An Emergency Measure Completeness Evaluation Method Based on UGC Data

Hua Zhao, ShuJia Li, PeiXin Zhang, and QingTian Zeng

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

Correspondence should be addressed to Hua Zhao; huamolin@163.com

Received 9 May 2021; Revised 29 June 2021; Accepted 15 July 2021; Published 9 August 2021

Academic Editor: Yugen Yi

Copyright © 2021 Hua Zhao et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Complete emergency measures are one of the important guarantees for effective emergency response. An evaluation method based on User-Generated Content (UGC) data is proposed to assess the completeness of emergency measures. To implement this method, UGC data comprising emergency needs from the users and emergency plans comprising emergency measures for specific emergency were firstly collected using a crawler program. A feature template was then established to identify relationships between different features, and a Conditional Random Field (CRF) model was used to extract emergency measures from the emergency plan and emergency needs from UGC data. The Siamese network was applied thereafter to compute the similarity between the emergency measures and emergency needs. The missing emergency measures were obtained based on the similarities, and a quantitative method to calculate the completeness was defined. Finally, using province A as a case study, the emergency measure completeness was evaluated and the emergency measures that need to be strengthened and improved were identified.

1. Introduction

Emergency refers to unforeseen circumstances like natural disasters, accidents, public health incidents, and social security incidents that occur suddenly, causing or possibly causing serious social harm, and such require emergency measures to be taken in order to deal with them [1]. The COVID-19 pandemic is an example of the major public health emergency with a fast transmission rate, complex and multiple variants of the virus, wide infection range, and great difficulty in prevention and control [2]. To minimize the loss of people’s lives and property and to maintain social stability in the disaster areas, it is of great importance to ensure the efficient emergency measures after the occurrence of COVID-19 [3].

An emergency plan is a set of emergency measures made in advance to deal with possible emergencies, and it should include emergency measures as complete as possible [4]. The completeness of emergency measures means that the emergency plan should include all the emergency measures that need to be carried out after the occurrence of an emergency, so as to meet all the emergency needs of the public [5]. Therefore, the completeness of emergency measures must be evaluated to ensure that emergency plans can achieve ideal emergency disposal effects. Incomplete emergency measures will not only be difficult to meet practical needs but also lead to many emergency dilemmas [6]. For example, in the infant stages of the COVID-19 pandemic, there were a multitude of problems during the pandemic prevention and control such as imperfect emergency plans and insufficient analysis of emergency needs among other problems [7].

There were some methods to evaluate emergency measure completeness, but these methods have some specific disadvantages such as high cost. With the development of Web2.0 technology, various social platforms have become the first choice for people to express their personal opinions and suggestions. Thus, when an emergency occurs, people may also choose these platforms to publish their personal emergency needs and these needs reflect the completeness of emergency measures to a certain extent [8]. Therefore, this paper proposes evaluating the completeness of emergency measures in the emergency plan based on UGC data.

The contributions of this paper are as follows:
(1) The paper proposes evaluating the completeness of emergency measures from the perspective of UGC data, which is more economical than traditional methods. At the same time, it defines a quantitative method for computing the completeness of emergency measures.

(2) A Chinese corpus is constructed to evaluate the completeness of emergency measures, which can be shared with the researchers.

(3) The Siamese network similarity calculation model is applied to match emergency needs and emergency measures, and the completeness of emergency measures is quantified. Not only is the completeness of emergency measures intuitively seen, but also the specific aspects that need to be strengthened and improved are found in this paper.

2. Related Work

A lot of recent research has been done on emergency plans. From the public emergency response in recent years, it is of great practical significance to evaluate the emergency measures of emergency plans. Previous studies have suggested that the evaluation can be conducted in terms of completeness (demand), emergency responsibility matrix (departmental responsibility), and operability (task) [9]. The completeness of emergency measures focuses on evaluating the corresponding relationship between emergency needs and emergency measures in the emergency process [10].

There are two main methods of emergency measure completeness evaluation: one is simulation or plan drills and the other is evaluation index systems. For the first method, researchers usually conducted various large-scale exercise experiments with the support of government agencies that provided a large amount of real and effective data for evaluating the effectiveness and completeness of emergency measures in the plan [11]. Curnin et al. [12] took the Indonesian Tsunami and Hurricane Katrina as examples to analyze the factors that may hinder the implementation of an emergency plan during the emergency response process. Kaji et al. [13] applied the medical drill tool to evaluate the emergency drills of several hospitals in Los Angeles and used statistical analysis of the drill results to determine whether the emergency measures of the hospitals were comprehensive. Anna et al. [14] took the agricultural drought in Guatemala as an example and discussed how emergency drills can help identify problems in an emergency plan. Although the operation and implementation of response measures can be observed during drills, not all emergency plans are suitable for drills, since the cost of drills is very high, which makes them impractical.

Other researchers mainly use simulation software or establish an evaluation index system for either one or a class of emergency plans and then combine the evaluation methods to build an evaluation model. An et al. [15] constructed an evaluation index system from five dimensions of legality, scientificity, comprehensiveness, cost, and operability of the plan and evaluated the emergency plan for production safety accidents based on the hesitating fuzzy set method. Sun [16] built a fuzzy evaluation model by the fuzzy comprehensive evaluation method to evaluate and analyze the overall quality of Xi’an subway fire emergency plan. Chen et al. [17] proposed an emergency simulation system of multiperson cooperative urban rail station based on virtual reality, which can improve the problem of insufficient emergency response. Although costs are saved by constructing an evaluation index system and simulation method, the data sources are relatively weak. In addition, these evaluation methods usually fail to point out the deficiencies of the emergency plan content and cannot provide specific and targeted suggestions for revising and improving the emergency plan.

The purpose of this paper is to evaluate the emergency measure completeness based on UGC data. This article thinks that UGC data provides more real emergency needs from people and best reflects the completeness of emergency measures. Therefore, emergency-related UGC data is collected first, and then matches are made with the emergency measures in the emergency plan to realize the evaluation of the emergency measures completeness and find the defects in the emergency plan.

3. Overall Framework

Traditional evaluation methods consume a lot of human and material resources, so this paper proposed a solution based on UGC data to evaluate the completeness of emergency measures. The evaluation process is decomposed into three steps, as shown in Figure 1.

As displayed in Figure 1, the framework proposed in this paper mainly includes the three following parts:

1. **Data Acquisition**. The data used in this paper includes two parts: one is user comment data (UGC data) obtained from microblogs and the other is emergency plans. For the comment data, a crawler program was designed to retrieve and grab relevant comments on Sina Weibo. For emergency plans, this paper collected the emergency plans for public health emergencies issued by some provinces. The final preprocessing mainly includes two steps: The first is to divide the comments into sentences and extract the sentence in the form of “emergency agency - emergency measures” from the emergency plan. The second is deduplication, Chinese word segmentation, dependency analysis, and other operations.

2. **Need and Measure Extraction**. This article regarded emergency measures or emergency needs as the opinion target in sentiment analysis research and used CRF model to extract them. Need and measure extraction is mainly divided into three steps: the first is to select training features according to the characteristics of emergency measures, followed by designing the feature template and, finally, using the CRF++ toolkit for model training in order to obtain the emergency needs in UGC data and the emergency measures in the emergency plan.
Matching. The main task here is to match the extracted emergency needs and emergency measures using the text-similarity calculation model and find out the missing emergency measures. The paper defined a comprehensive score of emergency measures according to the matching degree as an evaluation index for the completeness of emergency measures, as shown in the two following equations:

\[ C_{\text{incompleteness}} = \frac{n}{(m + n)} \times 100\% \]  \hspace{1cm} (1)

\[ C_{\text{completeness}} = 1 - C_{\text{incompleteness}} \]  \hspace{1cm} (2)

where \( C_{\text{completeness}} \) is the completeness rate of emergency measures; \( C_{\text{incompleteness}} \) represents the incompleteness rate of emergency measures, where the lower the value, the better the completeness of emergency measures; \( n \) is the number of missing emergency measures through similarity calculation; and \( m \) is the number of emergency measures provided in the emergency plan.

4. Extraction Method of Emergency Measures and Emergency Needs Based on CRF

As shown in Figure 1, this paper needs to extract emergency measures from the emergency plan and emergency needs from UGC data and then calculate the similarity between them to get the score of the completeness of emergency measures. This article regards emergency measures and emergency needs as opinion targets and realizes their extraction through the opinion targets extraction method.

Currently, CRF achieves good performances in Chinese word segmentation, part-of-speech tagging, named entity recognition, and other natural language processing tasks, especially in opinion target extraction [18]. This paper uses the CRF model to identify emergency needs and emergency measures and converts them into a maximum probability sequence solution problem. The extraction method mainly includes the following work.

Syntax structure analysis can summarize the rules and make a rule template to extract specific information [19]. Therefore, the paper used the dependency parsing interface of LTP (Language Technology Platform) to analyze the dependency relationship between the components in a sentence and to reveal its syntactic structure, such as the phrase structure “subject-predicate-object” and “attribute-adverbial-complement.” The core verb in a sentence is the central component that dominates other components, but it is uncontrolled by any other component [20]. The example and its notes are shown in Figure 2.

It can be seen from the analysis results that the subject of the sentence is “the government,” the core predicate is “regulate,” and the object is “market prices.” It can be concluded that emergency agencies generally exist in the subject of the sentence, while the emergency measures are generally performed by the core predicate and its object. That is, from “HED” to the end of the sentence is used as the task description text of the department. Therefore, the emergency measures or emergency needs in the text can be extracted by selecting the features such as dependency relationships. Table 1 lists all the features used in the experiment.

By analyzing the characteristics of the text, this paper chose the word-based BIEO tagging system to complete the feature tagging work. After the manual feature labeling, this article designed a feature template to establish the relationship between each feature and obtain the optimal model of sequence marking. Finally, CRF++ is trained and tested based on the collected data.
5. Matching Method Based on the Siamese Network

5.1. Task Description. The matching task is to compare the extracted emergency needs and emergency measures, which will have three possible results: in both the plan and comments, only in the plan, and only in the comments, as shown in Figure 3. The first is that emergency measures already contain emergency needs; the second indicates that emergency needs and emergency measures are independent of each other; the last means that the existing emergency measures cannot meet the emergency needs. This paper mainly focuses on the third result, which judges the completeness of emergency measures based on the matching results.

5.2. Similarity Computation. With the rapid development of deep learning technology, especially after the emergence of Word2Vec in 2013, the calculation of semantic text similarity has made breakthrough progress, and, therefore, the calculation method of semantic text similarity based on deep learning has gradually become the mainstream method in this field [21]. Currently, the classic similarity calculation method in deep learning technology is based on the framework of the Siamese network. Figure 4 shows the general process of the whole network [22].

One of the main characteristics of the Siamese network is the shared weights across subnetworks, which not only reduce the number of parameters but also reduce the tendency of overfitting. First of all, two sentences are input on the left and right, respectively, and the representation vector of the sentence to be tested through two identical embedding layers is obtained. After each sentence has been processed by an LSTM (Long Short-Term Memory) network, the distance between the two output vectors is then calculated by some distance measurement, as shown in equation (3). The distance calculation methods commonly used in text similarity calculation include Euclidean distance, Cosine distance, Manhattan distance, and Chebyshev distance [23]. In general, the smaller the distance is, the more similar the two texts are. The goal of training is to minimize the distance between similar sentence pairs in the embedding space and maximize the distance between different sentence pairs. Finally, the similarity score of the two texts to be matched is obtained.

\[
y = \exp \left( -\| h_T^{(a)} - h_T^{(b)} \| \right) \in [0, 1],
\]

where \( h_T \) is the last hidden state of the model, representing the final representation of the sentence. Thus, \( h_T^{(a)} \) is the representation of Sentence 1 and \( h_T^{(b)} \) is the representation of Sentence 2.

As shown in Figure 5, although the LSTM sequentially updates a hidden state representation, these steps also depend on the memory unit containing four components (real-valued vectors): the memory state \( c_t \), the output gate \( o_t \) that determines how the memory state affects other units, and an input (and forgetting) gate \( i_t(f_t) \) that controls the memory content according to each new input and current state, as shown in Figure 5. The following is an update of LSTM parameterized by weight matrices \( W_f, W_o, W_r, U_f, U_o, U_r \), and \( U_o \) and bias vectors \( b_f, b_o, b_r \) and \( b_o \) in each \( t \in \{1, \ldots, T\} \):

\[
i_t = \sigma(W_c x_t + U_f h_{t-1} + b_f),
\]

\[
f_t = \sigma(W_f x_t + U_f h_{t-1} + b_f),
\]

\[
\tilde{c}_t = \tanh(W_r x_t + U_r h_{t-1} + b_r),
\]

\[
c_t = i_t \cdot \tilde{c}_t + f_t \cdot c_{t-1},
\]

\[
o_t = \sigma(W_o x_t + U_o h_{t-1} + b_o),
\]

\[
h_t = o_t \cdot \tanh(c_t),
\]

\[
\sigma(z) = \frac{1}{1 + e^{-z}}.
\]
Figure 3: Matching results.

Figure 4: The Siamese network framework.

Figure 5: LSTM network diagram.
\[
\tanh(z) = \frac{e^z - e^{-z}}{e^z + e^{-z}}.
\] (11)

The parameter update formula of the gradient descent method is as follows:
\[
\theta_{\text{cur}} = \theta_{\text{pre}} - \eta \nabla L(\theta),
\] (12)

where \( \eta \) is the learning rate that determines the size of the moving pace; \( \theta \) refers to a bunch of parameters, such as those mentioned above; \( L \) is the loss function.

Specifically, each sentence (represented as a sequence of word vectors) \( x_1, \ldots, x_T \) is passed to LSTM, and LSTM updates its hidden state at each sequence index through equations (4)–(11) and uses its final hidden state as a vector of each sentence. Then the similarity between these vector representations is used as a predictor of semantic similarity, and, finally, the similarity score is obtained. If the score is greater than a certain threshold, they are considered to be similar; otherwise, they are not.

6. Experiments

6.1. Experiment of Emergency Measure and Emergency Need Extraction

6.1.1. Datasets. There were two kinds of data used in this experiment: the first was the user comment text about COVID-19 prevention and control, and the second was the public health emergency plans of some provinces. The details of the data are shown in Table 2.

In the experiment, LTP was selected to carry out language processing tasks such as word segmentation, part-of-speech tagging and dependency parsing on the text, and transforming the text into the data format required by CRF model. Table 3 gives a labeling instance.

6.1.2. Experimental Settings. In this experiment, CRF++ 0.53 was used to implement the Conditional Random Field model, in which the model parameter values, such as -c-f, were set according to manual experience. The window size of words and parts of speech was 3, and the current window was used for other features. In order to establish the internal relationship between the above four features (word, PoS, head, and relationship), 18 different feature templates were designed by different combinations of the four features.

6.1.3. Experimental Results. For the results of opinion target extraction, the experiment adopted accurate evaluation. That is, the extraction result is correct only when the extracted opinion target and the annotation answer match exactly. Precision, recall, and \( F_1 \) value were used as the evaluation criteria. The calculation formula is shown in formulas (13)–(15). \( TP \) (true positive samples) represents the number of labels correctly identified in the test set; \( FP \) (false positive samples) represents the number of misidentified labels; and \( FN \) (false negative samples) represents the number of labels that are not identified in the test set.

\[
\text{Precision} = \frac{TP}{TP + FP} \times 100\%,
\] (13)

\[
\text{recall} = \frac{TP}{TP + FN} \times 100\%,
\] (14)

\[
F_1 = \frac{2 \times P \times R}{P + R} \times 100\%.
\] (15)

\( \text{Precision} \) indicates the proportion of predicted positive cases that are actual positives. \( \text{Recall} \) is the proportion of actual positive cases that were correctly predicted. \( F_1 \) is the harmonic mean of \( \text{precision} \) and \( \text{recall} \).

The experimental results are shown in Table 4. It can be seen that the effect of the model is significantly improved by adding dependency features, which can effectively extract the emergency needs and emergency measures.

6.2. Experiments of Matching

6.2.1. Experimental Data and Environment. The experiment realized the text similarity calculation based on the Siamese network. Since there is no publicly available corpus, this paper manually sorted and marked 5165 pairs of sentences about emergency response extracted by the CRF model. The labeled samples used in training the Siamese network are sentence pairs (sentence1, sentence2) and their label is similar_score(0, 1). similar_score = 0 means that the two sentences are not similar; otherwise they are similar. Python 3 was used in the experiment, and Keras, a deep learning framework based on TensorFlow, was selected for the construction of the Siamese network.

6.2.2. Experimental Design

(1) Word Embedding Learning. In the word embedding learning module, the paper used a 300-dimensional Word2Vec model pretrained on an external corpus and then mapped each word to a fixed-size vector. More than 30,000 pieces of data used in word vector training are from Chinese Wikipedia, and the training was performed after pre-processing steps such as word segmentation. The basic training settings are shown in Table 5. When an unregistered word appears, it was initialized to a 300-dimensional 0 vector.

(2) Parameter Setting. Adadelta optimizer [24] was selected as the optimization algorithm, and the formula is shown as follows:

\[
\Delta \theta_t = \frac{\text{RMS}[\Delta \theta_{t-1}]}{\text{RMS}[g_t]} g_t,
\] (16)

\[
\theta_{t+1} = \theta_t + \Delta \theta_t,
\]

where \( \Delta \theta \) is the parameter update vector; \( g_t \) is the gradient of the parameters at the \( t \)-th iteration; and \( \text{RMS}[g_t] \) is the root mean square of the gradient at time \( t \).
In addition, gradient clipping was carried out by a threshold value of 1.25. EarlyStopping was added to each layer to prevent overfitting. To solve the problem of sample category imbalance, class_weight was used to map the class index to the weight value, which was used to weight the loss function. The number of hidden nodes of LSTM was set to 50. The batch parameter size was set to 64 and the number of training rounds was set to 20. Through the selection method of the similarity threshold shown above, the similarity threshold was 0.5. The loss function is MSE (Mean Square Error), and the formula is as follows:

$$\text{MSE}(y, y') = \frac{1}{n} \sum_{i=1}^{n} (y_i - y'_i)^2.$$  

(17)

6.2.3. Experimental Result. To evaluate the performance of different distance measurement methods on the task similarity prediction problem, this paper used the Matplotlib library to plot the accuracy of the model, as shown in Figures 6 and 7.

The final experiment adopted Manhattan distance, because it can be seen from Figures 6 and 7 that the cosine has a poor effect, while the Manhattan measurement effect is better. The reason may be that Cosine only calculates the angle between the two vectors, but Manhattan distance also saves the length information of the two vectors.

The Manhattan distance between two n-dimensional vectors \(a(x_{11}, x_{12}, \ldots, x_{1n})\) and \(b(x_{21}, x_{22}, \ldots, x_{2n})\) is

$$d(a, b) = \sum_{i=1}^{n} |x_{1i} - x_{2i}|.$$
The calculation formula is shown in formulas (19) and (20), where 

\[ R = \frac{\sum_{i=1}^{n} (y_i - \bar{y})(\hat{y}_i - \bar{y})}{\sqrt{\sum_{i=1}^{n} (y_i - \bar{y})^2 \sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2}} \]  

\[ P = 1 - \frac{6 \sum_{i=1}^{n} d_{i}^2}{n(n^2 - 1)} \]  

where \( R \) is used to measure the correlation between the predicted label and the true label and \( P \) is used to measure the correlation between rank variables.

The comparison of the results of each model is shown in Table 6. By comparing it to other models, it can be observed that the Siamese network used in this paper is better than the other three models in the evaluation index.

### 7. Case Study

In order to more intuitively verify the feasibility of the evaluation method for the emergency measures completeness proposed in this paper, the emergency plan for public health emergencies in province A was taken as an example to evaluate the completeness of emergency measures according to the above process method.

First of all, 24 emergency needs and 69 emergency measures are obtained through the CRF model. Then, the similarity calculation model was used to match them and obtain the similarity score of each sentence pair. For example, the matching score of "ensuring price stability in closed managed areas" and "ensuring price stability in market supply in closed areas" was 0.97. The matching score of "set up report telephone to strengthen prevention and control" and "set up temporary traffic health quarantine at the entry and exit ports of traffic stations" was 0.101. These matching scores are more consistent with manual judgment and are more objective to judge the completeness of emergency measures.

Therefore, emergency needs were matched, one by one, with all emergency measures in the plan. If all the similarity scores are less than 0.5, it can be determined that this emergency need is the missing emergency measure in the plan. Thus, this paper found that there are six emergency measures missing, as shown in Table 7.

Finally, according to formula (1), the incompleteness of emergency measures was calculated as follows:

\[ C_{\text{incompleteness}} = \frac{n}{(m + n)} \times 100\% = \frac{6}{(6 + 69)} \times 100\% = 8\%. \]

Therefore, the incompleteness rate of the emergency measures of the plan is 8%; that is, the completeness reaches 92%.

### 8. Conclusion

A UGC-based method to evaluate the emergency measure completeness has been proposed in this paper. The method begins with collecting emergency plans and user comments about public health emergencies, followed by extracting emergency needs and emergency measures using a CRF model. The CRF model is trained to extract emergency needs and emergency measures by selecting features based on designed feature templates. A similarity calculation model was then established to calculate the similarity between the emergency needs and the emergency measures. Finally, the emergency measure completeness or incompleteness of the emergency plan was quantified according to a defined
formula. In order to verify the feasibility of the proposed evaluation method, province A was used as a case and the experimental results prove the effectiveness of the proposed evaluation method.

This article regards the emergency measures taken by various departments as a set. For the emergency needs in the comments, the emergency measures that are missing in the set are found, instead of studying the emergency situations of specific departments; that is, the set of emergency department and the set of emergency measures are not corresponding to each other, so that it is impossible to locate the emergency measures that are lacking in specific departments. In the future, further studies can be carried out on specific departments to improve the responsibilities of each department. It can also conduct emotional orientation research on emergency measures existing in the plans and comments based on UGC data, so as to provide support for improving the effectiveness of emergency measures.

Data Availability

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

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (Grant nos. 61602278 and 61702306); Sci.&Tech. Development Fund of Shandong Province of China (ZR2017MF027); the Social Science and Humanity on Young Fund of the Ministry of Education (nos. 16YJCZH154, 16YJCZH012, and 18YJAZH017); the SDUST Research Fund (no. 2015TDJH102); Open Project Fund of Key Lab of Network Data Science and Technology (no. CASNDST201706); SDUST Cultural Construction Project (no. WHJSZ-1705); and Qingdao Philosophy and Social Science Planning Research Project (no. QDSKL1901124).

References

[1] Y. Y. Long, Research on Government Guidance Mechanism of Internet Public Opinion in Public Emergencies, University of Electronic Science and Technology, Chengdu, China, 2016.
[2] L. Y. Sun and Y. H. Sun, “Epidemic prevention and control highlights the institutional advantages of China’s emergency management,” Journal of Tianjin University (Social Science Edition), vol. 22, no. 3, pp. 200–205, 2020.
[3] S.-P. Wan, W.-B. Huang Cheng, and J.-Y. Dong, “Interactive multi-criteria group decision-making with probabilistic linguistic information for emergency assistance of COVID-19,” Applied Soft Computing, vol. 107, Article ID 107383, 2021.
[4] G. C. E. Kafoutis and I. M. Dokas, “MIND-VERSA: a new methodology for identifying and determining loopholes and the completeness value of emergency response plans,” Safety Science, vol. 136, pp. 1–13, 2021.
[5] S. Q. Wang, Z. M. Zhang, and P. J. Zheng, “Completeness evaluation of emergency plan for ship pollution accident based on Bayesian network,” Journal of Shandong Administration (Science and Technology Edition), vol. 33, no. 4, pp. 97–102, 2020.
[6] Y. J. Sun, “The implications of COVID-19 epidemic prevention and control,” Journal of Shandong Administration University, vol. 172, no. 3, pp. 48–51, 2020.
[7] Z. D. Liu, H. W. Gao, Y. Q. Wang, and Z. B. Jing, “Thoughts on China public health emergencies in the background of coronavirus disease 2019,” Journal of Central University of Finance and Economics, no. 94, pp. 109–115, 2020.
[8] X. F. Cai, Research on Public Participation in Internet Public Opinion in Emergencies, South China Agricultural University, Guangzhou, China, 2018.
[9] P. J. Zhang, J. Chen, and J. F. Liu, “Research on the evaluation method of natural disaster emergency plan in my country (III): operability evaluation,” China Safety Science Journal, vol. 18, no. 10, pp. 16–25, 2008.
[10] J. F. Liu, P. J. Zhang, Z. F. Chen, and J. Chen, “Research on the evaluation method of natural disaster emergency plan in my country (I): completeness evaluation,” China Safety Science Journal, vol. 18, no. 2, pp. 5–11, 2008.
[11] Y. H. Song, Y. Zeng, W. T. Chen, Y. Liu, M. Y. Xu, and W. Lv, “Study on the effectiveness evaluation of emergency plan considering completeness and operability,” Journal of Wuhan University of Technology (Information and Management Engineering Edition), vol. 42, no. 5, pp. 388–394, 2020.
[12] S. Curnin, C. Owen, D. Paton, and B. Brooks, “A theoretical framework for negotiating the path of emergency management multi-agency coordination,” Applied Ergonomics, vol. 47, pp. 300–307, 2015.
[13] A. H. Kaji and R. J. Lewis, “Assessment of the reliability of the johns hopkins/agency for healthcare research and quality hospital disaster drill evaluation tool,” Annals of Emergency Medicine, vol. 52, no. 3, pp. 204–210, 2008.
[14] A. Müller, V. Mora, E. Rojas et al., “Emergency drills for agricultural drought response: a case study in Guatemala,” Disasters, vol. 43, no. 2, pp. 410–430, 2019.
[15] J. W. An, X. An, L. K. Wang, and L. Li, “Evaluation of emergency plans for production safety accidents based on hesitating fuzzy sets,” Journal of Safety Science and Technology, vol. 13, no. 5, pp. 128–133, 2017.
[16] Y. B. Sun, Research on Quality Evaluation and Optimization of Emergency Plan for Xi’an Subway Fire Accident, Northwest University, Kirkland, WA, USA, 2019.
[17] T. Jiang, Y. Qiu, and F. Li, “A multi-person collaborative simulation system for subway emergency based on virtual reality,” IOP Conference Series: Materials Science and Engineering, vol. 750, no. 1, Article ID 012227, 2020.
[18] S. Liu, T. H. He, and J. H. Dai, “A survey of CRF algorithm based knowledge extraction of elementary mathematics in Chinese,” Mobile Networks and Applications, vol. 15, pp. 1–13, 2015.
[19] Y. F. Qiu and S. Y. Zhou, “Extraction of evaluation objects based on syntactic structure and dependence,” Application Research of Computers, vol. 37, no. 4, pp. 1054–1058, 2020.
[20] Y. Xia, Evaluation Object Extraction Based on Conditional Random Field, Huazhong University of Science and Technology, Wuhan, China, 2016.
[22] A. Thyagarajan, “Siamese recurrent architectures for learning sentence similarity,” in Proceedings of the Thirtieth Aaai Conference on Artificial Intelligence, pp. 2786–2792, Phoenix, AZ, USA, February 2016.

[23] N. Othman, R. Faiz, and K. Smali, “Manhattan siamese LSTM for question retrieval in community question answering,” in Proceedings of the the 18th International Conference on. Ontologies, DataBases, and Applications of Semantics, pp. 661–677, Rhodes, Greece, October 2019.

[24] M. Zeiler, “ADADELTA: an adaptive learning rate method,” 2012, https://arxiv.org/abs/1212.5701.