2581-4575

International Journal For Recent Developments in Science & Technology

Crossref

A Peer Reviewed Research Journal

RECOGNITION OF WEB RESOURCES USING DIFFERENT STATES OF EMERGENCY EVENTS

[1]PULI PRIYANKA , [2]N.SRINIVASA RAO

[1]Department of MCA, SKBR PG College, Amalapuram

[2]Assistant Professor, Department of MCA, SKBR PG College, Amalapuram

ABSTRACT. An emergency event is a sudden, urgent, usually unexpected incident or occurrence that requires an immediate reaction or assistance for emergency situations, which plays an increasingly important role in the global economy and in our daily lives. Recently, the web is becoming an important event information provider and repository due to its real-time, open, and dynamic features. In this paper, web resources based states detecting algorithm of an event is developed in order to let the people know of an emergency event clearly and help the social group or government process the emergency events effectively. The relationship between web and emergency events is first introduced, which is the foundation of using web resources to detect the state of emergency events imaged on the web. Second, five temporal features of emergency events are developed to provide the basis for state detection. Moreover, the outbreak power and the fluctuation power are presented to integrate the above temporal features for measuring the different states of an emergency event. Using these two powers, an automatic state detecting algorithm for emergency events is proposed. In addition, heuristic rules for detecting the states of emergency event on the web are discussed. Our evaluations using real-world data sets demonstrate the utility of the proposed algorithm, in terms of performance and effectiveness in the analysis of emergency events.

KEY WORDS: Events; emergency management; automatic states detection; web mining.

I.INTRODUCTION

No country or community or person is immune from emergency events [1]. An emergency event is a sudden, urgent, usually unexpected incident or occurrence that requires an immediate reaction or assistance for the emergency situation faced by social group (e.g., the corporations) or recipients of public assistance [2]. For example, in 2001, the “September 11 attack” caused nearly 3,000 deaths and its global impact continues to this day. In 2003, “Severe Acute Respiratory Syndrome (SARS)” spread from Hong Kong to infect individuals in 37 countries, which resulted in 8,422 cases and 916 deaths worldwide (10.9% fatality) according to the World Health Organization [3]. In 2008, the “Great Sichuan Earthquake” in China was a deadly earthquake killing an estimated 68,000 people. Therefore, how to prepare for, respond to, and recover from such emergency events is important. An apparent choice for processing an emergency event is to analyze its related information. Due to the popularity of the web, most emergency events are reported in the form of web resources. Especially, with the development of the social media, people can obtain/post more and more information about emergency events from/to the web in (near) real-time. In our view, using related web

International Journal For Recent Developments in Science & Technology
2581-4575
Crossref
A Peer Reviewed Research Journal

resources to analyze emergency events has three advantages.

(1) The web can provide related information immediately after an emergency event happens and keep updating the information in near real-time, which is a key requirement for keeping track of the sudden changing nature of an emergency event. Traditional media such as newspapers and magazines cannot report an emergency event immediately. In contrast, the web can deal with this issue properly. For example, the time at which Chinese web users came to know about the September 11 attacks is only 5 minutes later than the president of the United States [4]. Recently, with the rapid development of social media sites such as Twitter and Facebook, the web has become an important events information provider [5].

(2) The quick and easy spread of information through the web can provide a comprehensive perspective of emergency events. Traditional media usually provides content that has beencorroborated as well as commentary such as expert opinions to the public. In other words, traditional media usually tries to investigate and reach conclusions related to different aspects of emergency events rather than simply reporting them. Different from the traditional media, the web can provide different perspectives of an emergency event. Different users can give their own opinions about an emergency event. The open feature of the web ensures that users know about the different aspects of an emergency event including the different opinions/information about the event.

(3) The dynamic feature of web information can keep up with the evolution of emergency events. Of course, an emergency event is not static and the information about it may change with time. In some studies [6], the changing nature of an event is named as "event evolution." The event evolution generates a large volume of temporal data. This large amount of information should have an appropriate representation and accessibility to be useful. Besides the large volume of data, the information regarding an emergency event updates quickly. Different from our formal work [26-30] (which is given in the related work), in this paper, we introduce a new web mining task-- -states detection of emergency events imaged on the web.

Given an emergency event, the related web pages can be found, for example, web news, blogs, and forums. Based on the semantics of these web pages, the temporal features of an event can be identified. And then, the different states can be detected. Besides detecting the boundaries of different states, we also output the outbreak and fluctuation feature of each state. Herein, we divide the states of an emergency event into five stages, i.e., latent state, outbreak state, decline state, transition state, and fluctuation state. The detailed definitions of these states are defined in the following sections.

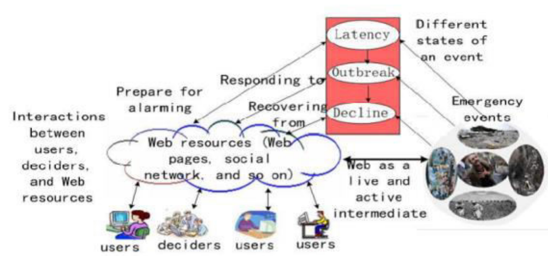

Fig. 1. The illustration of different periods of an event imaged on the web

Fig. 1 illustrates the basic states of an emergency event, in which, the web is seen as a live and active media which provides the interaction interface among the web users, the information regarding an emergency event imaged on the web, and the three basic states (i.e., latency state, outbreak state, and decline state). In the latent state, the number of related web pages is low; users usually do some prevention in this state. In the outbreak state, people usually try to avoid the emergency event because it may cause severe destruction to people.

Social groups or governments should respond to and lessen the effect of an emergency event. In the decline state, people usually undertake some work to recover from the emergency event. An intuitive method for detecting different states of an event is using algorithms from time series data mining, such as segmentation technologies [7]. However, the emergency event states detection algorithm should consider the sudden changes especially the dynamic feature. Unfortunately, current time series– based segmentation technologies do not focus on this issue. This is the gap that this paper seeks to address. In this paper, an algorithm is proposed for the states detection of emergency events imaged on the web. First, the related resources including web pages, keywords of an emergency event are collected by web search engines. Second, the outbreak power and the fluctuation power of an emergency event in timestamp t are computed. Based on the various temporal values, different states of an emergency event are detected. The major contributions of this paper are summarized as follows

(1) The temporal features of the emergency event imaged on the web are defined, which integrate the number of increased web pages, the number of increased keywords, the distribution of keywords, and the relationships between keywords. In addition, some heuristic rules are given to detect the different states of an emergency event.

(2) Different states of an emergency event have been proposed including latent, outbreak, decline, transition, and fluctuation states. Through these states, people can prepare, respond, and recover from an emergency event. Two factors including outbreak power and fluctuation power are presented to automatically detect different states of an emergency event.

(3) Using these two powers, an emergency event states detection algorithm is developed using web resources. Given an emergency event and its related web pages, this algorithm determines the different states of the event. The resources from web users can provide the real-time and semantic information related to emergency events.

II. RELATED WORK

The proposed states detection problem is similar to the research on Topic Detection and Tracking (TDT). Various methods have been proposed to manage news stories, spot news events, and track the process of events [8, 9, 10, 11, 24, 25]. Typically, the TDT approach generates a hierarchical structure of an event, which aims at clustering related news into it. Overall, TDT technologies attempt to detect or cluster news stories into these events, without focusing on or interpreting the sudden, urgent, and unexpected features of emergency events [12, 36]. Since event evolution technologies are similar to the emergency event states detection, we will list some related work. Event evolution proposed by Makkonen [13] is a subtopic of topic detection and tracking. In his study, two conclusions are reached:

2581-4575
International Journal For Recent Developments in Science & Technology
Crossref
A Peer Reviewed Research Journal

(1) a seminal event may lead to several other events; and

(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 ontologies to measure the similarity of events. However, these ontologies are difficult to get, which makes the work difficult to be used directly. Mei [14] investigated theme evolution, which is similar to event evolution. He proposed a temporal pattern discovery technique on the basis of timestamps of the text streams. The theme of each interval is identified, and the evolution of theme between successive intervals is extracted. However, the proposed technique does not consider the different states of an event, which may have an impact on its result. Wei [15] 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 our work handles the different states of emergency events imaged on the web.

Later, Yang [16] aimed to 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 [17] proposed a 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.

Earle et al. [31] assessed the use of Twitter for earthquake detection and mapping the affected area by using the tweets generated after the 30 March 2009 Morgan Hill, California, earthquake. Similarly, Crooks et al. [32] analyzed the spatial and temporal features of a 5.8 magnitude earthquake which occurred on the East Coast of the United States (US) on August 23, 2011 using Twitter messages. These messages are regarded as a hybrid form of a distributed sensor system that allows for the identification and localization of the impact area of the earthquake. Sakaki et al. [33] investigated earthquake-related messages on Twitter in real-time and proposed an algorithm to detect an event using tweets. In order to investigate the dynamics and evolution of online communities in social networks in response to emergency events, Liu et al. [34] collected three datasets from Twitter shortly before and after the 2011 earthquake and tsunami in Japan. Esposito et al. survey the available literature and practice on crisis information systems [40].

In order to detect and describe the real time urban emergency event, the 5W (What, Where, When, Who, and Why) model is proposed by Xu [26, 30]. This is somewhat similar to the concept underlying digital forensic investigations [37, 38]. Xuan [27] proposed a framework to identify the different underlying levels of semantic uncertainty in terms of web events, and then utilize these for web page recommendations. The basic idea is to consider a web event as a system composed of different keywords, and the uncertainty of this keyword

International Journal For Recent Developments in Science & Technology
2581-4575
Crossref
A Peer Reviewed Research Journal

system is related to the uncertainty of the particular web event. Liu [28] explored a Markov random field based method for discovering the core semantics of event.

This approach makes collaborative semantics computation for learning association relationship distribution and makes information gradient computation for discovering k redundancy-free texts as the core semantics of event. A crowdsourcing based burst computation algorithm of an urban emergency event is developed in order to convey information about the event clearly and to help particular social groups or governments to process events effectively [29]. Overall, the above methods have been shown to have good performance for general events other than emergency events. The emergency events possess dynamic, real-time, multi-states, sudden, and urgent features. In this paper, we consider the multi-states of an emergency event imaged on the web.

III. STATES DETECTION ALGORITHM

In this section, we discuss the proposed computation algorithm for detecting different states of an emergency event. Based on the five temporal features discussed in Section 3, the proposed computation algorithm is divided into three steps:

(1) Outbreak power computation. We compute the outbreak power, which reflects the influence degree of an emergency event to the society.

(2) Fluctuation power computation. We compute the fluctuation power, which reflects the change rate of an emergency event.

(3) States detection. Based on the outbreak power and fluctuation power, we detect the different states of an emergency event imaged on the web. Fig. 2 illustrates an overview of the proposed states detection algorithm.

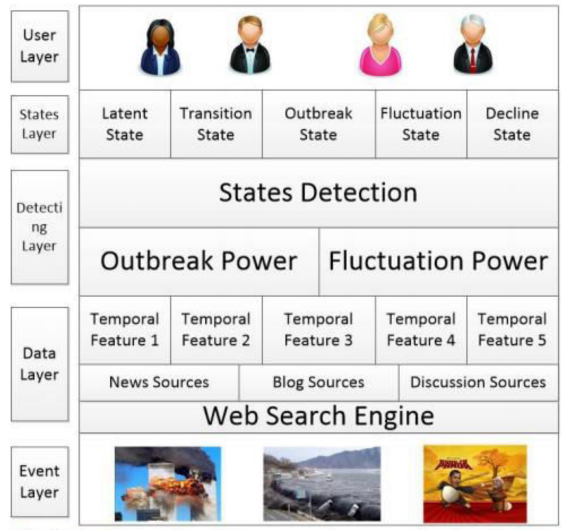

Fig. 2. The illustration of the proposed computation algorithm for detecting different states of emergency event

3.1 Computing Outbreak Power

After giving the basic temporal features of emergency events, we define outbreak power as follows. The detailed algorithm can be found in our earlier work.

Definition 1. The representative power of keywords,The representative power of keyword k is the probability of k to represent the event e correctly.

Definition 2. The confidence of web pages, d( ) The confidence of a web page d is the expected representative power of keywords provided by d. Based on common sense and the observations on real data, we have six basic corollaries which serve as the basis for our computation algorithm

3.2 Computing Fluctuation Power

After computing the outbreak power of an event, we define the fluctuation power as follows.

Definition 3. Fluctuation power, For an emergency event e, the fluctuation power is the change rate of web pages from i to j against the antecedent time interval. Fluctuation power is different from outbreak power, which is relevant to the antecedent time interval. Before we propose the computing algorithm for fluctuation power, we give an important definition as follows. Definition 4. The change rate of web pages, The change rate of a web page d is the dissimilarity of it against the web pages from the antecedent time interval. Similar to the outbreak power, we also give a heuristic rule to compute the fluctuation power. Heuristic rule 5. If the similarity bipartite graph of web pages between the adjacent time intervals is a complete graph, then the value of the fluctuation power is the lowest.

3.3 States Smoothing and Segmentation

The temporal feature in our work is obtained from web search engines. For example, the number of increased web pages is the page counts received from Google or other web search engines. This data may not be correct since the page counts given by web search engines are just estimates instead of a real number. Thus, the noise should be omitted. We use piecewise aggregate approximation (PAA) from [22] to smooth the temporal feature and take the properties of the emergency event into consideration. F the smooth result of the outbreak power of the event “Japan nuclear crisis”. The x-axis represents the date of the event, and the y-axis represents the outbreak power value. Besides the smoothing, we discuss below the method for achieving states segmentation of an event. The temporal segmentation problem has been widely studied within various disciplines. It is worth noting that the segmentation algorithm is based on the relative value of the outbreak power and fluctuation power. For example, if the outbreak power is between 0.366 and 0.396, then the segmentation algorithm can also perform well. The algorithm is based on the variation in the different stages other than the value difference.

3.4 States Detection Similar to computing the outbreak and fluctuation power, we also provide some heuristic rules to detect the states. Based on the definitions of latent state and decline state, we can derive Heuristic rule 6.

IV. RESULTS

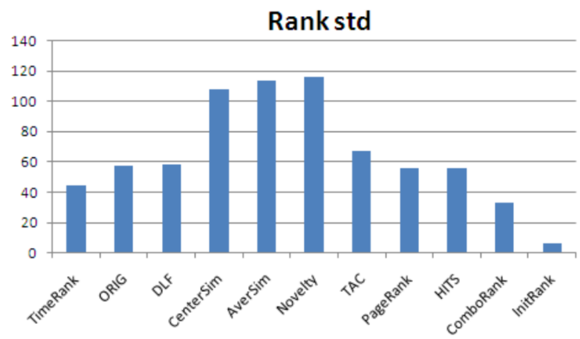

Fig. 3: STANDARD DEVIATION OF RANK

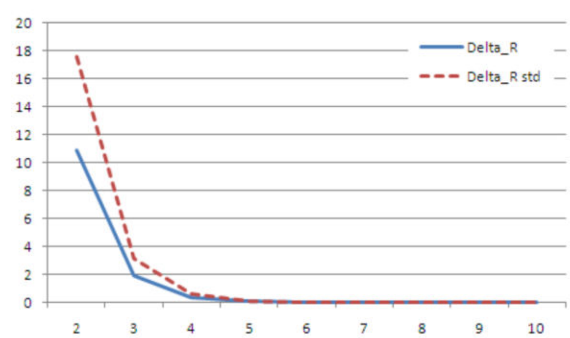

Fig. 4: CONVERGENCE OF INITRANK

V. CONCLUSION

One could never be fully prepared for an emergency event, and all countries, communities, and people are vulnerable to such events (e.g. terrorist attacks and natural disasters such as bush fire). A prudent choice for processing an emergency event is to analyze its related information. Due to the popularity of the web and the pervasiveness of Internet-connected consumer devices (e.g. Android and iOS devices), most emergency events are reported in the form of web resources (e.g. twitter and other social media feeds). In this paper, we proposed a novel algorithm to detect the different states of emergency events reported on the web. First, the related resources including web pages, keywords of an emergency event are collected using web search engines. Second, the outbreak power and the fluctuation power of an emergency event at different timestamps are computed. Based on the various temporal values, different states of an emergency event are detected. Future work will include extending our approach to other applications such as hot news analysis with the aims of further validation and refinement (if necessary).

VI. REFERENCES

[1] D. Haddow, A. Bullock, and P. Coppola. Introduction to Emergency Management, 2010.

[2]http://definitions.uslegal.com/e/emergency-event/, 2012.

[3] http://www.who.int/csr/sars/en/, 2012.

[4]http://en.wikipedia.org/wiki/September_11_attacks, 2012.

[5] F. Jess, M. Trina, T. Jarrod, et al. Riskr: A lowtechnological Web2.0 disaster service to monitor and share information. In the Proceedings of 15th International Conference on Network-Based Information Systems, pp. 1-8, 2012.

[6] J. Makkonen. Investigation on event evolution in TDT. In Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language, PP.43-48, 2003.

[7] C. Yang, X. Shi, and C. Wei. Discovering Event Evolution Graphs from News Corpora. IEEE Trans. on Systems, Man and Cybernetics—Part A: 39(4):850-863, 2009.

[8] J. Abonyi, B. Feil, S. Nemeth, and P. Arva. Modified Gath– Geva clustering for fuzzing segmentation of multivariate timeseries. Fuzzy Sets and Systems, Data Mining, Special Issue 149:39-56, 2005.

[9] X. Wu, Y. Lu, Q. Peng, and C. Ngo. Mining Event Structures from Web Videos. IEEE Multimedia, 18(1):38-51, 2011.

[10] Q. He, K. Chang, E. Lim and A. Banerjee. Keep It Simple with Time: A Reexamination of Probabilistic Topic Detection Models. IEEE Transactions on Pattern Analysis and Machine Intelligence, 32(10):1795-1808, 2010.

[11] C. Sung and T. Kim Collaborative Modeling Process for Development of Domain-Specific Discrete Event Simulation Systems. IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews, 42(4):532-546, 2012.

[12] J. Allan, G. Carbonell, G. Doddington, J. Yamron, and Y. Yang. Topic Detection and Tracking Pilot Study Final Report. In

Proceedings of the Broadcast News Transcription and Understanding Workshop, 1998.

[13] J. Allan. Topic Detection and Tracking: Event-Based Information Organization. Norwell, MA: Kluwer, 2000.

[14] Q. Mei and C. Zhai. Discovering evolutionary theme patterns from text: An exploration of temporal text mining. In Proceedings of the eleventh ACM SIGKDD international conference on Knowledge discovery in data mining, pp.198-207, 2005.

[15] C. Wei and Y. Chang. Discovering event evolution patterns from document sequences. IEEE Transactions