International Journal of Computer Trends and Technology
ISSN: 2231-2803 / https://doi.org/10.14445/22312803/IJCTT-V68I3P104
Volume 68 Issue 3, 16-20 March 2020

Original Article

Spatial Mining of Urban Emergency Events using Crowdsourcing

Sahana H R[1], M V Vijaya Kumar[2]

[1] *PG Student, Department of Information science and Engineering, Dr Ambedkar Institute of Technology, Bangalore, India*
[2] *Professor & Head, Department of Information science and Engineering, Dr Ambedkar Institute of Technology, Bangalore, India*

Received Date: 30 January 2020
Revised Date: 12 March 2020
Accepted Date: 15 March 2020

Abstract - *With the advancements of knowledge or information communication technologies, it's critical to enhance the efficiency and accuracy of emergency management systems through modern processing techniques. The literature has witnessed the tremendous technical advancements in Sensor Networks, Internet/Web of Things, cloud computing, Mobile/Embedded computing, Spatial/Temporal processing, Big Data, and these technologies have provided new opportunities and solutions to emergency management. The GIS (Geographic Information System) models and simulation capabilities are wont to exercise response and recovery plans during non-disaster times. They assist the decision-makers to understand near real-time possibilities during an occasion. Here the authors have proposed the Spatial Mining of Urban Emergency Events using Crowdsourcing. At First, the basic definitions of the proposed method are given, and secondly, the positive samples are selected to mining the spatial information of urban emergency events. Next, the location and GIS information is extracted from positive samples. The important spatial information is decided supported address and GIS information.*

Keywords - *Crowdsourcing, Geographic Information System, social media, urban computing.*

I. INTRODUCTION

For unexpected events like natural calamities and any quiet accidents, there's a requirement of developing strategies and establishing operations to decrease the potential impact. That is the most important aim of Emergency management. By quick response and rescue, one can save human lives from secondary disasters and enhance the steadiness of communities after disasters. Emergency management involves four stages: Planning and Mitigation, Preparedness, Response and Recovery. Geospatial applications, including GIS, are extensively utilized in each stage of emergency management. Decision-makers can utilize this geospatial information to develop planning and mitigation strategies. GIS models and simulation capabilities are applied in order to exercise response and recovery plans during non-disaster times. They assist the decision-makers in understanding the near real-time possibilities.

The recent evolution of the web/Internet has permitted an unusual increase in content created by the non-specialist user through platforms like WhatsApp, Twitter etc. Web users can provide their geographic information through social media. These are now commonly mentioned as Volunteered Geographic Information (VGI), having an enormous potential to interact with citizens in place-based issues and supply timely, significant and cost-effective sources for Geographer and the other spatially related fields of research and management.

Data mining is the process of extracting data from large data sets. Spatial data mining is the application of information or data mining to spatial models. In spatial data processing or mining, the analysts use geographical or spatial information to supply business intelligence. This needs specific resources and techniques to urge the geographical data into the relevant and useful formats.

Crowdsourcing involves acquiring information or opinions from a large group of people. Crowdsourcing may be a sourcing model wherein the individuals or organizations obtain goods and services, including ideas and finances, from an outsized, relatively open and sometimes rapidly-evolving group of internet users. It divides work between participants to realize a cumulative result. The word Crowdsourcing itself may be a combination of crowd and outsourcing. As a mode of sourcing, Crowdsourcing existed before the digital age, i.e. offline.

II. RELATED WORK

Over the years, many types of research have been conducted to deal with performance improvements, issues and challenges in unplanned mobile networks. Earle P, Guy M, Buckmaster R, Ostrum C, Horvath S, and Vaughan A in their Seismological Research Letter," OMG earthquake! Can Twitter improve earthquake response?"[7] the authors studied earthquake-related messages (tweets) and supplied geo-located earthquake detections, and rough maps of the corresponding felt areas. In regions with active Twitter users, the frequency of "earthquake" tweets often rises above the noise level within seconds after a felt earthquake.

Examining these tweets shows it's possible (in some cases) to roughly estimate the felt area of an earthquake using the geolocated Twitter responses. The tweets also provide (very) short first-impression narratives from people that experienced the shaking.

The main advantage of Twitter is speed, especially in sparsely instrumented areas. A Twitter-based system could potentially provide a fast notification that a possible event has occurred, which seismographically derived information will follow. The drawback of this approach is the lack of quantitative information it produces like epicentre, magnitude, and strong-motion recordings. Without quantitative data, prioritization of response measures, including building and infrastructure inspection, isn't possible.

T. Sakaki, M. Okazaki, and Y. Matsuo. In their work,” Tweet Analysis for Real-time Event Detection and Earthquake Reporting System Development”[12], To detect a target event, authors devised a classifier of tweets based on features such as the keywords in a tweet, the number of words and their context. Subsequently, they produced a probabilistic spatiotemporal model for the target event that can find the centre and the trajectory of the event location.

Researchers B. Longueville, R. Smith, and G. Luraschi, in their paper,” OMG, from here I can see the flames, a use case of mining location-based social networks to acquire Spatio-temporal data on forest fires.”[13] They observed that 1)Tweets might contain valuable Spatio-temporal data (and even more in the future). 2)The discrimination between primary and secondary information is uneasy but crucial. 3)Tweets contain many links that can be a good starting point for crawling. 4)Twitter offers a timeline spontaneously for event-related information. The authors used spectral clustering to obtain partitions that contain venues with similar activity profiles. They computed the difference between the activity profile of the partition and the activity profile of the regional level for every cluster and visualized all the differences.

III. SYSTEM DESIGN

System Design is defined as “The process of applying various techniques and principles for the purpose of defining a process or a system in sufficient detail to permit its physical realization”. Various design features are followed in order to develop the system. The design specification describes the features of the system, the components or elements of the system and their appearance to end-users.

A. Input Design

The input Design is the process of converting the user-oriented inputs into the computer-based format. The goal of designing input data is to make automation as easy and free from errors as possible. Providing a good input design for the application, easy data input and selection features are adopted. The input design requirements such as user-friendliness, consistent format and interactive dialogue for giving the right message and help for the user at the right time are also considered for the development of the project. Input design is a part of overall system design, which requires very careful attention. Often the collection of input data is the most expensive part of the system, which needs to be a route through a number of modules. It is the point where the user is ready to send the data to the destination machine along with a known IP address; if the IP address is unknown, then it may be prone to error.

B. Output Design

Quality output is one that meets the requirements of the end-user and presents the information clearly. In any system, results of processing are communicated to the users and to other systems through outputs. It is the most important and direct source of information to the user. Efficient and intelligent output improves the system's relationship with the source and destination machine. Outputs from computers are required primarily to get the same packet that the user has sent instead of a corrupted packet and spoofed packets. They are also used to provide a permanent copy of these results for later consultation.

C. System Development Methodology

The system development method is a process through which a product will get completed or a product gets rid of any problem. The software development process is described as a number of phases, procedures and steps that gives the complete software. It follows a series of steps that are used for product progress. The development method followed in this project is the waterfall model.

a) Model Phases

The waterfall model is a sequential software development process in which progress is seen as flowing steadily downwards (like a waterfall) through the phases of Requirement initiation, Analysis, Design, Implementation, Testing and maintenance. The reason why the waterfall model is used is a)Clear project objectives. b)Stable project requirements. c)Progress of the system is measurable. d)Strict sign-off requirements d)Helps you to be perfect. e)Logic of software development is clearly understood. f)Production of a formal specification. g)Better resource allocation h)Improves quality. i)The emphasis on requirements and design before writing a single line of code ensures minimal wastage of time and effort and reduces the risk of schedule slippage. j)Less human resources are required as once one phase is finished, those people can start working on the next phase. Fig. 1 shows the waterfall model.

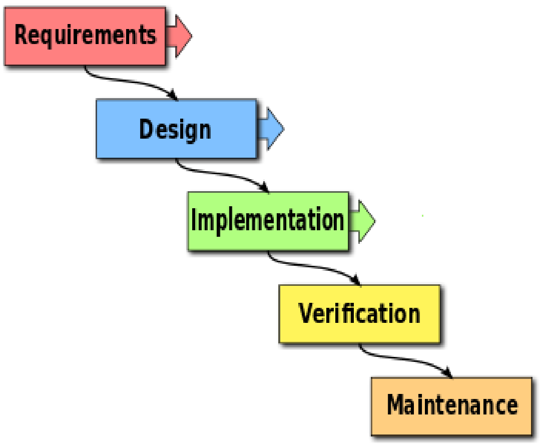

Fig. 1 Waterfall model

b) Architecture

The System architecture is as shown in fig. 2. Following are the functional module in the system. This is three-layered architecture, i) Selecting Positive samples. ii) Location and GIS Mining. iii) Spatial Information Mining.

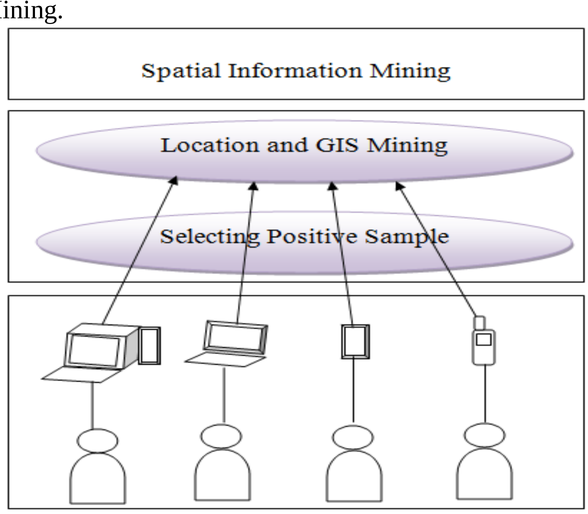

Fig. 2 Architecture

The working of each layer is as follows:

1) Social user layer

In this layer, the proposed method wants to collect the related data of urban emergency events. For example, if a user makes a message on Weibo about a fire occurrence, then the proposed method should collect this message.
For the collection of the message, the following heuristics algorithm is used:

- Message with the location or GIS information is prone to be a positive sample
- A message with the image or video information is prone to be a positive sample
- The original message is prone to be a positive sample

To identify Fake samples, we use the following approach

- Clustering of the information
- Estimate the density of clusters

c. Based on density lower than a threshold, we apply tweet user profile analysis to check if the user has an intention to spread fake and filter messages in that case.

2) Crowdsourcing layer

In this layer, the address and GIS information is mined from messages from social users. However, the location and GIS information related to the same urban emergency events should be clustered.

3) Spatial information layer

In this layer, the spatial information of the urban emergency event is mined and represented in visual form. The visualization will be in the form of a map with the location pin marked. The Pin colour is based on a number of tweets. Say Red for more comments. Blue for medium, yellow for less. Once any location is selected, the keyword categorization is displayed, so the user can select any keyword and view any information. Also, users can search within the displayed information.

IV. TESTING

System testing is basically a movement of different types of tests whose fundamental part is to totally hone the PC based structure. Testing is the last affirmation and acknowledgement activity inside the affiliation itself. At this stage, taking after the targets are endeavoured to fulfil the following: To affirm the way of the errand, To identify and get rid of any extra slip-ups from past stages, To endorse the item as a response for the main issue, To give the unflinching operational nature of the system, Amid testing the genuine activities are centred around the examination and modification of the source code.

A. Unit Testing

Generally, the Unit testing is done for each module. But since the system is developed in an Object-Oriented Programming model, the modules have factored to classes, and the authors have written unit testing for each module.

B. Integration Testing

Data can be lost transversely over the interface. One module can adversely influence another. Sub-limits, when solidified, should not diminish the looked for genuine limits. Joining testing is a conscious system for building up the venture structure. It addresses the issues associated with the twofold issues of check and program improvement. The central and main objective in this testing technique is to take the unit attempted modules and assemble a task structure that has been coordinated by arrangement. Table I shows how the Testing is done with different integrated classes.

Table 1. Integration Testing

Integrated Classes	Personalities in each class	Undergone tests	Conclusion
Class: Main	startJob()	All the functionalities in class were checked	Success
Class: Map	map()	All the functionalities in class were checked	Success
Class: Reduce	reduce()	All the functionalities in class were checked	Success
Class: TweetProcessor	processMsg() filterMsg()	All the functionalities in class were checked	Success
Class: Spatial Mining	clusterMsg() getGroupedInfo()	All the functionalities in class were checked	Success

C. Validation Testing

Towards the completion of joining testing, writing computer programs is done and accumulated as a pack. Interfacing errors are uncovered and helped. Endorsement testing can be portrayed from various perspectives. Here the testing acknowledges the item limit in a way that is sensibly expected by the customer. Table II illustrates how the Validation testing is done with different working functions along with its results.

Table 2. Validation testing with different working functions

Working Functions	Intake	Undergone Tests	Conclusion
Working of Crowdsource jar file	The user has to type the command Hadoop – jar crowdsource .jar in the terminal window	All partitions created, check the result message has to be displayed in the window.	Success
Working of Crowdsourcing Newsfront	User has to click the buttons called query by location, query by popularity and query by keyword	Based on the query, the result will display in the panel.	Success

V. RESULTS

In this paper, an emphasis was made to ensure the spatial data mining of the urban emergency events with the help of crowd-sourcing. With the advancements of knowledge or information communication technologies, it's critical to enhance the efficiency and accuracy of emergency management systems through modern processing techniques. Figure3 and figure 4 define the results and the outputs that have been obtained after the step by step execution of all the modules of the system.

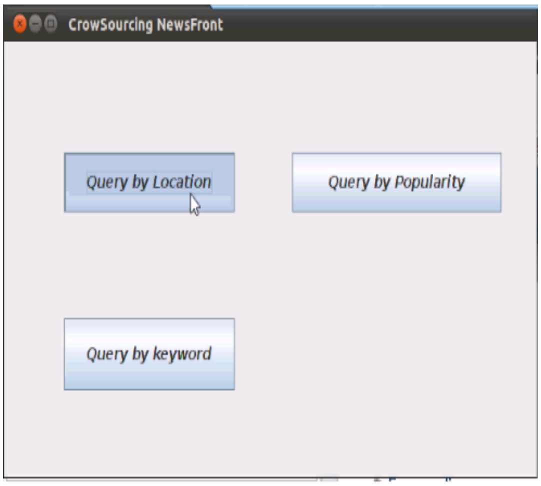

Fig. 3 New front code view window

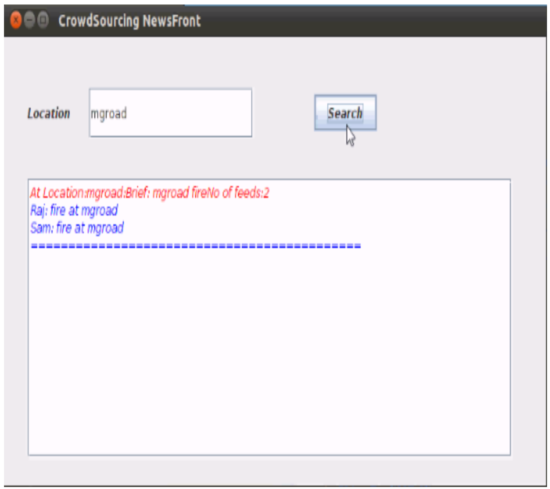

Fig. 4 The New front code now output

Enter the location, popularity, keyword and click on search. Based on the location, popularity and keyword message is sorted and displayed.

VI. CONCLUSION

In this work, the Crowdsourcing based model for mining the spatial information of several urban emergency events is introduced. At first, the basic definitions of the proposed method are given. Next, the positive samples are selected in order to mine the spatial data or information of urban emergency events. Thirdly, the location and the Geographic Information System's information are extracted from different positive samples. At last, the real spatial information is determined based on address and GIS information.

REFERENCES

[1] Goodchild, M.F., Citizens as Voluntary Sensors: Spatial Data Infrastructure in the World of Web 2.0, International Journal of Spatial Data Infrastructures Research 2 (2007) 24-32.

[2] (Business Models for the Next Generation of Software. http://www.oreillynet.com/pub/a/oreilly/tim/news/2005/09/30/what-is-web-20.html (2005).

[3] T. Sakaki, M. Okazaki, and Y. Matsuo. Earthquake Shakes Twitter Users: Real-time Event Detection by Social Sensors, In Proceedings of the 19th international World Wide Web conference, (2010) 851-860.

[4] V. Krishnamurthy and H. Vincentb Poor., A Tutorial on Interactive Sensing in Social Networks, IEEE Transactions on Computational Social Systems, early access, (2014).

[5] Field K and O'Brien J., Cartoblography: Experiments in using and organizing the spatial context of micro-blogging,Transactions in GIS, 15 (2010) 5–23.

[6] Stefanidis A, Crooks A, and Radzikowski J. Harvesting ambient geospatial information from social media feeds., GeoJournal, 77 (2012).

[7] Earle P, Guy M, Buckmaster R, Ostrum C, Horvath S, and Vaughan A. OMG earthquake! Can Twitter improve earthquake response? Seismological Research Letters 81 (2010) 246–51.

[8] Mummidi, L. and Krumm, J. , Discovering points of interest from users' map annotations, GeoJournal, 72(3-4) (2008) 215-227.

[9] Crandall, D., Backstrom, L., Huttenlocher, D., and Kleinberg, J., Mapping the World's Photos, In Proceedings of the 18th International World Wide Web Conference, (2009) 761-761.

[10] Pultar, E., Raubal, M., and Goodchild, M.F. GEDMWA: Geospatial Exploratory Data Mining Web Agent., In Proceedings of the 16th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, (2008).

[11] M. Trusov, V. Bodapati, and E. Bucklin., Determining influential users in social internet networks, J.Market.Res., 47 (2010) 643–658.

[12] T. Sakaki, M. Okazaki, and Y. Matsuo., Tweet Analysis for Real-time Event Detection and Earthquake Reporting System Development, IEEE Transactions on Knowledge and Data Engineering, 25(4) (2013) 919-931.

[13] A. Crooks, A. Croitoru, A. Stefanidis, and J. Radzikowski., Earthquake: Twitter as a Distributed Sensor System, Transaction in GIS, (2012) 1-26.

[14] B. Longueville, R. Smith, and G. Luraschi., OMG, from here, I can see the flames, a use case of mining location-based social networks to acquire Spatio-temporal data on forest fires., In Proceedings of the International Workshop on Location-Based Social Networks, (2009) 73-80.

[15] X. Jin, S. Spangler, R. Ma, and J. Han., Topic Initiator Detection on the World Wide Web., In Proceedings of the 19th international World Wide Web conference, (2010) 481-490.

[16] Abdul Razaque, Survey of Crowdsensing: Architecture, Classification and Security Challenges., SSRG International Journal of Electronics and Communication Engineering ., 6(10) (2019).

[17] Z. Xu, et al.., Knowle: a Semantic Link Network-based System for Organizing Large Scale Online News Events, Future Generation Computer Systems, 43(44), (2015) 40-50.

[18] L. Wang, J. Tao, et al. , G-Hadoop: MapReduce across distributed data centres for data-intensive computing., Future Generation Computer Systems, 29(3) (2013) 739-750.

[19] Z. Xu et al., Semantic-based representing and organizing surveillance big data using video structural description technology, The Journal of Systems and Software, 102 (2015) 217-225.

[20] Z. Xu, X. Luo, S. Zhang, X. Wei, L. Mei, and C. Hu. Mining Temporal Explicit and Implicit Semantic Relations between Entities using Web Search Engines, Future Generation Computer Systems, 37 (2014) 468-477.

[21] C. Hu, Z. Xu, et al. Semantic Link Network-based Model for Organizing Multimedia Big Data., IEEE Transactions on Emerging Topics in Computing, 2(3) 376-387.

[22] X. Luo, Z. Xu, J. Yu, and X. Chen., Building Association Link Network for Semantic Link on Web Resources. IEEE transactions on automation science and engineering, 8(3)(2011) 482-494.