 2019.
4. Zhong, S.; Su, G.; Wang, F.; Chen, J.; Zhang, F.; Huang, C.; Huang, Q.; Yuan, H. A preliminary research on incident chain modeling and analysis. *Commun. Comput. Inf. Sci.* **2013**, *399*, 171–180.
5. Liang, Z.; Fei, W.; Zheng, X. Complex network construction method to extract the nature disaster chain based on data mining. In Proceedings of the 2017 7th IEEE International Conference on Electronics Information and Emergency Communication (ICEIEC), Macau, China, 21–23 July 2017.
6. Dunbar, P.K. Increasing public awareness of natural hazards via the Internet. *Nat. Hazards* **2007**, *42*, 529–536. [[CrossRef](#)]
7. Peduzzi, P.; Herold, H. Mapping Disastrous Natural Hazards Using Global Datasets. *Nat. Hazards* **2005**, *35*, 265–289. [[CrossRef](#)]
8. Kuzey, E.; Vreeken, J.; Weikum, G. A Fresh Look on Knowledge Bases: Distilling Named Events from News. In Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management, Shanghai, China, 3–7 November 2014; pp. 1689–1698.

9. Trivedi, R.; Dai, H.; Wang, Y.; Son, L. Know-Evolve: Deep Temporal Reasoning for Dynamic Knowledge Graphs. In Proceedings of the 34th International Conference on Machine Learning, Sydney, Australia, 6–11 August 2017; Volume 70, pp. 3462–3471.
10. Gruber, T.R. Toward principles for the design of ontologies used for knowledge sharing? *Int. J. Hum.-Comput. Stud.* **1995**, *43*, 907–928. [[CrossRef](#)]
11. Studer, R.; Benjamins, V.R.; Fensel, D. Knowledge Engineering: Principles and methods. *Data Knowl. Eng.* **1998**, *25*, 161–197. [[CrossRef](#)]
12. Shi, G.; Barker, K. Extraction of geospatial information on the Web for GIS applications. In Proceedings of the IEEE 10th International Conference on Cognitive Informatics and Cognitive Computing (ICCI-CC'11), Banff, AB, Canada, 18–20 August 2011; pp. 41–48.
13. Li, B.; Liu, J.; Shi, L.; Wang, Z. A method of constructing geo-object ontology in disaster system for prevention and decrease. In Proceedings of the International Symposium on Spatial Analysis, Spatial-Temporal Data Modeling, and Data Mining, Wuhan, China, 13–14 October 2009.
14. Lu, H.; Xian-gang, L.; Hong-qi, C.; Jing, P.; Yi, C. A Visualization Model of Geological Disaster Emergency Scheme Based on Ontology. *Open Cybern. Syst. J.* **2014**, *8*, 393–398. [[CrossRef](#)]
15. Madhyastha, H.V.; Balakrishnan, N.; Ramakrishnan, K.R. Event information extraction using link grammar. In Proceedings of the Seventeenth Workshop on Parallel and Distributed Simulation 2003, San Diego, CA, USA, 10–13 June 2003; pp. 16–22.
16. Sagcan, M.; Karagoz, P. Toponym Recognition in Social Media for Estimating the Location of Events. In Proceedings of the 2015 IEEE International Conference on Data Mining Workshop (ICDMW), Atlantic City, NJ, USA, 14–17 November 2015; pp. 33–39. [[CrossRef](#)]
17. Sagcan, M.; Karagoz, P. Neural Architectures for Named Entity Recognition. 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. 260–270.
18. Chiu, J.; Nichols, E. Named Entity Recognition with Bidirectional LSTM-CNNs. *Trans. Assoc. Comput. Linguist.* **2016**, *4*, 357–370. [[CrossRef](#)]
19. Huang, Z.; Wei, X.; Kai, Y. Bidirectional LSTM-CRF Models for Sequence Tagging. *Comput. Sci.* **2015**, *3*, 56–69.
20. Ma, X.; Hovy, E. End-to-end Sequence Labeling via Bi-directional LSTM-CNNs-CRF. In Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics, Berlin, Germany, 7–12 August 2016; pp. 1064–1074.
21. Xu, M.; Zhang, X.; Guo, L. Jointly Detecting and Extracting Social Events from Twitter Using Gated BiLSTM-CRF. *IEEE Access* **2019**, *7*, 148462–148471. [[CrossRef](#)]
22. Devlin, J.; Chang, M.; Lee, K.; Toutanova, K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies 2019, Minneapolis, MN, USA, 2–7 June 2019; pp. 4171–4186.
23. Zhang, W.; Jiang, S.; Zhao, S.; Hou, K.; Liu, Y.; Zhang, L. A BERT-BiLSTM-CRF Model for Chinese Electronic Medical Records Named Entity Recognition. In Proceedings of the 2019 12th International Conference on Intelligent Computation Technology and Automation (ICICTA), Xiangtan, China, 26–27 October 2019; pp. 166–169.
24. Hou, J.; Li, X.; Yao, H.; Sun, H.; Mai, T.; Zhu, R. BERT-Based Chinese Relation Extraction for Public Security. *IEEE Access* **2020**, *8*, 132367–132375. [[CrossRef](#)]
25. Li, Z.; Zhang, M.; Che, W.; Liu, T.; Chen, W.; Li, H. Joint models for Chinese POS tagging and dependency parsing. In Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing, EMNLP 2011, Edinburgh, UK, 27–31 July 2011; pp. 274–286.
26. Suleiman, D.; Awajan, A.A.; al Etaiwi, W. Arabic Text Keywords Extraction using Word2vec. In Proceedings of the 2019 2nd International Conference on new Trends in Computing Sciences (ICTCS), Amman, Jordan, 9–11 October 2019; pp. 1–7.
27. Guo, W.; Zeng, Q.; Duan, H.; Ni, W.; Liu, C. Process-Extraction-Based Text Similarity Measure for Emergency Response Plans. *Expert Syst. Appl.* **2021**, *1*, 115301. [[CrossRef](#)]
28. Partner, J.; Vukotic, A.; Watt, N. *Neo4j in Action*; Manning Publications Co.: Shelter Island, NY, USA, 2014.