Geoinformatica
https://doi.org/10.1007/s10707-018-0319-4

Efficient CPS model based online opinion governance modeling and evaluation for emergency accidents

Xiao Long Deng[1] **· Yin Luan Yu**[2] **· Dan Huai Guo**[3] **·**
Ying Tong Dou[1]

Received: 25 January 2017 / Revised: 5 January 2018 / Accepted: 19 March 2018

Abstract In the last decades, there have been much more public crisis accidents in the world such as H1N1, H7N9 and Ebola outbreak. It has been proved that our world has come into the time while public crisis accidents number was growing fast. Furthermore, crisis response to these public emergency accidents is always involved in a complex system consisting of cyber, physics and society domains (CPS Model). In order to collect and analyze these emergency accidents with higher efficiency, we need to design and adopt some new tools and models to analysis the online opinion. In this paper, we have proposed a new CPS Model based Online Opinion Governance system which constructed on cellphone APP for data collection including GIS information and online opinion and decision making in the back end. Our contributions include the graded risk classification method and accident classification method. Besides, we propose the group opinion polarization analysis method consisting two models and make promotion of the relative conditional entropy based context key word extraction method. Basing on these, we have built an efficient CPS Model based simulated emergency accident replying and handling system. It has been proved useful for emergency response in some real accidents in China such as Tianjin Explode accident and Haiyan Typhoon in recent years with detailed and vivid analysis result.

Keywords Mobile data · CPS model · Online opinion analysis · Emergency disaster · Situation analysis and evaluation

✉ Xiao Long Deng
shannondeng@bupt.edu.cn

✉ Yin Luan Yu
yinluanyu@bnu.edu.cn

1 Key Lab of Trustworthy Distributed Computing and Service of Education Ministry, Beijing University of Posts and Telecommunications, Beijing, 100876, China

2 China Academy of Social Management, Beijing Normal University Beijing, Beijing, China

3 China Scientific Data Center, Computer Network Information Center, Chinese Academy of Sciences Beijing, Beijing, China

Published online: 12 April 2018

1 Introduction

There has been much more public crisis and accidents around the world in the past decade (i.e. H1N1, Ebola and Typhoon Haiyan). Emergency accidents with severe effect like Explosion of Tianjin Port have lead a huge fluctuation of online opinion in China. And it has made the evolution of online opinion about emergency accidents becomes a vital factor in crisis management for our government [1, 2].

In this article, in order to get the more comprehensive information from all round side in Cyber Domain, Social Domain and Physical Domain, we have promoted a CPS model based online opinion governance system and evaluation methods of emergency accidents. At first, our system would collect the online opinion from three-dimension space (cyber, physical and society) which was proposed by professor Zhang in 2014 [3–5]. In the physical space, we collect some physical information like the data about damages and deaths. For the cyber space, with the aid of distributed data mining system, we collect and analyze online opinion from Sina Micro blog, Baidu News, Tianya BBS and Twitter. In the society space, we analyze the trend of online opinion from social network. According to the information collected from several spaces, the calculation and analysis will be applied for emergency handling. At last, the decision will be made based on different situations to deal with various kinds of emergencies according to our simulation computing. In detail, our contributions can be found in the following:

1. We propose and implement a CPS Model based Online Opinion Governance system which constructed on cellphone APP for data collection and decision make in the back end and it has some very good performance;
2. In addition,we propose our UNPE and DNPE model to analyze Group Opinion Polarization Process and has evaluated it on real data of Tianjin Explosion Accident;
3. We propose a detailed risk classification result with five layered risk. Furthermore, we propose the accident types classification method. They are useful for emergency accident risk assessment;
4. We put forward and supply a detailed analysis of Tianjin Explosion in 2015 and Typhoon Haiyan in 2013 which combing the cyber information, physical information and society information together and it can provide vivid demonstration of the two accidents;

The outline of this paper is as follows: In Section 1, the research background is introduced. In Section 2, the related work on CPS Model and online opinion ming for emergency handling will be described. In Section 3, we offer detailed information of CPS model based analysis system and our APP based system framework. In Section 4, more detailed demonstration of key technology in our CPS model analysis system can be found. Furthermore, the detailed risk classification result and accident types classification will be found in Section 5. Then, we offered the opinion mining result of 2015 Tianjin Explosion Accident and 2013 Typhoon Haiyan in Section 6. And Section 7 gives a conclusion.

2 Related work

2.1 CPS model

The CPS model [4] has three main steps. The first step is the collection of online opinion. The second step is the calculation and analysis of collected data and the last step is decision making and feedback. Here we will introduce the details of CPS model (Fig. 1).

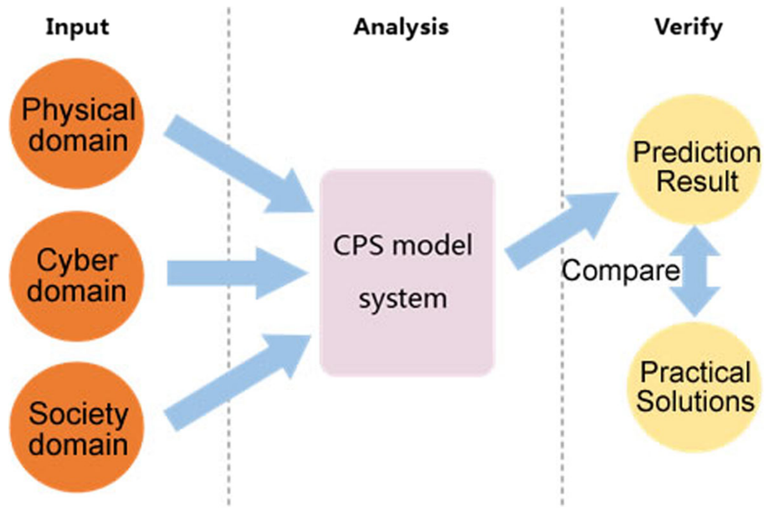


Fig. 1 CPS model based risk reply

The basic way to evaluate the model is analyzing the event data that have been input to the model. The conclusion of event will be compared with the actual process and verify whether they are similar or not.

Crisis response is involved in a complex system consisting of cyber, physics and society domains. Different domains are closely coupled and dynamically evolved. Cross domain analysis is applied for better response of natural disasters and other emergency accidents. A data-model integration method is proposed by us to support situation awareness and emergency response planning. Models are used for the prediction of crisis scales and its impact on infrastructure system and local society. Data mining is used to sense sentiment of affected citizens in society. The integration of cross-domain data mining and emergency modeling is essential for supporting scientific decision-making (including planning and response) [3–5]. But the current research results of CPS Model is still in the theoretical level and not been applied to the real accidents.

2.2 Online opinion ming for emergency handling

Online opinion ming is useful for emergency accident reaction in emergency accidents especially in disasters. There are some famous software and developing teams concentrated on this such as SAFRR (Science Application for Risk Reduction) (https://www2.usgs.gov/natural_hazards/safrr/contact.asp). The SAFRR Project is the continued evolution of the successful Multi-Hazards Demonstration Project (MHDP), started in 2006 and lasted for only five years. The SAFRR team helps to build resilience to natural hazards such as earthquakes, floods, wildfires, landslides, tsunamis, and coastal erosion by working with decision making and emergency response efforts across different nations including online opinion data handling. But the SAFRR team has not apply it to China and other developing countries. The another one is Open Source Disaster Management Software named Sahana (https://sahanafoundation.org/) to solve concrete problem and bring efficiency to disaster response coordination between government, civil society and the victims themselves by using mobile phone APP. Sahana's open-source codebase has provided a flexible, modular platform for rapidly deploying information management systems for disaster management and humanitarian use cases. In 2008, IBM has used Sahana to save lives in the WenChuan earthquake. The WHO group(World Health Organization) (http://www.who.int/entity/en/) also has played an important role in emergency handling including many projects to help people to survive in disasters by mobile phone APPs. But the related work mentioned above has no Chinese Edition and has no efficient opinion data processing tool in the back end and we need to develop some useful Chinese APPs and softwares to help Chinese people to save their lives in emergency accidents.

2.3 Summary

For the reason that CPS Model is very new in the decade and Online opinion Ming for emergency Handling is a crossing interdisciplinary researching area, there is not much related work of it and the common practice is to apply the traditional text ming method to analysis the emergency online opinion and short tweets. So it very necessary for us to propose a new integrated extracting and analysis framework for emergency online information handling.

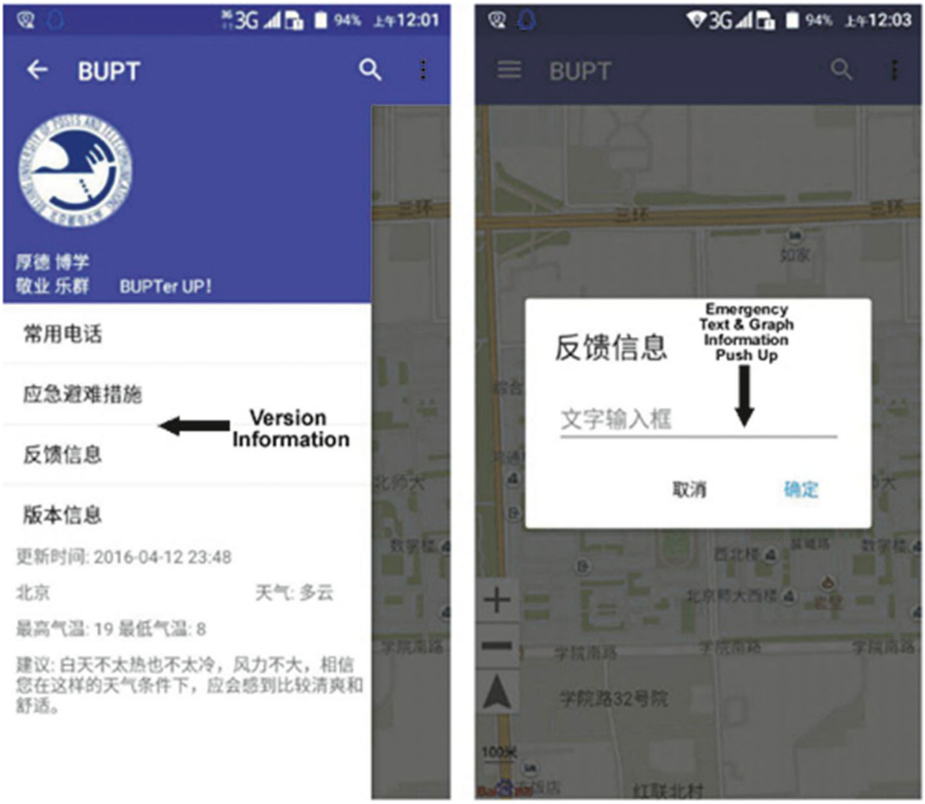


Fig. 2 APP GUI

3 CPS model based online opinion governance structure

3.1 Opinion collection APP

To collect the online opinion, we combine mobile APPs and web crawlers together. Our mobile APP collects the information of events happened in the real world (i.e. photos, locations, time and so on). The web crawler gets the online opinion in the web. The APP-based information collection is a user-driven information pushing mechanism (i.e. the client-server structure). Users send information (including longitude and latitude information) of emergency accidents via our specific APP like which can be found in Fig. 2 to the server where can store and analyze the accident data.

3.2 System framework

It can be found from Fig. 2 that user can send out the accident information in emergency and this is a demo GUI for emergency report for our university campus (www.bupt.edu.cn) for simple demonstration and can be used in many other areas to send information by users. In our system, user can send out text information, photo information and location information (including latitude and longitude information).After the information was sent to back end server, it can be extracted and helpful for further data ming in Fig. 3.

The web crawler is a server-based online information collection process. The web crawler can runs 7*24 hours on the server to collect the online information about the emergencies from BBS, micro-blogging and others websites. Besides, the crawler can also collect information from news website and other online libraries according to the specific

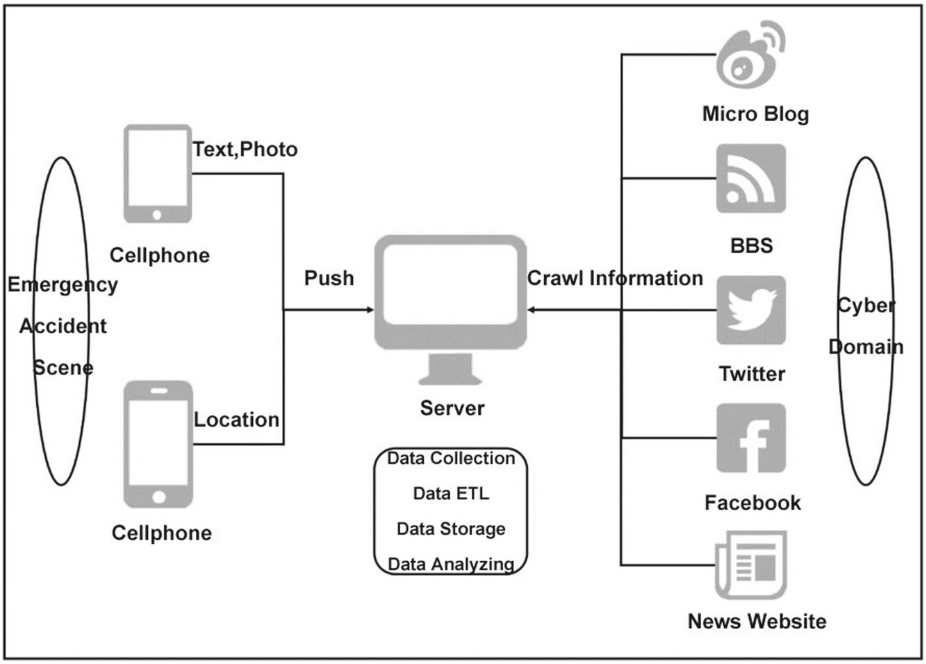


Fig. 3 Opinion collection system framework

emergency accident such as Sina Micro Blog, BBS, Twitter, and Facebook. As a result, we can offer some suggestions to the government to handle opinion situation in these emergency accidents [6–9].

4 Key technology in our CPS model analysis system

4.1 Relative conditional entropy based context key word extraction

The word frequency method is the simplest method for extracting the key words, taking the frequency scale of key words as a criterion to choose the feature words while considering that a word with increasing frequency of occurrence in text is more representative. In our System, mutual information entropy based method and relative conditional entropy based method are implemented to extract the main key words of some special emergency accidents.

1) Mutual information entropy method: The mutual information entropy based method is to measure the relevance between the feature item and a particular category at the time of extracting feature words. If the calculated value of mutual information in formula (1) is greater, the relevance between the corresponding feature item and the particular category is higher.

$$\begin{aligned} I(X;Y) &= H(X) + H(Y) - H(X,Y) \\ &= H(X) + H(Y) + \sum p(x,y)\, log_2\,(p(x,y)) \\ &= \sum_{x\in X}\sum_{y\in Y} p(x,y) log_2 (\tfrac{p(x,y)}{p(x)p(y)}) \end{aligned} \tag{1}$$

In formula (1), $p(x, y)$ stands for the joint distribution of two random variables (X, Y) and the marginal distribution is $p(x)\,p(y)$. The mutual information $I(X; Y)$ is the relative entropy of the joint distribution $p(x, y)$ with the marginal distribution $p(x)\,p(y)$. $H(X, Y)$ is the joint entropy. The mutual information entropy based method is mainly used to calculate the relativity of analyzed short text sent by online social network user such as Facebook and our Opinion Collection APP, with some topic recognized articles of emergency accidents. And the relativity can be shown by the calculated result of information quantity. In case of hot discussion on some accidents, the information quantity would reach climax peak to show the intensive attention of social network on the accident.

2) Relative conditional entropy based context key word extraction method: In information theory, the entropy is to measure the expected value of a random variable. The entropy value H of a random variable X in range of $\{x_1, ...x_n\}$ is defined as $H(x) = E(I(X))$. And $I(X)$ is the self-information of random variable X. At the same time, another expression form of entropy H can be defined in formula (2) according to the definition of expectation and self-information $I(X)$:

$$H(X) = \sum_{i=1}^{n} p(x_i)I(x_i) = -\sum_{i=1}^{n} p(x_i) logp(x_i) \tag{2}$$

The main idea of Relative Conditional Entropy based Context Key Word Extraction method is to extract several key words or phrases which can stands for the main topic of the analyzed article. The result of this method can be used to refined reading, semantic query and fast topic match.

The main process of this method can be described in the followings:

Firstly, based on some scale of corpus, several Focus Words [10] would be selected which can cover the most scope of the target article for key words extraction. Then the words on the m position before and behind each Focus Word would be selected to compose the "Focus Word Context Matrix". W stands for the Focus Words Set and CW is the Words Set of Context Words. The "Focus Word Context Matrix" can be defined as:

$$\begin{aligned} M_w = [w_{i,j}]_{n\times(2\times m+1)} i = 1, 2, ..., n;\ j = 1, 2, ..., (2 \times m + 1); \\ w_{i,(m+1)} \in W, w_{i,j} \in CW, j \neq m + 1. \end{aligned} \tag{3}$$

In formula (3), $n = \sum_{w\in W} fre(w)$ stands for the word frequency of all Focus Word $w \in W$ in the corpus library. $w_{i,1:m}$ are the nearest m words before Focus Word w in the i_{th} context paragraph and $w_{i,m+1:2\times m+1}$are nearest $m$words behind Focus Word w in the i_{th} context paragraph.

Secondly, the Matrix M_w of all Focus Words Set would be transmitted into a system of symbols. Corpus Focus Words set $w \in W$ and the context word distance to Focus Word w will be formalized separately formalized to the information source and information sink of information system. And then, the provided information gain by each position of different context words will be calculated precisely to the whole formalized information system. The information gain will be calculated as follows:

$$IG_p = H(W) - H(W|V_p) \tag{4}$$

In formula (4), IG_p is the information quantity of context position p and IG_p stands for the decrement of information entropy $H(W)$ from whole system to $H(W|V_p)$ which stands for the relative conditional entropy.

$H(W)$ can be calculated as the information entropy of the Corpus Focus Words which is defined as the information source in formula (5):

$$H(W) = -\sum_{\omega\in W} p(\omega) \times \log_2 p(\omega) \tag{5}$$

$p(\omega)$ is the counted word frequency of Focus Word ω and $p(\omega)$ is defined as:

$$p(\omega) = \frac{fre(\omega)}{n} \tag{6}$$

And $H(W|V_p)$ in formula (4) is the relative conditional entropy of context position V_p which is defined as:

$$H(W|V_p) = \sum_{c\omega\in V_p} p(c\omega) \times H(W|c\omega) \tag{7}$$

In formula (7), $p(c\omega)$is the counted word frequency of context words $c\omega$ in context position p. $H(W|c\omega)$ is the conditional entropy where the distribution of context words $c\omega$ is attained and can be defined as:

$$H(W|c\omega) = -\sum_{\omega\in W} p(\omega|c\omega) \times \log_2 p(\omega|c\omega) \tag{8}$$

After calculation the values of parameters from formula (2) to (8), the information quantity gained of each word in the context position to the all Focus Word $w \in W$ can be found out. Besides the main idea of our relative conditional entropy based context key word extraction method, this algorithm can extract the representative words or phrases with high preciseness on behalf of the main soul of the whole article in fast speed. In handling short text of social network, it has been proved of good efficiency.

4.2 Group opinion polarization analysis in social network structure

The group opinion polarization process is always influenced by the user structure of social network. Among the well-known network models such as ER Random Network Model [11], Small World Network Model [12] and Scale-Free Network Model [13], the Scale-Free Network Model has been proved as the most suitable model for the structure of online social network [14] in 2010. And the Local World Evolving Network Model was proposed by Li and Chen in 2003 to simulate the local network evolving mechanism and it can be used to research on the group opinion polarization process of social network [15]. Besides these models, the Hopfield model is often cited as the classical model for group opinion polarization process [16–19] which rest assure that the changing process of single person would always be dominated by the opinion pressure of other people in the same group.

Basing on the related work of group opinion polarization mechanism and Hopfield model, we propose the UNPE Model (Undirected Network Polarization Evolution) and DNPE Model (Directed Network Polarization Evolution) for group opinion polarization process of online social network group.

In our experiment, the UNPE, DNPE model and group polarization model are integrated to study in a given social environment while how social networks lead to group polarization and how to control the network polarization phenomenon by controlling the various parameters of social network in case of emergency for public opinion governance.

The UNPE Model is constructed to simulate the network polarization phenomenon on undirected networks.(e.g. Renren.com and Facebook) while the DNPE Model focus on classical directed social network such as Sina Micro-Blog and Twitter. Both of them adopted the method of nonlinear preferential selection to describe the structure and evolution mode of online social network more accurately.

The nonlinear preferential selection probability of UNPE model can be found in formula (9), $P(i)$ is the possible being linked probability of node k_i and N stands for the node number in the whole model. As for the directed DNPE Model, k_{i_out} is the out degree of node k_i and k_{i_in} is the in degree of node k_i need to be calculated separately to UNPE model, as shown in formula (10) and (11).

$$P(i) = \frac{k_i^{1+0.5\lg(k_i)}}{\sum_{j=1}^{N} k_j^{1+0.5\lg(k_j)}} \tag{9}$$

$$P_{in}(i) = \frac{k_{i\_in}^{1+0.5\lg(k_{i\_in})}}{\sum_{j=1}^{N} k_{j\_in}^{1+0.5\lg(k_{j\_in})}} \tag{10}$$

$$P_{out}(i) = \frac{k_{i\_out}^{1+0.5\lg(k_{i\_out})}}{\sum_{j=1}^{N} k_{j\_out}^{1+0.5\lg(k_{j\_out})}} \tag{11}$$

4.2.1 Network evolution model

The whole evolution process of the UNPE model and DNPE model includes three steps: the joining of new nodes, the generation and the disappearance of new links. Meanwhile, compared with the relevant classic theories of ER Random Network Model and Small World

Network Model, the generation of new links in UNPE and DNPE are more likely to establish between the new node and the “nodes with more related links”, so the whole evolution process is as follows:

(1) Network initialization: The initial network is a stochastic network with m_0 nodes (with the maximum node number M), e_0 edges, which ensures that each node has at least one link connection and there is no isolated node in the network.
(2) Network evolution: Within each time step, select $1/M$ of the whole network as a local target network randomly, and repeat the following procedure:

   1) New nodes joining: According to an experienced probability p_1(always $p_1 = 0.5$), add the new generated node to the selected local target network and connect the new node to the existing m_1 nodes by higher probability to the nodes ”with more related links”. Then the node connection matrix of the whole network will be updated automatically.
   2) Mechanism I of new link generation: With probability p_2, m_2 new links will be generated in the network while each of these links will select a node and its neighbor’s neighbor to connect within the network randomly. New link generation of Mechanism I will be repeated for m_2 times, and then the node connection matrix of the whole network will be updated automatically.
   3) Mechanism II of new link generation: With probability p_3, m_2 new links will be generated in the network. As it for ttogetherhe UNPE model, each of these m_2 links will select a node and another node with larger $P_{in}(i)$ to connect together within the network randomly according to formula (9). As it for the DNPE model, we will select a node as a followers (e.g. in Twitter) with larger $P_{in}(i)$ according to the formula (10), and select another node with larger $P_{out}(i)$ according to the formula (11) to be the one which is be followed. Then it would generate m_2 directed links.
   4) The disappearance of new links: With probability p_4 we remove an edge from the network randomly. If the death of this edge results in isolated nodes, the isolated node will be connected by the above step 2 and step 3. If death of this edge results make the network turn into several disconnected small group, the remove of this edge will be abandoned.

4.2.2 Group opinion polarization process

According to the classical theory of group polarization, priority of being access of social network user can be expressed by PageRank value to evaluate the felt pressure of each user [20]. According to the connection matrix $A_{N\times N}$, the PageRank matrix $R_{N\times N}$ can be achieved. Basing on some society comparing theory, the distance between two nodes can influenced the ”credit” [21]. Therefore , the distance of shortest path d_{ij} (matrix is D_{ij}) between two nodes can be used to stands for the ”distance” and it is assumed to be related with the felt pressure of each node.

In the group polarization model, the felt pressure of user i from another node j is directly proportional to the influence ability of node j (expressed by PageRank value R_j), and inversely proportional to d_{ij}. The felt pressure of node i can be presented as I_{ij} which can be calculated in matrix $I_{N\times N}$ in formula (12).

$$I_{ij} = R_j/D_{ij} \tag{12}$$

The average cumulative felt pressure of node i in group network towards some public event can be found as P_i in formula (13). It comes from the Hopfield network model [16, 22].

$$P_i = \frac{\sum_{j=1}^{N} I_{ij} S_j}{N} \tag{13}$$

N is the total number of nodes in the network, S_j is the emotional tendency flag of node j and can be assigned value of 1(stands for "support"), -1(stands for "not support") or 0(stands for "indifferent"). S_j of all nodes will compose the group emotional tendency matrix $S_{1\times N}$. The final stable group emotional tendency is not only related to the group average cumulative pressure, but also depends on the initial perspective emotional tendency. S_i is the initial emotional tendency of node i, and the perspective tendency change of node i is also related to the parameter $\beta S_i + (1-\beta)P_i$ in which β is the adjustable parameters of $0 < \beta < 1$ and always be assign with β=0.5 in our simulation experiment. Our simulation experiment can be divided into two parts. The first part is to generate the network dataset using model in Section 4.2.1. The second part is group opinion polarization process with the random threshold value π_{thresh}. When $\beta S_i + (1-\beta)P_i > \pi_{thresh}$, the initial emotional tendency value S_i of node i would be updated to 1 while $\beta S_i + (1-\beta)P_i < -\pi_{thresh}$, S_i would be updated to -1 otherwise value of S_i would remain unchanged. The second part of our experiment will be carried iteratively until the emotional tendency status of the whole network comes to the final stable status.

Based on the above theory, we can get the simulation results of the UNPE and DNPE group polarization model based on the network simulation of the Tianjin explosion events, and compare it with the true twittering relation graph of Tianjin explosion in Section 6.1.

5 Graded risk list for emergency accident handling

5.1 Risk classification

According to the accumulated emergency accidents by our research team in recent years, we have proposed a detailed risk classification method with five layered risk which can be found in Table 1.

Risk grade judgment is based on content analysis. An important step in the content analysis is to extract and quantify information entropy of twittered text. The extracted content builds an identifiable vector relating to the key words of risk grade in the content of Sina Micro Blog and WeChat at the initialization of pre-judgment model. Furthermore, we have has been constructed a method combining the word frequency method, the mutual information entropy method and the relative conditional entropy based context key word extraction method together. We use them to extract and identify content sentiment eigenvector and finally transfer the eigenvector to judge the risk grade in the content of Sina Micro Blog and WeChat [6, 9].

Using the above three methods, it can carry out the extraction and identification to some particular content key words when searching, and then calculate the oriented risk grade. If there is a basic conformity, a primary early- warning can be launched for government; if not, other operations could be continued. For example, in the primary risk judgment procedure, key words, such as 'great', 'bad', 'injuries and deaths', can be marked when building the identifiable vector of key words of Sina Micro Blog content, using the word frequency method to decrease the data dimension, to abolish the noisy data and to reduce the scope. To

Table 1 Graded emergency accident risk

Risk grade and characteristics	Events types	Events examples
Primary risk Characteristics: events that cannot be prejudged and controlled by humankind and that are unexpected and greatly harmful to the safety of life and property as well as extremely bed effects on the stability of country and society.	Natural disaster, Geologic and meteorological disasters, Ecological disaster, Violence and terror incidents	Typhoon HaiYan in 2013, Hagupit Typhoon in 2014, Wenchuan earthquake in 2008, SARS in 2003 , KunMing Railway Station terrorist attack in 2014.
Secondary risk Characteristics: unexpected mega-events due to human factor, resulting in certain damage to the safety of life and property and to the economy and society.	Safety accidents, Traffic accidents, Environmental pollution in a large scale and so on	Fire in Harbin in 2016, Explosion at Tianjin Port in 2015.
Tertiary risk Characteristics: events that has no threat to safety of life and property, but a great influence on the social stability and public opinion guidance	Issue of law about livelihood, Implementation of policy and so on, National and even global events caused by technological economy	Delayed Retirement Act in China in 2015, Anticorruption Act in CCP launched in 2014, The crash of stock market in China in 2015, Global financial crisis in 2013.
Quaternary risk Characteristics: events that has no threat to safety of life and property, but a certain influence on the social stability and public opinion guidance	Domestic and foreign policy, Diplomatic incidents	American presidential election in 2016, TPP launched by Obama in 2015 , Xi Jinping visit USA in 2015,the 2015 China Victory Day Parade of Counter-Japanese War (1937-1945).
Others Characteristics: informational events that has no threat to safety of life and property and no significant influence on the social stability and public opinion guidance	Event of ordinary public figures, Events like typical phenomenon of society and so on	Attack on a girl in Yitel Hotel in Beijing in 2016.

confirm the relevance between the Sina Micro Blog content and 'the primary risk category', the mutual information method and the relative conditional entropy based context key word extraction method can be used to calculate the mutual information value (i.e. relative entropy) of feature quantity from key words, while finally to confirm the contribution degree of the Sina Micro Blog content to the entire 'primary risk category' for choosing the selected situation of the feature quantity [7, 14, 23–25].

Basing on the collected online public opinion data of hot accidents in the past decade, we calculate and attain the graded risk list in Table 1 which divided into five types of risk and it may be useful for government to make decision in case of emergency. In addition, the hot accidents are classified into five different types which can be found in Table 2.

5.2 Accident types

Depending on the type of network public opinion of hot event, it can be divided into five areas, political, society, the people's livelihood, natural or man-made disasters and other.

6 Case evaluation with CPS model

The efficient CPS model based online opinion governance modeling and evaluation system proposed in this article has been applied into some real emergency accidents such as Typhoon HaiYan in 2013, Typhoon Hagupit in 2014 and Explosion at Tianjin Port in 2015.Our system has been proved useful and helpful for government officers to make decisions in facing emergency accidents and disasters.

6.1 Explosion at Tianjin port in 2015

The 2015 Tianjin explosion accident composed of a series of explosions which killed over one hundred people at a chemicals storage company in Tianjin Port on Wednesday, August 12 2015. It has been regarded as the most harmful and dangerous explosion accident after the foundation of PR China since 1949. The first two explosions occurred within 30 seconds at the chemicals storage company, which located in the BinHai New Distric of Tianjin. The second explosion was much larger and caused by about 800 tons of ammonium nitrate. Fires caused by the initial explosions continued to burn out of control throughout the weekend,

Table 2 Accident types

The type of event	Instructions
Political	The events involving diplomatic or political incidents happened at homeland and abroad.
People's livelihood	The events involving the livelihood of people, such as laws enacted and implementation of policy.
Society	The events involving the public safety, social celebrities and social phenomenon etc.
Natural/Man-made disasters	Natural disasters, accidents, transportation accidents and environment pollution etc.
Other	The events triggered by the science, technology and economic factors.

Fig. 4 Word frequency of SINA micro blog

repeatedly causing some secondary explosions. On 12 September 2015, the government reported 173 deaths, 8 missing, and 797 non-fatal injuries in this accident.

6.1.1 *Cyber domain*

Figure 4 demonstrates the key word frequency calculated by our system from August 12 to August 31 by Sina Micro Blog data. And we can found that the calculation result precisely shows the attention focus of social network user. The circled focus Chinese word stands for the explosion location and the explosion event can be found in the central area in Fig. 4. And the calculated word frequency result has been proved useful to find out focus of online opinion.

We can get event data area of physical domain and the initial situation. Our system (in adjustable intervals) can collect the online information from social network in each 5 minutes, and the user can monitor events status in real-time development, according to meteorological situation, hydrological situation, mass evacuation routes and deployment of firefighters' disposal program.

Tables 3 and 4 shows the Word Frequency Rank in the early stage and the later stage. It can be found that public opinion focus on the accident victims when the incident occurred. Lately, people become concerned about conducting a thorough investigation and show respect to the sacrifices of the heroes of fire fighter. All the words in Column "Word" of Tables 3 and 4 has been translated from Chinese characters to English words for the comprehensive convenience of English speakers.

We use a method for automatic summarization based on LDA model and information entropy for Chinese document. It uses LDA model to do shallow semantic analysis work on documents and gets the distribution of topics under each document. Through analyzing the topics of document, we get the topic which has the best expression of central idea for document. Meanwhile, our paper has proposed a new method to compute the sentence weight and extract the most important sentence based on measuring the information entropy for each sentence. It treats the sentence as a random variable and calculates the information

Table 3 Word frequency rank in the early stage

Words	Part of speech	Weights	Word Frequency
Explosion	Verb,Noun	5.68	2753
Occur	Verb	5.82	874
Tianjin BinHai New Area	New Noun	6.52	599
Explosion	New Noun	10.07	463
Clifford	Verb	6.24	457
Dangerous goods stores	New Noun	3.81	338
Hope	Verb	6.7	300
Tianjin Binhai	New Noun	4.49	280
Accident	Noun	6.71	271
Victims	Intransitive Verb	4.95	269
Hospital	Noun	6.7	260
Help	Verb,Noun	6.75	246
Know	Verb	5.78	232
Store	Noun	3.67	227
Injured	Intransitive Verb	4.86	218
Blessing	Verb	7.06	191
Binhai New Area	New Noun	7.89	191

entropy for every random variable. Experimental results show that this method can pick out the most important sentence in the document.

By the mentioned above table, it can be found that, at the beginning of the Tianjin Explosion event, people greatly concerned about the incident itself. And lately people have expressed concern about the injured persons while focus has changed.

6.1.2 Society domain

In Fig. 5, it can be found that the hot search distribution of Tianjin Explosion Event which has demonstrated the human behavior of paying attention from the society domain. And we can found that the ”Topic heat” is focus on ”Tianjin” and ”Explosion” while the fastest rising search words is the location of the Tianjin Explosion Event which is ”Changchengji street”. Furthermore, another very clear district of Tianjin to the explosion point which named ”Dagang” is found in the list to show that the people also care about the the neighbors to the explosion event. Figure 5 has shown us the online searching hot words and its distribution to reflect the attention focus of people.

Figure 6 is the overall trend line which based on the result of Baidu index when “Tianjin explosion” as the search keyword. As it can be seen from the chart, the incident search index on day of August 13 has reached its peak while in the next two days to August 15 the search index reached a valley. And before August 17, the search index has more moderate ups and downs while finally has dropped down in 8 days after August 24.

We have found that more than 3,210,000 points of Baidu search results and Baidu news over 3,050,000 articles by keywords of ”Tianjin explosion” by September 9, 2015. And the news has been broadcasted over 66,962,864 times by Sina Weibo and over 519,465 times by Tencent Micro Blog. We have collected the data and analyzed the case by some useful methods [26, 27].

Table 4 Word frequency rank in the late stage

Word	Part of speech	Weights	Word Frequency
Explosion	Verb,Noun	6.23	7306
Tianjin	Noun	8.78	4942
Explosion	New Noun	8.23	1409
Occur	Verb	7.36	1176
Victims	New Noun	8.73	1020
Accidents	Noun	8.63	973
Tianjin harbor	ns	6.76	881
Tianjin Binhai New Area	New Noun	8.29	861
Ruihai company	New Noun	9.3	695
News	Noun	7.97	674
Government	Noun	9.09	582
Hope	Verb	7.69	528
Clifford	Verb	7.01	518
Disaster	Noun	7.45	486
Marina	Noun	7.09	470
The company	Noun	8.65	469
Life	Noun	7.56	446
Know	Verb	6.7	437
Help	Verb,Noun	8.03	419
Survey	Verb	6.47	419
Warehouse	Noun	6.07	404
Binhai New Area	New Noun	9.21	391
Country	Noun	8.52	386
Hero	Noun	6.82	377
Jobs	Verb,Noun	7.53	369

6.1.3 Group opinion polarization analysis

Based on the Group Opinion Polarization Analysis simulation method mentioned above, we get the simulation results of the UNPE group polarization model based on the network simulation of the Tianjin explosion events, and compare it with the Barabási group polarization Model [28] and the true Twittering relation graph of Tianjin explosion from the crawled Sina Micro Blog data. The true Twittering relation graph includes 55 focus vertex and 200 edges which stand for the very active people in the discussion towards the explosion.

It could be seen from Figs. 7 and 8 in the experiment result that in the simulation of group polarization network, the UNPE model conforms to the actual situation more in the PapeRank index and HITS (Hyperlink Induced Topic Search) index of group polarization effect, but shows inconformity with the actual network in Katz centrality index. Through the research and predicting of the polarization phenomenon of mass viewpoint of public online opinion, we can make a pre judgment and conduct early screening and management towards the tendency of public opinion in different social networks in cyber domain. And UNPE model is useful to prevent the spread of rumor information, to promote the dissemination of effective information on the aspect.

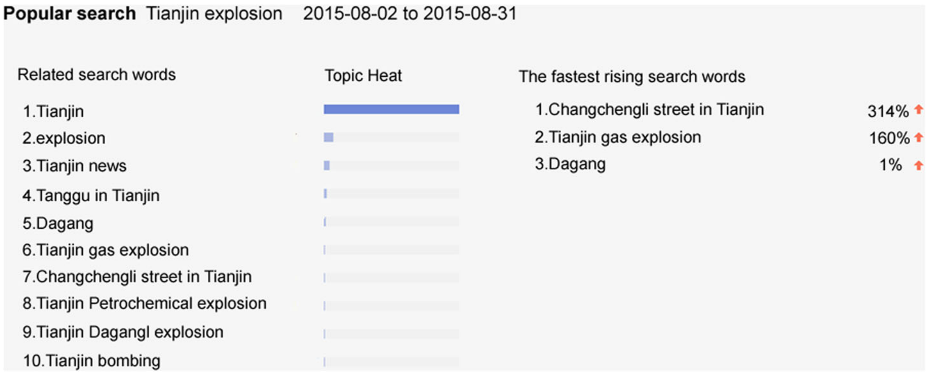


Fig. 5 Hot search distribution of Tianjin event

6.2 Typhoon HaiYan in 2013

Typhoon Haiyan happened in November, 2013. We have separated the moving process of Haiyan into three periods. The first period was before Haiyan landed, which was from November 5th, 2013 to November 6th, 2013. During that period, typhoon was formed in the Northwest Pacific ocean and moved towards Northwest. The intensity of Haiyan was increased and became a super typhoon when it got closed to Philippine. The maximum wind power reached to the level of 17 in the center. The second period was on November 7th, 2013 when Haiyan landed on Philippine. The intensity was increased by the super power typhoon. The last period was from November 8th, 2013 to November 11th, 2013. Haiyan moved to the South China Sea and its intensity has become weakened. On November 10th, 2013, it grazed Hainan and caused storm. Next day, it landed on Guangxi and its intensity has continually been decreased.

6.2.1 Cyber domain

We collect the real-time microblog messages from SINA by crawled the required data through keyword on SINA API. The data we crawled was from November 3rd, 2013 to

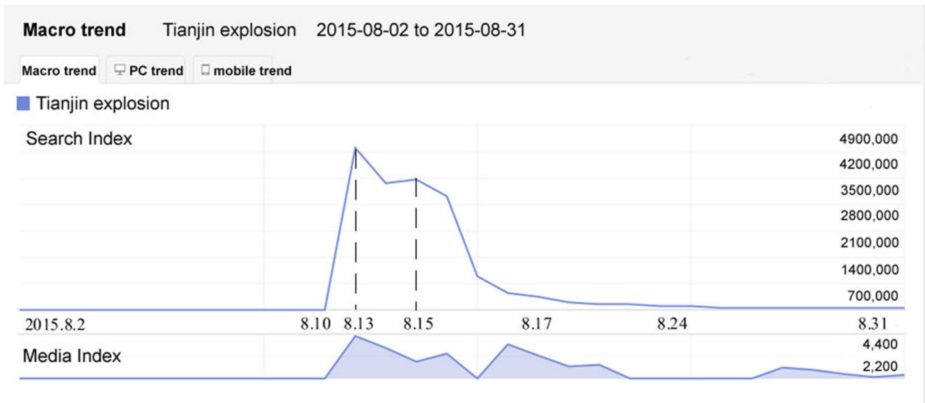
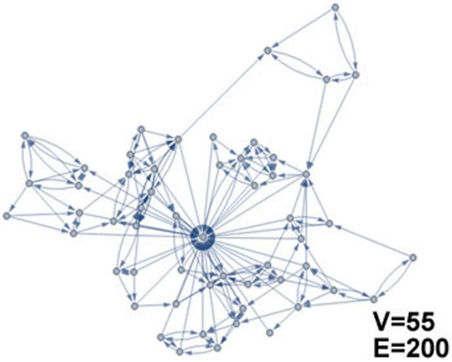
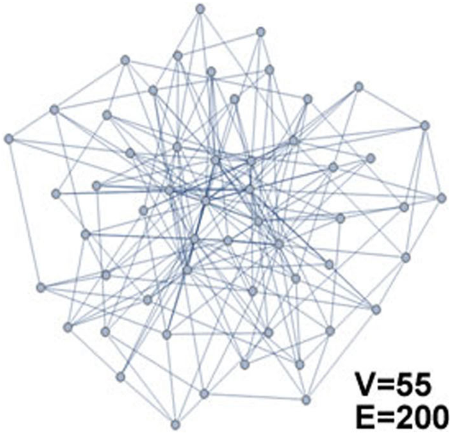


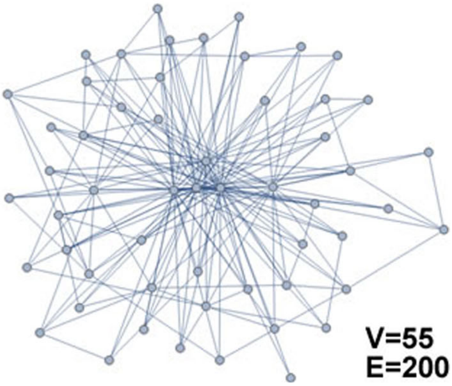
Fig. 6 Online public opinion data



(a) True Twittering relation graph



(b) Simulation results by Barabási model



(c) Simulation results by B UNPE model

Fig. 7 Group polarization graph

November 15th, 2013. We found that the number of the microblog was changed according to different situation changes in Haiyan and the related measures taken by the government. For example, in Fig. 9, it is clear that the number of microblog posts was increased to 1200 on November 11th 2013, which was also the date when typhoon Haiyan landed on Guangxi. When the number reached to the maximum value, it decreased dramatically because during those days, government had taken some related measures to handle this emergency accident. And from Fig. 9, we can see that measures worked and have efficiently made the discussion less hot.

Since Haiyan was only landed in Hainan and grazed in Guangxi, the following data mining such as the word frequency analysis and other related analysis based on the cyber information space is only executed in Hainan and Guangxi.

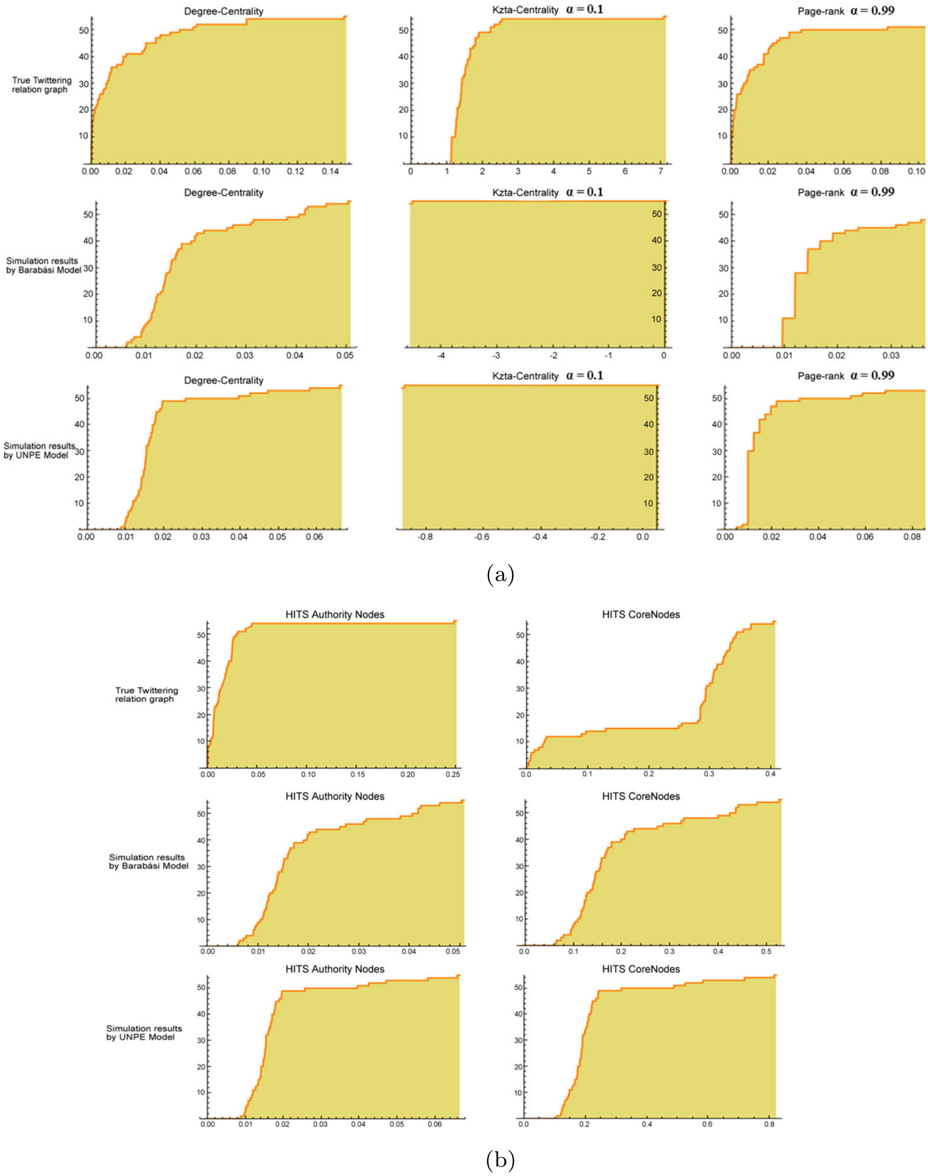


Fig. 8 **a** Degree, Kzta, Pagerank statistic result, **b** HITS statistic result

After obtained the opinion data, we make analysis, such as data cleaning, segmentation and analysis of the word frequency. Firstly, we do the data cleaning to remove the unrelated and useless data and find out that the number of microblog about Haiyan is 2539. Next, we collect words about the same meaning into one group to find out the valuable information

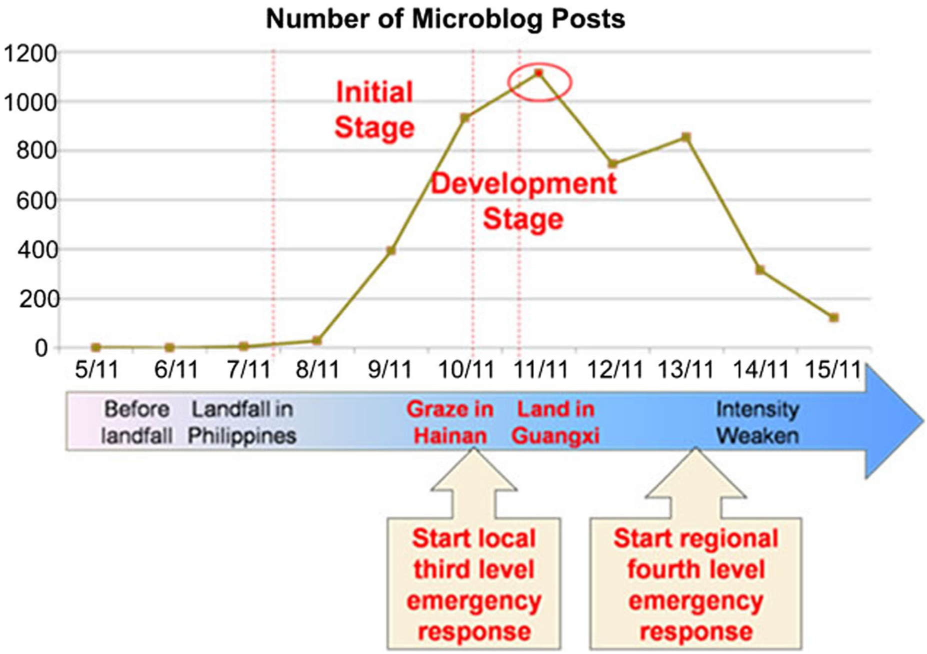


Fig. 9 Opinion development timeline of Haiyan

given by microblog. After that, we have ranked these words and conducted the word frequency analysis. The outcomes are showed in Tables 5 and 6.

From Table 5, we can clearly find out the high frequency words posted by Hainan citizens. These hottest discussion words include ”power outage”, ”Ledong County”, ”Wanquan River”, ”rescue” and ”without water”. It tell us that citizens in Hainan pay their attention much on the damage of the infrastructure, such as power, water, food supply, affected location and the recovery plan. From Table 6, the hottest discussion words are ”rain storm”, ”umbrella”, ”attend class”, ”go to work” and ”power recovery”, which indicates that citizens in Guangxi province focus on the storm caused by typhoon Haiyan.

Comparing with Tables 5 and 6, we can discover that citizens in Hainan have much more concern on their basic needs, such as rescue, food, water, power. However, people in Guangxi concerns much about the coming storm, recovery and they also want to go out soon. It has a reason because the damage in Hainan is much more severe than the damage in Guangxi by using the Maslow’s Theory of the Level Demand. Interestingly, it is Guangxi rather than Hainan that the typhoon Haiyan landed.

6.2.2 Society domain

From Fig. 10, we can see that most of the Chinese citizens hold the negative emotions towards Haiyan especially on the day when it landed on Guangxi. It made Chinese people become much fear because it is so close to their real-life.

Table 5 High frequency words posted by Hainan citizen

Number	Keywords	Frequency
1	Haiyan	142
2	strong typhoon	94
3	power outage	88
4	Hainan	74
5	@News Head	7
6	distance	7
7	you	7
8	@Daily	6
	International Tourism	6
9	Ledong County	6
10	@CCTV	6
11	16 o'clock	6
12	about (‘’)	6
13	Kilometer	6
14	Cry	5
15	trapped (‘’)	4
16	Wanquan River	4
17	Rain&wind	4
18	village	4
19	rescue	4
20	without water	2

Table 6 High frequency words posted by Guangxi citizen

Number	Keywords	Frequency
1	Haiyan	111
2	typhoon	86
3	@News Head	51
4	Guangxi	47
5	morning	43
6	domestic (jingnei)	41
7	enter	41
8	cry	15
9	go out	14
10	rain storm	11
11	Ledong County	10
12	attend class	10
13	go to work	7
14	shocked (chi jing)	6
15	candle	6
16	umbrella	6
17	recovery	6
18	attack	4
19	power recovery	4
20	affected (shouzai)	4

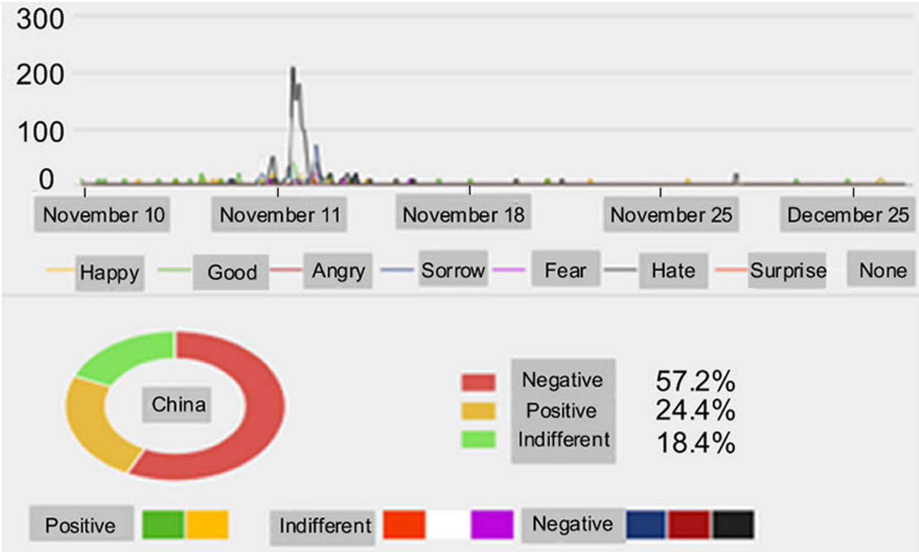


Fig. 10 Sentiment analysis on typhoon Haiyan

7 Conclusion

In this paper, we have proposed a new CPS Model based Online Opinion Governance system which constructed on cellphone APP for data collection and decision make in the back end. Our contributions are as followed: we propose the new graded risk classification method and new accident classification method. We propose the new group opinion polarization analysis method and have made very good promotion of the relative conditional entropy based context key word extraction method. Based on our contribution, we have built an efficient CPS Model based simulated emergency accident replying and handling system. It has been proved that integrated data from cyber domain, physical domain, and society domain will help us to make the final management decision. Furthermore, we proposed the detailed analysis of 2015 Tianjin Explode accident and Haiyan Typhoon in 2013.

Acknowledgments Thanks to Philosophy and Social Science Project of Education Ministry (No. 15JZD027), National Culture Support Foundation Project of China (2013BAH43F01), and National 973 Program Foundation Project of China (2013CB329600) in social network analysis. We appreciate direction from professor Hui Zhang andhis aid form Joint-Operated project from National Natural Science Foundation of China (NSFC) (Grants No. 91224008-14) from Tsinghua University.

References

1. Joseph K, Landwehr PM, Carley KM (2014) An approach to selecting keywords to track on twitter during a disaster. In: Proceedings of the 11th International ISC. Conference – University Park. https://doi.org/10.1037/0003-066X.36.4.343RAM
2. Slavkovikj V, Verstockt S, Hoecke SV et al (2014) Review of wild fire detection using social media. Fire Safe J 68:109–118. https://doi.org/10.1016/j.firesaf.2014.05.021

3. Deng Q, Liu Y, Zhang H et al (2016) A new crowdsourcing model to assess disaster using microblog data in typhoon Haiyan[J]. Nat Hazards J Int Soc Prevent Mitigation Nat Hazards 84(2):1–16. https://doi.org/10.1007/s11069-016-2484-9
4. Ma YF, Deng Q, Wang XZ, Liu JQ, Zhang H (2014) Keyword-based semantic analysis of microblog for public opinion study in online collective behaviors, web-age information management lecture notes in computer science 2014, pp 44–55. https://doi.org/10.1007/978-3-319-11538-2_5
5. German N, Leonie R, Astrid M et al (2014) Psychosocial functions of social media usage in a disaster situation: a multi-methodological approach. Comput Hum Behav 34:28–38. https://doi.org/10.1016/j.chb.2014.01.021
6. Onorati T, Malizia A, Diaz P, Aedo I (2014) Modeling an ontology on accessible evacuation routes for emergencies. Exp Syst Appl 41:7124–7134. https://doi.org/10.1016/j.eswa.2014.05.039
7. Hadiguna RA, Kamil I, Delati S, Reed R (2014) Implementing a web-based decision support system for disaster logistics: a case study of an evacuation location assessment for Indonesia. Int J Disaster Risk Reduct 9:38–47. https://doi.org/10.1016/j.ijdrr.2014.02.004
8. Campos V, Bandeira R, Bandeira A (2012) A method for evacuation route planning in disaster situations. Procedia - Soc Behav Sci 54:503–512. https://doi.org/10.1016/j.sbspro.2012.09.768
9. Ndiaye IA, Neron E, Linot A, Monmarche N, Goerigk M (2014) A new model for macroscopic pedestrian evacuation planning with safety and duration criteria. Transport Res Procedia 2:486–494. https://doi.org/10.1016/j.trpro.2014.09.064
10. Li R, Zhang H-p, Zhao Y-p, Shang J-y (2014) Automatic text summarization research based on topic model and information entropy[J]. Comput Sci 41(11a):298–332
11. Erdös P, Rényi A (1960) On the evolution of random graphs [J]. Publ Math Inst Hung Acad Sci Ser A 5:17–61. 05C80
12. Watts DJ, Strogatz SH (1998) Collective dynamics of “small world” networks [J]. Nature 393(4):440–442. https://doi.org/10.1038/30918
13. Bababasi AL, Albertr (2512) Emergence of scaling in random networks [J]. Science 286(509):1999. https://doi.org/10.1126/science.286.5439.509
14. HaiBo H (2010) Research on the structure, evolution and dynamics of online social networks [D]. Shanghai Jiao Tong University, Doctoral dissertation
15. Li X, Chen G (2003) A local world evolving network model [J]. Phys A 323:274–286. https://doi.org/10.1016/S0378-4371(03)00604-6
16. Macy MW, Kitts JA, Andreas F (2003) Polarization in dynamic networks: a Hopfield model of emergent structure [J]. The National Academies Press, (163)
17. Maria MC, Neville AS, Ioannis M (2015) The concept of risk situation awareness provision: towards a new approach for assessing the DSA about the threats and vulnerabilities of complex socio-technical systems. Saf Sci 79:126–138. https://doi.org/10.1016/j.ssci.2015.05.012
18. Newman MEJ, Clauset A (1863) Structure and inference in annotated networks. Nat Commun 7(1):2016. https://doi.org/10.1038/ncomms11863
19. Xiao Z, Travis M, Newman MEJ (2015) Identification of core-peripherystructure in networks. Phys Rev E 91:032803. https://doi.org/10.1103/PhysRevE.91.032803
20. Wang L (2009) Amelioration of PageRank Algorithm [D]. Shanghai Jiao Tong University, Unpublished master’s thesis
21. Latané B (1981) The psychology of social impact [J]. Amer Psychol 36:343–365. https://doi.org/10.1037/0003-066X.36.4.343
22. Hopfield JJ (1982) Neural networks and physical systems with emergent collective computational abilities [J]. Proc Nat Acad Sci 79:2554–2558. https://doi.org/10.1073/pnas.79.8.2554
23. Sun D, Zhang LH, Chen FX (2013) Comparative study on simulation performances of CORSIM and VISSIM for urban street network. Simul Model Pract Theory 37:18–29. https://doi.org/10.1016/j.simpat.2013.05.007
24. Wilensky H (2014) Twitter as a navigator for stranded commuters during the great East Japan Earthquake. In: Proceedings of the 11th International ISCRAM conference. University Park
25. Seppänen H, Virrantaus K (2015) Shared situational awareness and information quality in disaster management. Saf Sci 77:112–122. https://doi.org/10.1016/j.ssci.2015.03.018
26. Mora K, Chang J, Beatson A, Morahan C (2015) Public perceptions of building seismic safety following the Canterbury earthquakes: a qualitative analysis using Twitter and focus groups. Int J Disaster Risk Reduct 13:1–9. https://doi.org/10.1016/j.ijdrr.2015.03.008
27. Gaspar R, Pedroc C et al (2016) Beyond positive or negative: qualitative sentiment analysis of social media reactions to unexpected stressful events. Comput Hum Behav 56:179–191. https://doi.org/10.1016/j.chb.2015.11.040
28. Barabási A-L, Albert R (1999) Emergence of scaling in random networks [J]. Science 286:509–512

Xiao Long Deng, born in 1977, PHD, is interested in data mining from Beijing University of Post and Telecommunication(BUPT). His research interests include data mining and emergency management.

Yin Luan Yu, born in 1974, a professor in China Academy of Social Management, Beijing Normal University Beijing. Her research interests include social computing and emergency management.

Dan Huai Guo, born in 1975, an associate professor in Computer Network Information Center of Chinese Academy of Sciences. His research interests include social network analysis and data science.

Ying Tong Dou, born in 1994, a senior student majoring in Telecommunication Engineering with Management at BUPT International School. He is a research assistant at the Key Laboratory of Trustworthy Distributed Computing and Service and interested in social network analysis and data science.