Crowdsourcing based Description of Urban Emergency Events using Social Media Big Data

Zheng Xu, Member, IEEE, Yunhuai Liu, Member, IEEE, Neil Y. Yen, Member, IEEE, Lin Mei, Member, IEEE, Xiangfeng Luo, Member, IEEE, Xiao Wei, and Chuanping Hu, Member, IEEE,

Abstract—Crowdsourcing is a process of acquisition, integration, and analysis of big and heterogeneous data generated by a diversity of sources in urban spaces, such as sensors, devices, vehicles, buildings, and human. Especially, nowadays, no countries, no communities, and no person are immune to urban emergency events. Detection about urban emergency events, e.g., fires, storms, traffic jams is of great importance to protect the security of humans. Recently, social media feeds are rapidly emerging as a novel platform for providing and dissemination of information that is often geographic. The content from social media usually includes references to urban emergency events occurring at, or affecting specific locations. In this paper, in order to detect and describe the real time urban emergency event, the 5W (What, Where, When, Who, and Why) model is proposed. Firstly, users of social media are set as the target of crowd sourcing. Secondly, the spatial and temporal information from the social media are extracted to detect the real time event. Thirdly, a GIS based annotation of the detected urban emergency event is shown. The proposed method is evaluated with extensive case studies based on real urban emergency events. The results show the accuracy and efficiency of the proposed method.

Index Terms—Crowdsourcing, emergency events, social media, big data, urban computing

I. INTRODUCTION

A. Motivation and Purpose

Crowdsourcing is a process of acquisition, integration, and analysis of big and heterogeneous data generated by a diversity of sources in urban spaces, such as sensors, devices, vehicles, buildings, and human [1]. With the help of cloud computing [2, 3, 39, 40], internet of things [4, 5], and Big Data [6, 7], crowdsourcing connects unobtrusive and ubiquitous sensing technologies, advanced data management and analytics models, and novel visualization methods, to create solutions that improve urban environment, human life quality, and city operation systems [8]. Especially, nowadays, no countries, no communities, and no person are immune to urban emergency events [9]. An urban emergency event is a sudden, urgent, usually unexpected incident or occurrence that requires an immediate reaction or assistance for emergency situations faced by social group (e.g., the corporations) or the recipients of public assistance [10]. For example, the urban resident may face fires, storms, traffic jams and so on. Thus, it is important to detect, resistant, and analyze these real time urban emergency events to protect the security of urban residents.

Crowdsourcing is also an emerging computing paradigm that tasks everyday mobile devices to form participatory sensor networks. It allows the increasing number of mobile phone users to share local knowledge acquired by their sensor-enhanced devices, e.g., to monitor pollution level or noise level, traffic condition, etc. The sensing data from volunteer contributors such as social network users can be further analyzed and processed, and leveraged in many areas such as environment monitoring, urban planning, emergency management, as well as public healthcare/safety. Weibo1, a popular Chinese micro blogging service similar to Twitter2, has received much attention recently. This online social network service is used by about 500 millions of people in China to remain socially connected to their friends, family members, and colleagues through their computers and mobile phones. The user of Weibo concerns one question, “What’s happening?” The poster of each user must be fewer than 140 Chinese words. A status update message is often used as a message to friends and colleagues. A user can follow other users; that user’s followers can read her messages on a regular basis. An important feature of Weibo service is its real time nature. The large number of posted messages includes urban emergency events such as storm, fire, traffic jam, riots, heavy rainfall, and earthquakes. In fact, a Weibo user can be seen as a sensor of an urban emergency event. By urban emergency events, we mean important phenomena with a local and temporal dimension in the physical world [13]. For example, if a user makes a message in Weibo about a fire or crash, then she/he can be seen as a “fire sensor” or “crash sensor”. The social network such as Weibo can be seen as a sensor receiver. Usually, the Weibo users can be as “social sensors [11].” A social sensor is defined as an

1 www.weibo.com
2 www.twitter.com

This work was supported in part by the National Science and Technology Major Project under Grant 2013ZX01033002-003, in part by the National High Technology Research and Development Program of China (863 Program) under Grant 2013AA014601, in part by the National Science Foundation of China under Grant 61302020, in part by the China Postdoctoral Science Foundation under Grant 2014M560085, and in part by the Science Foundation of Shanghai under Grant 12ZR1452900.

Zheng Xu is with The Third Research Institute of the Ministry of Public Security and Tsinghua University, Beijing, China (e-mail: xuzheng@shu.edu.cn).

Y. Liu, L. Mei, and C. Hu are with The Third Research Institute of the Ministry of Public Security, Shanghai, China.

Neil Y. Yen is with University of Aizu, Japan. X. Luo is with Shanghai University. X. Wei is with Shanghai Institute of Technology.
agent that provides information about its environment on a social network after interaction with other agents [12]. The sensing message from social sensors can be used during a live fire emergency or traffic updates.

B. Significance in Contributions

In this paper, in order to describe the real time urban emergency event based on crowdsourcing using Weibo, the 5W (What, Where, When, Who, and Why) model is proposed. The 5W model provides five basic elements of an urban emergency event, which is summarized as follows.

(1) What. The most important element of the proposed 5W model is to detect what happened in the urban environment. For example, if a user posts a message in Weibo about a fire or crash occurrence, the proposed 5W model must detect that real time urban emergency event.Besides the short text provided by Weibo user, multimedia data such as images and short videos can also be get. For example, a user may upload the real time image of a fire when he sees it.

(2) Where. Besides detecting what happened in the urban environment, it is needed to reveal the location information of the urban emergency event. Fortunately, microblog services have become a location information platform of users’ daily life. In China, the major location based service applications Jiepang\(^3\) and Dianping\(^4\) both allow their users provide check-in data through Weibo. Stefanidis et al. [14] reported that approximately 16% of the Twitter feeds in their experiments had detailed location information with it in the forms of coordinates, while about 45% of the tweets they collected had some geolocation information at the city level. Fig. 1 shows the check-in information of China and Beijing\(^5\). The check-in information usually appears in the modern city of China and the center of Beijing. Usually, the possibility of the appearance of urban emergency events is higher in the modern city and the center of the city. The check-in information can be used as the location information of the urban emergency event.

(3) When. Weibo has a very good real time feature. Some important instances verify this nature: a damage fire in Guangzhou\(^6\) is first published in Weibo and the image of a riot between polices and residents in Wenzhou are also transmitted by Weibo. Each Weibo message has a timestamp, which can be used for revealing the occurrence time of an urban emergency event. Besides the occurrence time, the proposed 5W model wants to show the timeline of an urban emergency event. For example, at the beginning, an event may be in a latent state. The number of Weibo messages about it may be low, only a few social sensors focus on it. When some milestones things happen and post it on the main websites as headline news, an event may be in an outbreak state. So many social sensors talk about it. Of course, at last, an event may be in a decline state. The number of Weibo messages about it may be low again.

(4) Who. Different person act different roles in an urban emergency event. Social sensors may act as the witness of an urban emergency event since they are at the place of the urban emergency event. For example, when a Weibo user takes a picture of a happening fire event, he can be seen as the witness of that fire. Besides the witness, some people act as the participator of the urban emergency event. For example, a person may act as the suspect of a robbery event. The proposed 5W model wants to provide the witness and participator of an urban emergency event.

(5) Why. An emergency event is a sudden, urgent, usually unexpected incident or occurrence that requires an immediate reaction or assistance for emergency situations. Since the huge damage and influence of the urban emergency event, it is important to collect the reason after the decline of that event. The upload message by social sensors may reveal the reason for an urban emergency event. For example, a Weibo user may post a message “OMG, I saw a car crash a man who cross the red light”. The message posted by witness or participator can be used to investigate the potential reason of an urban emergency event.

The major contributions of this paper are summarized as follows.

a) This paper proposes a 5W model for describing urban emergency events. The proposed 5W model includes what, where, when, who, and why elements to detect and analyze the urban emergency event.

b) The proposed model is based on crowdsourcing, which uses the real time nature of Weibo users. The proposed model is applied into the emergency management field, which can provide useful information to analyze and resist urban emergency events.

c) Case studies on real data sets show the proposed model has good performance and high effectiveness in the analysis and detection of urban emergency events.

The rest of the paper is organized as follows. In the next section, the related work is given. Section 3 gives the problem formulation. Section 4 presents the proposed 5W model. Case studies on real data sets are conducted in Section 5. The last section gives the conclusion of our work.

II. RELATED WORKS

In this section, two aspects including event detection and social media based application are introduced. The summaries of challenges of using social sensors are also given.

---

\(^3\) www.jiepang.com

\(^4\) www.dianping.com

\(^5\) The date of the check-in information is from 2014.4.1-2014.5.1 provided by Weibo user

\(^6\) Guangzhou and Wenzhou are two cities in China.
A. Event Detection

Event detection based on prior user queries is reported in [15, 16]. Fung et al. [15] proposed to first identify the bursty feature related to the user query and then organize the documents related to those bursty features into an event hierarchy. In [16], a user specifies an event (or a topic) of interest using several keywords as a query. The response to the query is a combination of streams (e.g., news feeds, emails) that are sufficiently correlated and collectively contain all query keywords within a time period. The proposed work is also related to event detection using click-through data [17]. Event ranking with user attention is reported in [18] where the events are firstly detected from news streams. User attention is then derived from the number of page-views (collected through web browser toolbars) for all the news articles in the same event. Leskovec et al. [19, 20] proposed the method for outbreak detection based on cost-effective function.

Event evolution proposed by Makkonen [34] is a subtopic of topic detection and tracking. In his study, two important conclusions are given: (1) a seminal event may lead to several other events; (2) the events at the beginning may have more influence on the events coming immediately after than the events at the later time. Makkonen used the ontologies to measure the similarity of events. However, these ontologies are difficult to get, which makes the work difficult to be used directly. Mei [35] investigated theme evolution which is similar to event evolution. He proposed a temporal pattern discovery technique on the basis of the timestamps of text streams. The theme of each interval is identified, and the evolution of theme between successive intervals is extracted. Unfortunately, the proposed technique did not consider the different states of an event, which may impact on its result. Wei [36] proposed an event evolution pattern discovery technique which identifies event episodes together with their temporal relationships. An event episode is defined as a stage or sub-event of an event. The above study differs from this paper: their study deals with an event and their event episodes, whereas this paper handles the above study differs from this paper: their study deals with an event episode is defined as a stage or sub-event of an event. The event episodes together with their temporal relationships.

Later, Yang [37] aimed at discover event evolution graphs from news corpora. The proposed event evolution graph is used to present the underlying structure of the events. The proposed method uses the event timestamp, event content similarity, temporal proximity, and web pages distribution proximity to model the event evolution relationships. Recently, Jo [38] studied the method to discover the evolution of topics (i.e., events) over time in a time-stamp document collection. He tried to capture the topology of topic evolution that is inherent in a given corpus. He claimed that the topology of the topic evolution discovered by his method is very rich and carries concrete information on how the corpus has evolved over time.

B. Social Media based Application

Recently, with the high speed development of the social networks such as Twitter and Weibo, many researchers have published their work of using the data from social networks including special events for targeted advertising [26], marketing [27], localization of natural disasters [28], and predicting sentiment of investors [29, 30]. Sakaki et al. [21] investigated the real-time nature of Twitter, put particular attention to event detection. The twitter users are regarded as sensors. Their messages are used for detecting earthquake. A reporting system is developed for use in Japan by their proposed methods. Crooks et al. [22] thought Twitter as a distributed sensor system. They analyzed the spatial and temporal features of the Twitter feed activity responding to a 5.8 magnitude earthquake. Their experimental results argued that the Twitter users represent a hybrid form of a sensor system that allows for the identification and localization of the impact area of the event. Longueville et al. [23] used Twitter as a source of spatial-temporal information. By focusing on a real-life case of forest fire, they aimed to demonstrate its possible role to support emergency planning, risk and damage assessment activities. Besides the emergency events management, other researchers use the spatial and temporal information from social networks to support location based services. Liu et al. [24] presented MoboQ, the location-based real-time question answering service that is built on top of microblogging platform. Qu and Zhang [25] used Twitter user generated mobile location data for trade area analysis. Their model includes three key processes: identifying the activity center of a user, profiling users based on their location history, and modeling users’ preference probability.

Unfortunately, automatically detecting and resisting real time urban emergency events using social media is not that easy. The potential challenge is summarized as follow.

<table>
<thead>
<tr>
<th>Event</th>
<th>Time</th>
<th>UTC</th>
</tr>
</thead>
<tbody>
<tr>
<td>00:00</td>
<td>00:41</td>
<td>16:41</td>
</tr>
<tr>
<td>00:20</td>
<td>01:01</td>
<td>17:01</td>
</tr>
<tr>
<td>00:36</td>
<td>01:19</td>
<td>17:19</td>
</tr>
<tr>
<td>00:49</td>
<td>01:21</td>
<td>17:21</td>
</tr>
<tr>
<td>00:49</td>
<td>01:30</td>
<td>17:30</td>
</tr>
<tr>
<td>00:56</td>
<td>01:37</td>
<td>17:37</td>
</tr>
<tr>
<td>01:34</td>
<td>02:15</td>
<td>18:15</td>
</tr>
<tr>
<td>01:41</td>
<td>02:22</td>
<td>18:22</td>
</tr>
<tr>
<td>02:49</td>
<td>03:30</td>
<td>19:30</td>
</tr>
<tr>
<td>03:47</td>
<td>04:34</td>
<td>20:34</td>
</tr>
</tbody>
</table>

Figure 2. The timeline of disappearance of Malaysia Airlines Flight 370 (MH370) from Wikipedia.
Table 1. The variables and parameters used in the model.

<table>
<thead>
<tr>
<th>Name</th>
<th>Description</th>
<th>Name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>emergency event</td>
<td>( e )</td>
<td>Concepts set</td>
<td>C</td>
</tr>
<tr>
<td>life course</td>
<td>( T )</td>
<td>Positive samples</td>
<td>PS</td>
</tr>
<tr>
<td>Location information</td>
<td>( L )</td>
<td>Spatial information</td>
<td>SI</td>
</tr>
<tr>
<td>starting timestamp</td>
<td>( ST )</td>
<td>ending timestamp</td>
<td>ET</td>
</tr>
<tr>
<td>participator</td>
<td>( P )</td>
<td>witness</td>
<td>W</td>
</tr>
</tbody>
</table>

(1) The data volume of all Weibo users is up to 1 TB every day. The Weibo is still adding about 100 million messages per day. The huge volume of social media data brings the big challenge for processing and analyzing social media data.

(2) Social sensors are nosier and more redundant than real sensors. Different from physical sensors, social sensors are activated by specific events. For example, when a social network user makes a poster about “school shooting in Virginia Tech”, then the user can be considered as a “Virginia school shooting sensor”. On the other hand, social sensors may post incorrect information due to the user did not activate by a school shooting or the user did not see it when they are in Virginia Tech.

(3) The social media data usually has high value. For example, in the criminal investigation systems, the social media data may help the police to find the suspect [33]. In the traffic surveillance system, the social media data can provide the real time condition of the road network. On the other hands, the huge volume of social data brings the challenge for mining the value from the social media data. The phenomenon of “high volume, low value” from the big data area also exists in the social media data.

(4) The social media device is with fast data in/out. The velocity of collecting social media data is faster than that of processing and analyzing them. The high velocity of social media devices brings the big challenges for processing and analyzing social media data.

III. PROBLEM FORMULATION

In this section, the problem formulation of the proposed work is given. The basic elements of the proposed 5W model are given in the first aspect. The second aspect gives basic definitions of the proposed model.

A. The Problem Formulation

What is a good description for an urban emergency event? Let’s see one example from Wikipedia\(^7\), which is shown in Fig. 2. In Fig. 2, the description of Malaysia Airlines Flight 370 (MH370) is structured as follows. First, the timeline of MH370 is presented. In each timestamp, what was happened in the emergency event is given. Second, the spatial information of some timeline is given, which can be seen in the fifth and ninth line of the timeline of MH370.

Analogously, if the temporal and spatial information of an urban emergency event is provided, it will be very helpful for detecting and resisting it. Thus, inspired by examples from Wikipedia, the description of an urban emergency event should include the following elements.

(1) The temporal information (When). The occurrence time and timeline of an urban emergency event is necessary to describe it.

(2) The spatial information (Where). The happened place is necessary to describe it.

(3) The semantic information (What). The semantic information such as what happened is necessary to describe an urban emergency event.

(4) The person information (Who). The participator and witness are needed to describe an urban emergency event. It is noted that the person information is not essential. For example, a huge fog or air pollution does not have particular participators.

(5) The reason information (Why). The reason of an urban emergency event is needed to describe it. Similar to the person information, the reason information is not essential. For example, it is difficult to give a clear reason of huge air pollution.

In this paper, Weibo is used for data collection since it is the biggest social media platform in China. Weibo has more than 500 million users and about 50 million users are active users. Besides, Weibo users usually provide the useful information, which can be used for extracting the related 5W element of the proposed model.

B. Basic Definitions

An event is something that happens at some specific time, and often some specific places [31]. In the proposed 5W model, the time and place of an urban emergency event can always be identified since some messages sensing by social sensors have an exact timestamp and location. An urban emergency event is defined as follows.

**Definition 1. Urban emergency event, \( e \)**

An urban emergency event \( e \) is a tuple \([T, L]\), where \( T \) and \( L \) is the life course and location of \( e \).

**Definition 2. Life course of an urban emergency event, \( T \)**

The life course of an urban emergency event is a time range from the starting timestamp to the ending timestamp.

The life course of an urban emergency event is related to the when element of the proposed 5W model.

**Definition 3. Location of an urban emergency event, \( L \)**

The location of an urban emergency event is a set of location stamps which are involved by the urban emergency event.

Table 1 shows the variables and parameters used in the following discussion.

IV. THE PROPOSED 5W MODEL

In this section, the detail of the proposed 5W model is given. In the first aspect, the basic framework of the proposed model is illustrated. In the next four aspects, the method and technology

\(^7\) www.wikipedia.com
for generating the 5W elements of the proposed model are given.

A. The Overview of the Proposed 5W Model

Crowdsourcing or participatory sensing may be a potential solution solving the description of urban emergency events. The proposed 5W model aims at collecting and analyzing the information from social sensors. The social network can be seen as a sensor receiver. Usually, the social network users can be seen as social sensors. The proposed 5W model is set as a hierarchical data model including three different layers. The different layers of the proposed method are illustrated in Fig. 3.

(1) Social sensors layer. In this layer, the proposed 5W model wants to collect the related data of urban emergency events. For example, if a user makes a message in Weibo about a fire occurrence, then she/he can be seen as a “fire sensor”. The social network such as Weibo can be seen as a sensor receiver. Usually, social media provides API for downloading the real time data.

(2) Crowdsourcing layer. In this layer, basic elements (what, when, where, who, and why) of the proposed 5W model are extracted from the sensing data of the social sensors layer. Knowledge base and positive samples of the urban emergency event are implemented in this layer, which are used for improving the accuracy of this layer.

(3) 5W based description layer. In this layer, the detection and description of the urban emergency event is launched. Of course, the spatial and temporal information of this event is also given. A GIS based description of the detected urban emergency event is shown.

B. Generating the “What” Element

In this section, the method for generating the “what” element of the proposed 5W model is given. For example, if a user posts a message in Weibo about a fire occurrence, the proposed 5W model must detect that real time urban emergency event.

The message sensing by Weibo users is used to detect real time urban emergency events. Usually, Weibo provides API to support searching messages.

Definition 4. Concepts set of an urban emergency event, \( C \)

Concepts set consists core keywords of an urban emergency event, which is denoted as

\[
C = \{ c_1, c_2, \ldots, c_k \}. \tag{1}
\]

Usually, these keywords can be used to search the urban emergency event from social sensors. For example, the searching result of using “着火 (on fire)” is shown in Fig. 4. The accurate of searching results relies on the quality of the query. Thus, the initializing of the concept is operated manually in order to ensure the accurate of detecting real time urban emergency events. For example, initial concepts of the “on fire” urban emergency event are {“着火”, “火灾”, “起火”, “失火”, “走火”}. It is noted that initial concepts are synonyms since social sensors may use different words to post the same urban emergency event.

Definition 5. Positive samples of an urban emergency event, \( PS \)

Positive sample is the set of accurate messages posted by social sensors, which is denoted as

\[
PS = \{ p_{s_1}, p_{s_2}, \ldots, p_{s_{|P|}} \}. \tag{2}
\]

For example, “I see a car crash now”, “A car crash is happening” can be seen as positive samples for detecting real
The positive sample can be used for detecting real time urban emergency events. For example, in Fig. 4(a), a real time fire event is reported by a Weibo user. Based on common sense and our observations on real data, we have three heuristics that serve as the base of selecting positive samples from Weibo search results.

**Heuristic 1:** The Weibo message with the location information is prone to be a positive sample.

Since the location information is an important aspect of an urban emergency event, the message with the location information can be thought as a positive sample. For example, the Weibo message in Fig. 4(d) provides the location information.

**Heuristic 2:** The Weibo message with the check-in information of the user is prone to be a positive sample.

Different from the location information in Weibo message, the check-in information of the user means the spatial information provided by the mobile device. For example, the Weibo allows users upload their location information by mobile device such as iPhone or iPad. The Weibo user in Fig. 4(e) provides the check-in information.

**Heuristic 3:** The Weibo message with the image or video information is prone to be a positive sample.

The image or video information is very important for knowing the real-time situation of an urban emergency event. For example, the Weibo message in Fig. 4(e) provides three images of a real-time fire event. With the help of concepts of an urban emergency event and three heuristics, the positive sample of an emergency event can be got accurately. For example, if we want to detect whether a real-time fire event happen or not, the concepts about “fire” event are used to search related Weibo messages. Three heuristics are used to select positive samples. Of course, the positive sample can not only tell us what happened but also where and when happened. In next sections, the positive sample is used to generate other elements of the proposed 5W model.

**C. Generating the “Where” Element**

Besides detecting what happened in the urban environment, it is needed to reveal the spatial information of the urban emergency event. Usually, the posted message from social sensors can be revealed as the spatial information of the urban emergency event, which is defined as

**Definition 6. Spatial information of an urban emergency event** $e, SI$

The spatial information can be extracted from the posted messages from Weibo, which is denoted as

$$SI = \{s_{i1}, s_{i2}, ..., s_{ip}\}. \tag{3}$$

For example, in Fig. 5, a user may post a message “I see a fire at Huaihai Road”, the posted spatial information “Huaihai Road” can be extracted. In the proposed 5W model, the spatial information detection is based on the Baidu Map$^{10}$, which can detect whether a word is a location name or not.

**Definition 7. GIS information of an urban emergency event** $e, GI$

Different from the spatial information, the GIS information is extracted from the check-in information from Weibo. The GIS information is denoted as a two tuples with the longitude and latitude. The GIS information can be mined from the element of HTML page. For example, in Fig. 5, the GIS information of the message is “geo=121.46, 31.22”.

**Heuristic 4:** The spatial information is prone to be the location of an urban emergency event other than the GIS information

The GIS information is extracted from the check-in information. The check-in information is the location information of the social sensor other than the urban emergency event.

Since the noise of the social sensor, an urban emergency event may have different spatial information extracted from the posted message or GIS information. For example, in Fig. 6, two different users mention two different locations of the same “fire” event. Since they are in the different location, they may see the different condition of that “fire” event. Of course, the uploaded image of that “fire” event is in the different angle. It is important to collect the different location information and image of an urban emergency event. Different GIS information can reflect the spread condition of an urban emergency event.

---

10 Openapi.baidu.com
a robbery is happening at the 5th gate of the Shanghai south station of underground railway.

Figure 7. The illustration of “Who” element.

The translation of the first message is “a robbery is happening at the 5th gate of the Shanghai south station of underground railway”. The translation of the second message is “In the 9th o’clock tonight a robbery is happening at the 5th gate of the Shanghai south station. I call the police and advise the robber to put down his knife.”

For example, in Fig. 6, the GIS information of two users is “North Jiefang Road” and “No. 5 Zhongshan Road”. These two GIS information can reflect that fire is spread to these two roads.

D. Generating the “When” Element

Besides the spatial information, the temporal information is another basic factor of an urban emergency event. Similar to the spatial information, the temporal information can also be extracted from the posted messages. For example, in Fig. 5, the timestamp of the posted message is 12:57 at April 11, 2014.

Definition 8. The starting timestamp of an urban emergency event \( e \), \( ST \)

The starting timestamp means the happened time of an urban emergency event. Of course, the extracted timestamp usually is not the starting timestamp. The extracted timestamp is usually the posted time of the message other than the happen time of the urban emergency event.

Definition 9. The ending timestamp of an urban emergency event \( e \), \( ET \)

The ending timestamp means the finished time of an urban emergency event.

The detection of the starting and ending timestamp is not an easy thing. Intuitively, the extracted timestamp of an urban emergency event can be ranked as the descendant order. The earliest and the latest timestamp can be seen as the starting and the ending timestamp. Unfortunately, since the existence of the noise and redundancy of the social sensor, the intuitive method based on ranking is inappropriate.

Heuristic 5: The valid timestamp should be extracted from the posted message following at least two heuristics of 1, 2, and 3.

Heuristic 6: The valid timestamp should be extracted from the original posted message other than forward messages.

The valid timestamp should be extracted from the messages with spatial information, the GIS information, or uploaded image. The social sensor that provides spatial, GIS, or image information is prone to be a witness of the urban emergency event. Of course, if a message is a forward message other than an original message, the message is not a valid message since the social sensor is not a witness of the urban emergency event.

The starting timestamp can be obtained from the ranking of valid timestamps. Besides the starting timestamp, the proposed 5W model aims at describing the timeline of an urban emergency event. Each timeline should describe a new condition of an urban emergency event.

Heuristic 7: The timeline of an urban emergency event should provide new spatial information or check-in information.

The new spatial information or check-in information can be seen as the spread of the urban emergency event. For example, in Fig. 6, these two users give the check-in information “North Jiefang Road” and “No. 5 Zhongshan Road”. These two messages can be seen as the timeline of the “fire” emergency event.

E. Generating the “Why” Element

Unlike the “what”, “where”, and “when” elements, the “who” element is an alternative element of an urban emergency event. For example, a “hijack” event may have the participator. A “fire” event may only have the witness.

Definition 10. The participator of an urban emergency event \( e \), \( P \)

The participator means the person who engages the urban emergency event. For example, in Fig. 7, the people in the uploaded image are participators of the “hijack” and “fight” events. In the proposed model, DPM [32] model is used for detecting the person in the uploaded image. The participator of the urban emergency event is detected in the red rectangle using DPM model.

Definition 11. The witness of an urban emergency event \( e \), \( W \)

The witness means the person who sees the urban emergency event. For example, in Fig. 7, these two Weibo users who upload the image are witnesses of the “hijack” and “fight” events.
Table 2. The timeline of the “fire” event. (The last row is the name of Weibo users, which is Chinese.)

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Where (Spatial)</td>
<td>No. 5 Zhongshan Road</td>
<td>Xiaoma Zhan</td>
<td>Beijing Road</td>
<td>Xihu Road</td>
<td>Beijing Road</td>
<td>Beijing Road</td>
<td>Xiaomazhan</td>
<td>Danan Road</td>
<td>Beijing Road</td>
<td>East Huifu Road</td>
</tr>
<tr>
<td>Where (check-in)</td>
<td>North Jiefang Road</td>
<td>North Haizhu Road</td>
<td>No. 4 Zhongshan Road</td>
<td>Xinmin Road</td>
<td>No. 6 Zhongshan Road</td>
<td>Xihu Road</td>
<td>Huifu Road</td>
<td>Xihu Road</td>
<td>East Huifu Road</td>
<td></td>
</tr>
<tr>
<td>Who (Witness)</td>
<td>K仔牛肉拉肠</td>
<td>婷 nicky</td>
<td>Vincent</td>
<td>迟半拍之愚</td>
<td>珠水微帆</td>
<td>色水</td>
<td>汪子零时工</td>
<td>乌托伊甸</td>
<td>海珠顽石</td>
<td>温超00</td>
</tr>
</tbody>
</table>

Figure 9. The illustration of GIS information of witnesses.

Heuristic 8: The witness usually provides the check-in information and upload image or video information.

The person who provides the check-in information and uploads the image of an urban emergency event is prone to be the witness. They can provide the real condition of an urban emergency event.

F. Generating the “Why” Element

The upload message by social sensors may reveal the reason of an urban emergency event. For example, a Weibo user may post a message “OMG, I saw a car crash a man who crossed the red light”. The message posted by witness or participator can be used to investigate the potential reason of an urban emergency event.

Heuristic 9: The message forwarded by large number of users is prone to mention the reason of an urban emergency event.

The message with a high forward number usually reveals the reason of an urban emergency event. For example, in Fig. 8, these two Weibo users reveal the reason of two urban emergency events. According to the heuristic 9, these two messages are with a high forward number.

Heuristic 10: The message posted by official Weibo users is prone to mention the reason of an urban emergency event.

Different from Weibo users, the official Weibo user11 is an information provider of an organization such as local government, new site, and so on. The trust of the official Weibo user is usually high than real user.

V. CASE STUDIES

In this section, two case studies about “fire” emergency event and “hijack” emergency event are given using the proposed 5W model. The proposed 5W model aims to describing the urban emergency event using Weibo social media.

A. The “Fire” Urban Emergency Event

The “fire” emergency event is a major damage to the urban especially the big city. For example, in Shanghai, there are about 796 “fire” alarms from May 16 to May 26 in 2014. Thus, it is important to detect and analyze the real time “fire” event. For example, the location of an real time “fire” event can be pushed to the nearby Weibo user. The detected image of a real time “fire” event can be transferred to the firemen to help them know the condition of that “fire”. The extracted message from social sensors can help the firemen to know the reason of the “fire”.

The case event. We select a “fire” event happened in 15:20 at “Beijing Road” on May 29 in Guangzhou13. The proposed 5W model is used to detect and describe that event.

What. Five concepts about the “fire” are used to search in Weibo from 15:00 to 20:00. The search location is set as Guangzhou14. Totally, 246 messages are returned. According to heuristics 1, 2, and 3, 21 messages provide location information, check-in information, and image. These 21 messages can be used to detect another four elements (When, Where, Who, and Why).

Where. According to the heuristic 4, the spatial information extracted from these 21 messages are “No. 5 Zhongshan Road”, “Beijing Road”, “Xihu Road”, “Xiaomazhan”, and “Jiaoyu Road”. The spatial information “Beijing Road” appears 12 times of these 21 messages, which is the real happened location of that “fire” event. The GIS information of some messages is shown in Fig. 9. In Fig. 9, five users are annotated in the map according to their GIS information. The location of the “fire” is annotated by the red circle. The uploaded images of these five users are shown along with the annotated user. The uploaded images are in the different angle of the “fire” event and show the different condition of the “fire” event.

When. It is noted that the first appearing time of these 21 messages is 15:24, which is only 4 minutes later than the appearing time of that “fire” event. The starting timestamp of that “fire” event using the proposed 5W model is 15:24.

11 In Weibo, the official user is usually noted as “V”
12 The biggest city with about 3 million people in China
13 A big city with about 15 million people in China
14 Weibo provides the API for searching in the different location
The fire at 38 Beijing Road is caused by the damage of the electric wire.

**Figure 10. The reason of the “fire” event.**
(The translation of the first message is “The fire at 38 Beijing Road is caused by the damage of the electric wire.”)

**Where.** The spatial and check-in information extracted from two messages are “South Railway Station Shanghai”.

**When.** It is noted that the first appearing time of two messages is 21:31, which is only 11 minutes later than the appearing time of that “hijack” event. The starting timestamp of that “hijack” event using the proposed 5W model is 21:31. The ending timestamp is 1:40 at April 29.

**Who.** In the “hijack” emergency event, besides the witness, participator is also important. The participator is denoted in the red rectangle using DPM model, which is shown in Fig. 11. The young man is the criminal of the “hijack”. The old man is also the participator of the “hijack”. In fact, he saves the girl and captures the knife from the hand of the criminal.

**Why.** A message posted by Xinmin News Weibo\(^{15}\) which is forwarded by 221 times reveals the reason of the “hijack” event. The criminal say that he does not eat anything and is very hungry for two days.

VI. CONCLUSION

Crowdsourcing is a process of acquisition, integration, and analysis of big and heterogeneous data generated by a diversity of sources in urban spaces, such as sensors, devices, vehicles, buildings, and human. The content from social media often includes references to urban emergency events occurring at, or affecting specific locations. In this paper, in order to detect and describe the real time urban emergency event, the 5W (What, Where, When, Who, and Why) model has been proposed. Firstly, users of social media are set as the target of crowd sourcing. Secondly, the spatial and temporal information from the social media are extracted to detect the real time event. Thirdly, a GIS based annotation of the detected urban emergency event is shown. The proposed method is evaluated with extensive case studies based on real urban emergency events. The results show the accuracy and efficiency of the proposed method. The proposed 5W model includes what, where, when, who, and why elements to detect and analyze the urban emergency event. The proposed model is based on crowdsourcing, which uses the real time nature of Weibo users. The proposed model is applied into the emergency management field, which can provide useful information to analyze and resist urban emergency events. Case studies on real data sets show the proposed model has good performance and high effectiveness in the analysis and detection of urban emergency events.

In the future work, we will extend our work on using the

\(^{15}\) An official Weibo of Xinmin Post (the largest newspaper in Shanghai)
crowdsourcing-based technology for other applications such as traffic analysis. Moreover, we may consider to do a comparison of same detection process between the social media (i.e., weibo, etc.) in China, and other similar ones (e.g., facebook, twitter, etc.) used by other countries. Perhaps we will find something interesting and different from the results that you obtained.

REFERENCES


Yunhuai Liu is a professor in the third research institute of ministry of public security, China. He received the PhD degrees from Hong Kong University of Science and Technology (HKUST) in 2008. His main research interests include wireless sensor networks, pervasive computing, and wireless network. He has authored or co-authored more than 50 publications and his publications have appeared in IEEE Trans. on Parallel and Distributed Systems, IEEE Journal of Selected Areas in Communications, IEEE Trans. on Mobile Computing, IEEE Trans. on Vehicular Technology etc.

Neil Y. Yen received Doctorate degree in human sciences at Waseda University, Totsukamachi, Japan, and in engineering at Tamkang University, New Taipei, Taiwan, in 2012. His Doctorate degree at Waseda University was funded by Japan Society for the Promotion of Science (JSPS) under RONPAKU program. He has joined the University of Aizu, Aizuwakamatsu, Japan, as an Associate Professor since April 2012. Dr. Yen has been engaged extensively in an interdisciplinary field of research, where the themes are in the scope of big data science, computational intelligence, and human-centered computing. Dr. Yen has been actively involved in the research community by serving as Guest Editor, Associate Editor, and Reviewer for international referred journals and as Organizer/Chair of ACM/IEEE-sponsored conferences, workshops, and special sessions. He is now a Member of IEEE Computer Society, IEEE System, Man, and Cybernetics Society, and Technical Committee of Awareness Computing (IEEE SMCS).

Lin Mei received his PhD. degree from Xian Jiaotong University, China. He is currently working in the third research institute of ministry of public security, China. He is the Dean Professor of the department of internet of things.

Xiangfeng Luo is a professor in the School of Computers, Shanghai University, China. Currently, he is a visiting professor in Purdue University, USA. His main research interests include Web Wisdom, Cognitive Informatics, and Text Understanding. He has authored or co-authored more than 50 publications and his publications have appeared in IEEE Trans. on Systems, Man, and Cybernetics-Part C, IEEE Trans. on Automation Science and Engineering, IEEE Trans. on Learning Technology, etc. He has served as the Guest Editor of ACM Transactions on Intelligent Systems and Technology. Dr. Luo has also served on the committees of a number of conferences/workshops, including Program Co-chair of ICWL 2010 (Shanghai), WISM 2012 (Chengdu), CTUW2011 (Sydney) and more than 40 PC members of conferences and workshops.

Xiao Wei received his BS degree in computer application technology from Shandong University, China. Now he is currently working toward the PhD. degree at Shanghai University. Con currently, he is a visiting scholar at City University of Hong Kong. His main research interests include Web Mining, Semantic Search, and E-learning.

Chuanping Hu received his PhD. degree from Tongji University, China. He is the Dean Professor of the third research institute of ministry of public security.