



**HAL**  
open science

# Information Tracking Extraction for Emergency Scenario Response

Hua Zhao, Xiaoqian Li, Peixin Zhang, Zhengguo Song

► **To cite this version:**

Hua Zhao, Xiaoqian Li, Peixin Zhang, Zhengguo Song. Information Tracking Extraction for Emergency Scenario Response. 12th International Conference on Intelligent Information Processing (IIP), May 2022, Qingdao, China. pp.203-215, 10.1007/978-3-031-03948-5\_17. hal-04178716

**HAL Id: hal-04178716**

**<https://inria.hal.science/hal-04178716v1>**

Submitted on 8 Aug 2023

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License



This document is the original author manuscript of a paper submitted to an IFIP conference proceedings or other IFIP publication by Springer Nature. As such, there may be some differences in the official published version of the paper. Such differences, if any, are usually due to reformatting during preparation for publication or minor corrections made by the author(s) during final proofreading of the publication manuscript.

# Information Tracking Extraction for Emergency Scenario Response

Hua Zhao <sup>\*[0000-0001-6467-6892]</sup>, Xiaoqian Li, Peixin Zhang, Zhengguo Song

(College of Computer Science and Engineering, Shandong University of Science and Technology, Qingdao, 266590, China)

huamolun@163.com

**Abstract.** With the frequent occurrences of emergencies, scenario response is proposed by researchers for emergency management. On the one hand, extracting emergency-related information is necessary and important for describing emergency scenario correctly; On the other hand, some emergency-related information evolve dynamically with time going on. To this end, a method for information extraction based on event tracking (called as information tracking extraction) for emergency scenario response is proposed. Based on the emergency tracking, it firstly judges the type of the emergency; secondly extracts time, location, casualty and loss based on multi strategies; finally adopts E-charts to visualize the extraction results. Experimental results show the proposed method can achieve high precision and recall, and the visualization of the tracking extraction results can display the dynamic evolution process better, which is very helpful to the scientific decision-making in the emergency management.

**Keywords:** Emergency, Information extraction, Event track, Scenario response.

## 1 Introduction

In recent years, there happened a series of emergencies, such as USA "911" terrorist incident, the Indian Ocean tsunami, the "5.12" Wenchuan earthquake and the "4.20" Lushan earthquake in China, which show that we are now in a social period with frequent emergencies. The frequent occurrence of emergency poses severe challenges to the emergency response and management. Many researches on emergency decision support have been carried out [1, 2]. Relevant researches showed that the traditional "prediction-response" management model is already difficult to adapt to the high timeliness, comprehensiveness, dynamic and interactive requirements of emergency management, and then proposed a new "scenario-response" emergency management model. The premise of the "scenario-response" management model is to monitor and analyze the development situation of emergency comprehensively, that is, to

---

\* Corresponding Author

accurately depict the current "scenario" of emergency, in which the information related to emergency is a key dimension.

For the above-mentioned reasons, this paper aims to extract emergency-related information from unformatted news reports, including time, location, casualty, loss and so on, all of which are indispensable dimensions for describing emergency scenario accurately. Based on the summary of existing research results and the analysis of emergency news reports, we find that compared with traditional information extraction, emergency-related information extraction has its own characteristics. Firstly, the emergency information has a strong timeliness, and emergency news reports are often published on the day when the emergency occurs. Secondly, the format of ordinary text is often diversified, and cannot design uniform rules for extraction. But the format of emergency news reports are highly standardized, not only their contents generally conform to the news standard, but also their attributes are relatively uniform, which ensures that rule-based methods can be used for emergency-related information extraction.

The contributions of this paper are as follows:

(1) The concept of information tracking extraction is proposed. Existing studies neglect the continuity and dynamics of emergency, that is, an emergency usually lasts for a certain period of time after it occurs, and meanwhile, some emergency-related information (casualties, losses, etc.) evolves dynamically with time going on.

(2) Based on the topic tracking method, an emergency tracking algorithm is proposed. The tracking algorithm represents the emergency news report as a vector space model, and draws a conclusion whether they are relevant by comparing the similarity between news reports with the threshold.

(3) A multistrategy-based information extraction method is proposed and implemented, which mainly includes the extraction of the world, location, casualties, and economic losses and so on.

## **2 Related Work**

The goal of information extraction is to obtain target information accurately and quickly from a large amount of data, thereby improving data utilization. Generally speaking, information extraction methods can be divided into statistical methods and rule-based methods [3]. With the frequent occurrence of emergency and the urgent need of emergency management in recent years, the research on emergency information extraction has gradually attracted extensive attentions of researchers at home and abroad.

Relevant studies abroad have been carried out earlier, for example, D Brunner [4] used emergency information extraction results in post-disaster loss assessment. Amaief et al. [5] extracted emergency information based on Ontology and maximum entropy statistical model, and then sent the results to an emergency response system based on mobile terminal.

Domestic research on emergency information extraction is relatively late, but some excellent results have been achieved. Chen et al. [6] proposed a method based

on the combination of rules and machine learning to extract chemical emergency information. Michele et al. [7] first identified the three-level event framework system through the analysis of the composition of the main and auxiliary information structure chain, then identified the event keywords, and finally identified and classified the event attributes. Jiang D. L. [8] implemented a rule matching-based method for extracting information from emergency news reports. Wang N. [9] introduced knowledge element into emergency field, designed customizable emergency case information extraction template, and then combined with rule matching, realized the emergency case information extraction method based on knowledge element.

### 3 Emergency Information Tracking Extraction Framework

The flow chart of our emergency information tracking extraction framework is shown in Fig. 1. It consists of three parts: data acquisition and pre-processing; information tracking extraction; and results visualization.

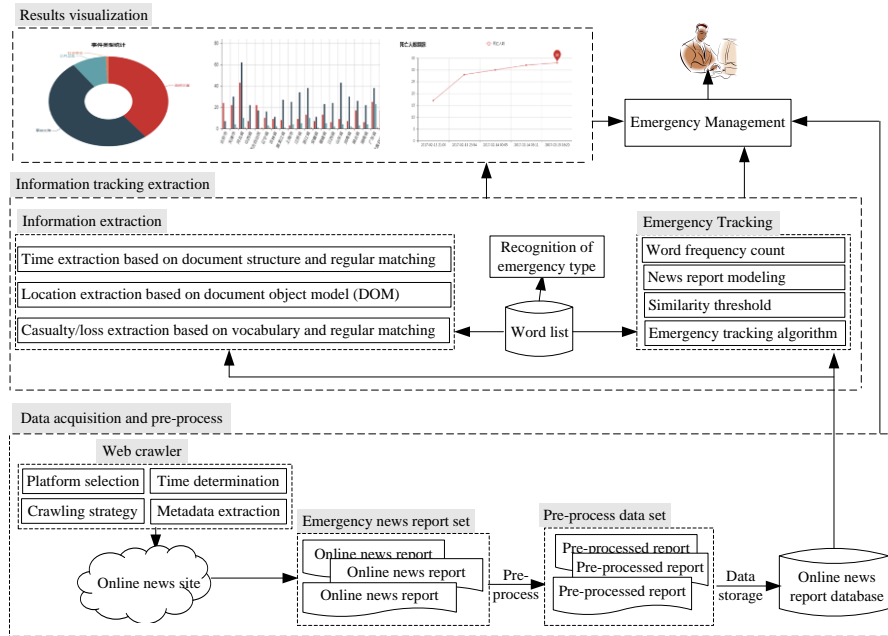


Fig. 1. Emergency information tracking extraction framework

**Information tracking extraction.** This is the core module of this paper, which is mainly responsible for information extraction, emergency type identification and event tracking. From Fig. 1, we can see that our proposed information tracking extraction method is to extract information based on emergency tracking. We will focus on this module in Section 4.

**Results visualization.** This module visualizes the results of information tracking extraction. The main purpose of data visualization is to use graphical methods and means to clearly and effectively convey relevant information and express relevant data. This paper uses Baidu's open source E-Charts visualization library for data visualization.

## 4 Emergency Information Tracking Extraction Method

### 4.1 Emergency Tracking Method

Based on the idea of topic tracking [10], this paper designs and implements a simple and effective method for tracking incidents. The steps are shown in Algorithm 1.

Equation (1) is a kind of TFIDF method [11] and Equation (2) is a classical cosine function. They are as follows:

$$sw_i = tf\_S(k_i) \times \log \frac{|ST|}{\delta(k_i)} \quad (1)$$

$$\cos(S, N_i) = \frac{\sum_{i=1}^n (sw_i \times sni_i)}{\sqrt{\sum_{i=1}^n sw_i^2} \times \sqrt{\sum_{i=1}^n sni_i^2}} \quad (2)$$

Where  $tf\_S(k_i)$  represents the word frequency of  $k_i$  in  $S$ ,  $|ST|$  denotes the number of news reports contained in  $ST$  and  $\delta(k_i)$  represents the number of news reports containing  $k_i$  in the set  $ST$ .  $sni_i$  can be obtained in the same way. Our experiments show that when the similarity threshold is set to 0.6, the performance of the emergency tracking algorithm is the best.

---

#### Algorithm 1 Emergency Tracking Algorithm

---

**Input:** A News Report  $S$ , News Report Collection  $ST$

**Output:** News reports related to  $S$

1. Repeat the following steps until all the news reports in  $ST$  are processed.
  2. Choose a piece of news report from  $ST$  as  $N_i (1 \leq i \leq |ST|)$ .
  3. Establish the vector space models of the news reports' title, assuming that  $T_S = (k_1, sw_1; k_2, sw_2; \dots; k_n, sw_n)$  represents the vector space model of the title of  $S$ , and the vector space model of the title of  $N_i$  is expressed as:  $T_{N_i} = (k_1, sni_1; k_2, sni_2; \dots; k_n, sni_n)$ , where  $k_i$  represents the keywords,  $sw_i$  and  $sni_i$  represent the weight of  $k_i$  in  $S$  and  $N_i$  respectively, which are computed using Equation (1).
  4. Calculate the similarity between  $T_S$  and  $T_{N_i}$  using Equation (2) and mark the similarity as  $\cos(S, N_i)$ .
  5. If  $\cos(S, N_i)$  is greater than or equal to the predetermined similarity threshold, then  $N_i$  is related to  $S$ , otherwise they are irrelevant.
-

## 4.2 Multi-Strategy Based Information Extraction

**Emergency Type Identification.** Recognize emergency type precisely is the premise of other information extraction. Because we just focus on emergency-related information, so in order to improve the efficiency of the system, only when an event is judged as an emergency can the relevant information be extracted. In October 2012, the General Administration of Quality Supervision, Inspection and Quarantine and the National Standardization Management Committee jointly issued the National Standard for Classification and Code of Natural Disasters, which was drafted by the National Disaster Reduction Center of the Civil Affairs Bureau. According to the standard, the natural disasters can be divided into 5 upper categories which include meteorological and hydrological disasters, geological and earthquake disasters, marine disasters, biological disasters and ecological environmental disasters. And these 5 upper natural disasters categories are then divided into 39 lower categories, which include drought, flood, and so on. Besides natural disasters, we add other three kinds of emergencies, which are accident, public health and social security, and then we collect the lower categories for these three emergencies, for example nuclear accident, poisoning and attack etc.

Type identification can be realized by classification method [12, 13] or cluster method [14]. Based on the keywords (emergency names) listed in Table 1, we use the following rules to identify the emergency type: If one of the above keyword appears in a news report, the news report is considered to belong to the emergency to which the keyword belongs. On the contrary, if there isn't any keyword appears in the news report, the news report doesn't belong to any kind of emergency.

**Time Extraction.** In general, there are three kinds of time expressions in a news report, which are the time of the event occurred, the time of the event reported, and the time of the news report released. The time of the news report released can be directly obtained through the web crawler. And according to the news report writing specification and style, the time of the event reported is usually mentioned in the first sentence of the news report, so it also can be extracted easily. So, the difficulty of time information extraction is the extraction of the time of the event occurred.

Through the analysis of a large number of news reports, we find that in online news reports, there are usually the following kinds of formats for the time of the event reported, and the time of the news report released:

- YYYY-MM-DD, such as “2018-08-01”
- YYYY/MM/DD, such as “2018/08/01”
- YYYY.MM.DD, such as “2018.08.01”

Based on the above foundation, we use the regular “ $d\{4\}(\-|\-|/|\.)\}d\{1,2\}\I\}d\{1,2\}$ ” to extract the time of the event reported, and the time of the news report released in this paper, in which the meaning of symbols is shown in Table 1. And the symbols and their meaning are also applicable to the subsequent regulars appearing in this paper.

**Table 1.** The regular expression syntax

Symbol	Meaning
\	Mark a character immediately following "\" as a special character, or as a literal character.
*	Match the sub-expression before "*" 0 times or N times, which is equivalent to {0,}.
+	Match sub-expressions in front of "+" once or N times, which is equivalent to {1,}.
?	Match the sub-expression before "?" 0 times or 1 time, which is equivalent to {0,1}.
{x,y}	Match the previous expression at least x times and at most y times.
(text)	Match the content of text and get the match, and the brackets can be understood as grouping.

As mentioned earlier, there are no special fixed formats and rules about the time of the event occurred, and this paper designs the following regulars for its extraction:

$$((num)())+(time)?((num)(time))+(time)?$$

where,

*num*: one, time two, twice, three, four, five, six, seven, eight, nine, ten, 0, 1, 2, 3, 4, 5, 6, 7, 8, 9

*time*: year, month, day, hour, hour, minute, second, morning, before dawn, a. m, afternoon, evening, night, dusk

Based on the above regulars, we adopt the following method to extract the time of the event occurred: all the time expressions appearing in the news report are extracted based on the above two rules, and then normalized into a unified format; and the earliest time is selected as the occurrence time of the emergency.

**Location Extraction.**Through analyzing a large number of news reports, we find out that location names usually appear in the following two places in news report:

- News headlines: the locations appearing in news headlines are usually the names of some provinces or municipalities;
- First sentence of news reports: the locations appearing in the first sentence are usually specific to the cities, districts, or counties.

In order to extract the locations, we create the hierarchical structure (provincial, municipal and district) of China administrative area, and use XML structure to store this hierarchy. Then we use the Document Object Model (DOM) to parse the above-mentioned location XML file, the steps are as follows:

- Create the object of DOM tree;
- Create the root node of DOM tree, that is, the national node, and use DOM object to add the root node;
- Use DOM objects to create the sub-nodes at next lower level, i.e. provincial and urban nodes;



- Add the next lower level child node to the object of the parent node to form the parent-child cascade relationship.

Based on the above DOM tree, we use the following method to extract the locations appearing in the news report:

- if the district (county) location name is found in the news report, the corresponding province/city/district (county) name is returned by searching the parent node in the DOM tree;
- if only the city location name is matched, the parent node in the DOM tree is searched, and return the province/city list, the district (county) part is blank;
- if only the provincial level place name is matched, the province list is returned, and the city/district (county) part is blank.

**Personnel Casualty Extraction.** Unexpected emergency would cause varying degree casualties, slight or serious. This is an important basis for launching different levels of emergency reserve plans. Accurately extracting casualty-related information is essential for emergency scientific decision-making. Through analysis, it is found that in the news reports, the format of casualty-related information is relatively fixed, such as "about 30 people with different degrees of injury", "two workers rescue invalid death". So it is suitable to use rule matching method to extract casualty-related information. To this end, this paper designed the following regular expressions, and the corresponding feature words are shown in Table 2.

*Death: (Degree adverbs)?(Numeral)+(Degree adverbs)?(Numeral)?(Disaster victims)?(Other words)?(Death keywords)*

*Injured: (Degree adverbs)?(Numeral)+(Degree adverbs)?(Numeral)?(Disaster victims)?(Other words)?(Injury keywords)*

**Table 2.** The feature vocabulary of casualty-related information extraction rules

Type	Characteristic words
Degree adverbs	about, almost, around, more than, at least.....
Numeral	one, two, three, four, five, six, seven, eight, nine, ten, hundred, thousand, ten thousand, 0, 1, 2, 3, 4, 5, 6, 7, 8, 9.....
Disaster victims	person, personnel, children, students, teachers, residents, teachers and students, workers, passengers.....
Death keywords	dead, death, casualty, wreck, lose life, drowning, be buried, die in an accident, be murdered, suffocation, shoot dead, killing, no sign of life.....
Injury keywords	wounded, injuries, serious injuries, minor injuries, infections, traps

The above casualty-related information extraction rules will result in a number of string, such as "One hundred and thirty people are injured, no one is dead". In order to facilitate subsequent statistical analysis, formatting is required to convert Chinese numbers and special words into Arabic numerals. For example, for the above results, "one hundred and thirty" needs to be converted to 130, and "no one" needs to be converted to 0.

**Economic Loss Extraction.** Emergency would also lead to a certain level of economic loss, which is a very important dimension for the correct assessment of the impact of disasters, and also an important consideration in emergency management. In the Web news report, the format of economic loss is relatively fixed, such as “the affected area is about 5.4 hectares” and “the economic loss is 150 million yuan”. Similar to the above casualty attribute extraction, this paper extracts economic loss information based on regular matching. The designed rules are as follows, and the corresponding feature vocabulary is shown in Table 3.

*(Extent of loss)+(Degree adverbs)?(Numeral)+(\,.)?(Numeral)?(Degree adverbs)?(Quantifiers)+*

**Table 3.** The feature vocabulary of economic loss information extraction rule

Type	Characteristic words
Degree adverbs	about, almost, around, more than, at least.....
Numeral	one, two, three, four, five, six, seven, eight, nine, ten, hundred, thousand, ten thousand, 0, 1, 2, 3, 4, 5, 6, 7, 8, 9.....
Extent of loss	economic losses, direct losses, indirect losses, collapsed houses, fields, livestock, poultry, landslides, damaged roads, affected areas, burned areas.....
Quantifiers	dollars, yuan, RMB, hectares, square meters, mu, kilometers, households, heads.....

## 5 Experiments and Analysis

### 5.1 Evaluation Metrics

In this paper, precision and recall are used as evaluation metrics for the results of emergency information extraction. Among them, the precision rate refers to the ratio of the correct amount of data extracted to the total number of extracted data. The recall rate refers to the correct amount of data extracted and the total number in database. The calculation equations for precision and recall are shown in Equation (3) and Equation (4), respectively.

$$Precision = \frac{A}{B} \times 100\% \quad (3)$$

$$Recall = \frac{A}{C} \times 100\% \quad (4)$$

Where A is the correct number in the information extraction result, B is the total number of extracted information, and C is the total number in database.

All our experiments are implemented in Python environment within Windows 10 operating system, and the main frequency is 3.3GHz, memory is 8GB.

## 5.2 Experimental Results and Analysis

Firstly, we evaluate our emergency tracking method. In order to evaluate the tracking performance, we randomly select 1,000 news reports from database, and then let three students to mark the emergency included in the each news report. And then, we select randomly 600 news reports from these 1,000 reports as the training corpus, and the other 400 news reports as the evaluation corpus. Based on training, we find that when the similarity threshold is set to 0.6, the performance of the emergency tracking method is best. We use this similarity threshold for the test, and experiments show that the tracking precision and recall are 0.8876 and 0.9023, respectively, which indicates that our proposed tracking method is simple and effective. The result of this experiment is better than that of the traditional topic tracking experiment. We think that this is because the format of emergency news report is more consistent and the news language is more standardized.

Secondly, we evaluate the performance of our method for emergency type identification and information tracking. Again, in order to evaluate, we randomly select 500 news reports from the database, and let three people to mark the emergency type, time, location, casualty and loss for each news report. The evaluation results are shown in Table 4.

**Table 4. Experimental results of emergency type identification and information extraction**

Information	Theoretical value	Actual value	Correct value	Precision	Recall
Type	500	478	466	0.9748	0.9320
Time	478	478	466	0.9748	0.9748
Location	478	477	341	0.7148	0.7133
Casualty	478	478	470	0.9832	0.9832
Loss	44	40	37	0.9250	0.8409

It can be seen from the results that:

(1) The precision and recall rate of emergency type identification are relatively high, which indicates that the types of keywords collected in this paper are comprehensive and can cover most types of emergency. Through the analysis of the experimental results, it is found that the reason for the identification error is mainly because some news reports contains keywords of cross-type emergencies, especially fire accidents. Some fires are caused by human factors and should belong to “accident disaster”, while some others, such as forests fires, are caused by high temperatures and should be classified as “natural disasters”, which lead to inaccurate type identification.

(2) The result of time extraction is better, which shows that the method of rule matching can play a very good role because the time expression format in Web news reports is relatively fixed and unified. The reason for the time extraction error is mainly because some related events are sometimes reviewed in the news report, so that there will be multiple times in the news report other than the time of the current emergency, resulting in the extraction error.

(3) The precision and recall rate of location extraction is low. Based on analysis, it is found that the algorithm of the Chinese word segmentation has certain limitations and cannot identify the location nouns perfectly. For example, it is easy to divide “Xinhuanet” into “Xinhua” and “net”, so the location extraction method will extract “Xinhua” as “Xinhua County”, which results in extraction error.

(4) Casualty extraction got the highest precision and recall, both 98.32%. The high precision is mainly due to the uniformity of the description of casualty attributes in news reports, while the precision is not 100% mainly because some news reports don't describe casualties directly, such as “the casualty died unfortunately on the way to the hospital”, in this case, the use of regular expressions cannot identify and derive a person's death.

(5) The performance of loss information extraction is also better, mainly due to the uniformity of loss description in news reports.

From the above experimental results, we can find that the rule-based information extraction method proposed in this paper is not only simple, but also can achieve good results.

### 5.3 Results Visualization

In order to display the information extraction results more intuitively, this paper uses E-charts as a tool for visualization. Based on the emergency type identification results of 1,634 news reports, a proportion statistical map of emergency types is obtained which is shown in Fig. 2. According to the identification results, accident disaster and natural disaster accounted for the largest proportion, which are 51% and 39% respectively; the proportions of public health accident and social safety accident are lower, which are 9% and 1%, respectively.

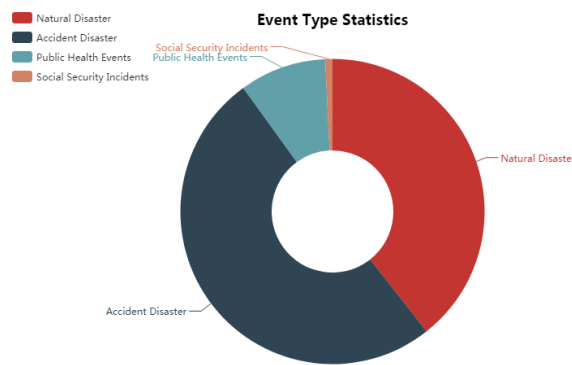
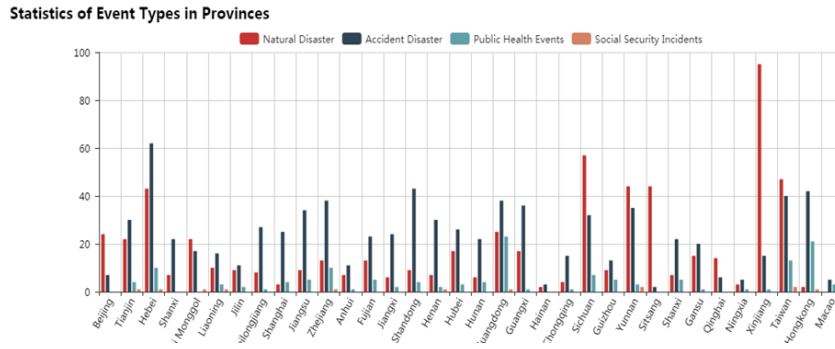


Fig 2. The proportion chart of different type of emergency

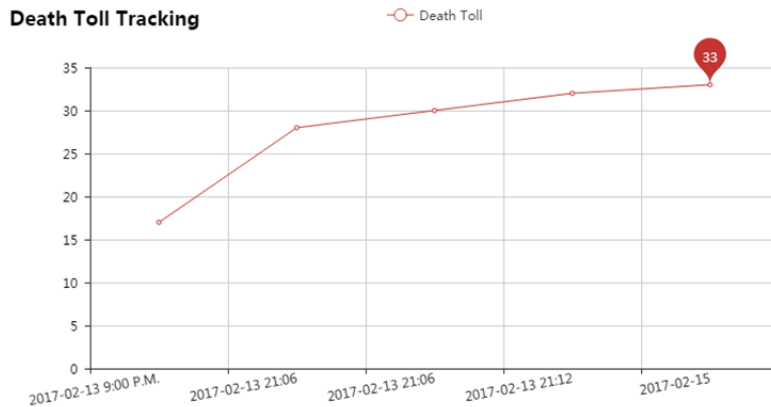
Based on the location recognition results of 1,634 news reports, the statistical proportion map of each event type in each province is shown in Fig. 3. As can be seen

from the figure, there are regional differences in the proportion of each type of emergency in each province. There are more natural disasters in Xinjiang and Sichuan Province, but more accident disasters in Hebei Province.



**Fig 3.** The statistical charts of the occurrence of each event type in each province

Based on tracking "Taiwan Tourist Car Overturning Accident", the time extraction results and the death toll extraction results, a broken line graph shown in Fig. 4 is obtained. Using the broken line chart, we can see the dynamic changes of the relevant indicators with time more intuitively, which can better assist emergency decision-making.



**Fig 4.** An example chart of a broken line for tracking the number of deaths

## 6 Conclusions

In order to meet the high timeliness, comprehensiveness, dynamics and interactivity requirements of emergency management, a new "scenario-response" emergency

management model was proposed. Accurate access to emergency-related information is an important dimension in building “scenario-response” emergency management model, and information such as casualties is dynamically changing. To this end, a method for emergency information tracking extraction is proposed. Based on the automatic tracking of emergency, the method first determines the type of emergency, and then extracts information such as time, location, casualty, and loss based on rule matching. Finally, results are displayed using E-charts. The experimental results show that because the news report has certain format when reporting the information related to the emergency, the extraction method based on rule matching can achieve good performance, and the dynamic change process of related information is displayed based on event tracking. It can help decision makers better grasp the development trend of emergency and provide technical support for scientific decision-making.

## **Acknowledgements**

This research was supported by the Shandong Natural Science Foundation Project (No. ZR2021MG038); Special Study on Cultural Tourism of Shandong Social Science Planning (No. 21CLYJ32); 2019 Qingdao Philosophy and Social Science Planning Research Project (No. QDSKL1901124); Shandong Postgraduate Education Quality Improvement Plan (No.SDYJG19075); Shandong Education Teaching Research Key Project (No.2021JXZ010); Qingdao City Philosophy and Social Science Planning Project (No.QDSKL2001128); National Statistical Science Research Project (No.2021LY053) and Shandong University of Science and Technology Education and Teaching Research "Stars Plan" Project (No.QX2021M29).

## **References**

1. Shi, W., Wang, H. W., He, S. Y.: Sentiment analysis of Chinese microblogging based on sentiment ontology: a case study of ‘7.23 Wenzhou Train Collision’. *Connection Science*, Vol 25, Issue 4, pp.161-178(2013).
2. Tran, Q. K., Song, S. K.: Learning Pattern of Hurricane Damage Levels Using Semantic Web Resources. *International Journal of Computational Science and Engineering*, Vol.20 No.4, pp.492-500(2019).
3. Guo, X. Y., He, T. T.: Survey about Research on Information Extraction. *Computer Science*, Vol, 42, No.2, pp.14-17 (2015).
4. Brunner, D.: Advanced Methods for Building Information Extraction from Very High Resolution SAR Data to Support Emergency Response. Unpublished PhD thesis, University of Trento, UniTrento, Italy(2009).
5. Amailef, K., Lu, J.: Mobile-based Emergency Response System Using Ontology-Supported Information Extraction. *Handbook on Decision*, Vol. 2, pp. 429-449(2012).
6. Chen, Z., Zheng, S.: Research on Chemical Emergency Information Extraction based on Multi Algorithm Fusion. *Computer & Digital Engineering*, Vol. 46, No.2, pp. 264-268(2018).

7. Michele, B., Michael, J. C., Stephen, S., Matt, B., Oren, E.: Open Information Extraction from The Web. Proceedings of the 2007 International Joint Conference on Artificial Intelligence, San Francisco, USA, pp. 2670-2676( 2007).
8. Jiang, D. L.: Research on Extraction of Emergency Event Information based on Rules Matching. Computer Engineering and Design, Vol. 31, No. 14, pp. 3294-3297(2010).
9. Wang, N., Chen, Y., Guo, W., Zhong, Q. Y., Wang, Y. Z.: A Method for Emergency Case Information Extraction based on Knowledge Element. System Engineering, Vol. 32, No. 12, pp. 133-139(2014).
10. Zhao, H., Zhao, T. J., Yu, H., Zheng, D. Q.: English Topic Tracking Research based on Query Vector. Journal of Computer Research and Development, Vol. 44, No. 8, pp. 1412-1417(2007).
11. Gu, Y. W., Wang, Y. R., Huan, J., Sun, Y. Q., Xu, S. K.: An Improved TFIDF Algorithm based on Dual Parallel Adaptive Computing Model. International Journal of Embedded Systems, Vol.13, No.1,pp.18-27(2020).
12. Li, X. O., Yu, W.: Imbalanced data classification via support vector machines and genetic algorithms. Connection Science, Vol 26, Issue 4, pp.335-348(2014).
13. Wang, F. F., Liu, Z., Wang, C.D.: An Improved kNN Text Classification Method. International Journal of Computational Science and Engineering, Vol.20, No.3, pp.397-403(2019).
14. Zhang, D. B., Shou, Y.F., Xu, J.M.: An Improved Parallel K-means Algorithm based on MapReduce. International Journal of Embedded Systems, Vol.9 No.3, pp.275-282(2017).