# THE APPLICATION OF VGI ON SPATIAL CLUSTER ANALYSIS OF TRAFFIC INCIDENTS

Yu-Chi Kao [1], Chao-Hsien Wei [2], Yung-Chung Chuang [3], Tsu-Chiang Lei [4]

Department of Urban Planning and Spatial Information, Feng-Chia University

No. 100, Wenhwa Rd., Seatwen, Taichung, 40724 Taiwan (R.O.C.)

Email [1]: kaoyc@fcu.edu.tw Email [2]: st_webber@gis.tw

Email [3]: yungcchuang@fcu.edu.tw Email [4]: tclei@fcu.edu.tw

**KEY WORDS:** Volunteered Geographic Information, Citizen as Sensors, Traffic Incident, DBSCAN

**ABSTRACT**

This study through "citizen as sensors" concepts of volunteered geographic information, combined with improved density-based spatial clustering of applications with noise (DBSCAN) method used for traffic incident identification. Cluster identification is taking time and space and created a three dimensional cluster model. Simulation of statistical data according to the pattern of roads, where the incident occurred is defined as five categories, including: level crossings, general road (line), circle, square. Outcome data are grouped according to each point in time from the events of the termination point and duration of the event record. In accordance with data validation test is the final event information as the event witnessed a termination condition. And join the search during the time parameter, determine the duration of the event, so as to obtain a flexibility to adjust event time determination. Results of simulation tests showed that when the search space scale, the higher the noise false positive rate, the lower the information notified missing rate; conversely, smaller-scale search, the false positive rate, the lower the noise, the higher the notification information leak rate. The future lies in the traffic incident briefing, the main reference index values missing values false and noise data for the event may be recognized.

## 1. Introduction

VGI rapid development in recent years, scholars have proposed VGI apply to emergency management and rewards (Chen Xu, 2010). Laura Spinsanti (2013) proposed GeoCONAVI (Geographic CONtext Analysis for Volunteered Information) method and system hardware through existing facilities to retrieve, process, analyze and evaluate large amounts of data. This method can effectively expand the amount of content VGI, and the use of clustering time and space to support and verify that VGI while both reliability and operability, used in emergency events. Also after the incident instant data collection is also very important, so researchers hope that through the use of mobile devices for the first time witnesses will be notified to send data to a central system, these emergency information through VGI thus acquired a large number of classification, to event management, reporting and purpose (Erkine and Gregg, 2012) release.

However, data on the distribution of these clusters may be arbitrary linear, spherical or have continuity of shape characteristics, etc., and therefore extremely important in the identification of a cluster algorithm spatial data. As for dealing with these large amounts of spatial data, DBScan data clustering method can segment the density in space-based, simply enter an appropriate parameter that is able to identify clusters of arbitrary shape. Compared to the common CLARANS algorithm, DBScan can more efficiently identify a large number of spatial data and processing

irrelevant noise value (Martin Easter, 1996). Furthermore If GeoCONAVI (Laura Spinsanti, 2013) of section properties, taking into account the situation of the time and space clustering, combined DBScan clustering method is bound to improve the accuracy of the information can be effectively recognizable, and thus enhance the credibility of the information. And Renato Assunção (2008) have proposed a method to quickly detect the emergence of space-time clusters, mainly in accordance with the minimum requirements for user-defined parameters, randomly generated model to describe the spatial clustering; This method simultaneously space and time clustering, detection a time of one or more space-time clusters began to appear, and statistical inference, and new cluster detected, in order to reduce false information. And John Steenbruggen (2012) in the study were asked the following data for Telecommunications (Telecom Data) in event management: (1) telecommunications data can provide diverse types of information, improve the ability of perception of time and space situational; (2) telecommunications data can be used as an effective source event detection (3) affected by the incident were able to provide a large number of real-time data, and spatial and temporal information of the severity of the event is closely related nature.

This study will be conducted for the following points in the study developed to quantify checklist of information mechanisms and methods, and proposes improvement objectives: (1) notification to provide information on the information technology (2) to resolve data timeliness, repeatability and reliability of problem (3) evaluation indicators and targets (4) inform the architectural model build.

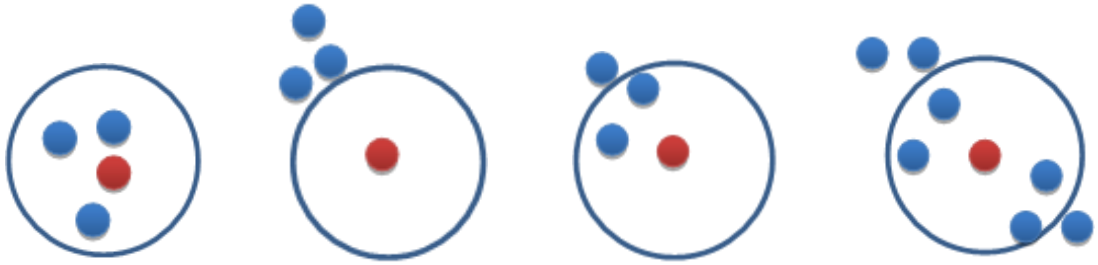
## 2. Research methods

VGI information as to when the traffic incident reporting basis, there will be three main issues to be overcome were informed repetitive events, traffic incidents witnessed the release of the timeliness and reliability of notification, expect space informed by witnesses when and time will be adjacent to each other's characteristics, to identify the relationship between the information in the notification similarity and clustering. Therefore, this study will be used as the core data clustering methods research methods and the improvement of its proposed suitable for use in identification of the traffic incident-based systems.

(A) Clustering method

(1). Spatial distribution - Space repeatability

When traffic incident witnesses who for distribution space in terms of the distribution of the location close to witnesses, the likelihood of the same event also witnessed relatively high, in addition to the use of characteristic DBScan can not belong to any cluster information data points considered noise to give shave, shown in Figure 1.



(a) (b) (c) (d)

Figure 1. Spatial distribution, a). forming a cluster. b). noise. c).join an existing cluster. d). connecting two clusters

(2) Time characteristics - Event timeliness

To recognize traffic incident for the purpose of starting the smallest unit of each set of traffic incidents distinction minutes, depending on time for unidirectional extension. The following Figure 2 as an example, in the briefing point in time five minutes before and after the search, contains more than three items of information when determining the formation of a cluster is considered the occurrence of an event; through continuing to wait for the search range of new information, conduct time extension notification information, it is the same group of data; while unable to continue its group data points in the last piece of data, the time point plus the search time interval compared with clustering end point, without any in determining the scope of the search there may be added to the information when it is considered an event lifting points.

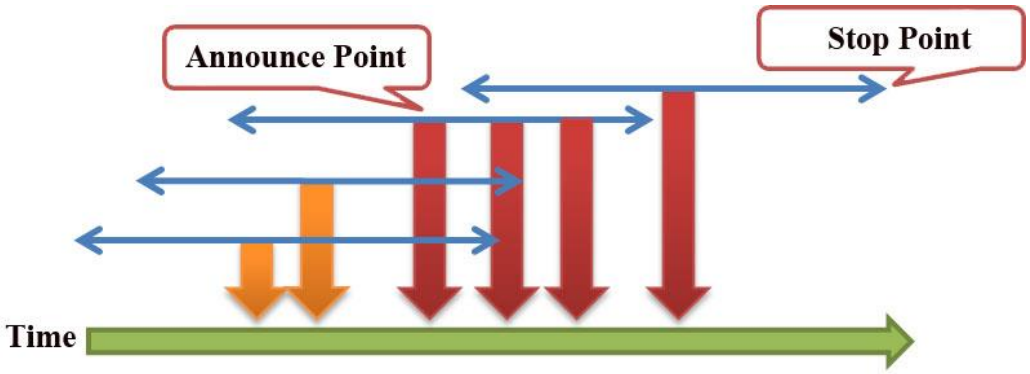



Figure 2. One-dimensional time clustering concept map

(3) Space considerations clustering

After receipt of notification data to determine the number of traffic incidents, location and time information shall also consider the time and space clustering, and therefore must be addressed, such as the differences between the two scales and conversion issues: Unit (1) time and space integration (2) the calculation of time and space scales (3) the search criteria to quantify. Therefore, the difference in handling the issue of time and space scales, the proposed order ellipsoid equation as modified DBScan range predicate to exclude information in time and space away from the information, but the use of extended properties DBScan, but were more than events detection, thereby achieve the purpose of identification of traffic incidents. In Figure 3, for example.

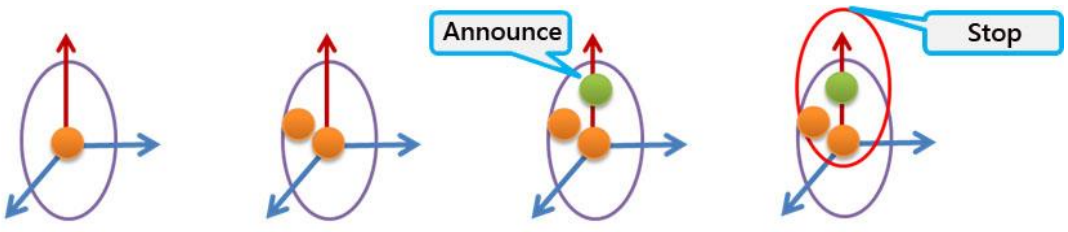



Figure 3. Three dimensional diagram of DBScan grouping, a). spheroid formation b). scope judge c). event publishing d). event lifted

(B) The data analysis methods

(1). Parameter comparison

In this study, the range of possibilities as reference events DBScan grouping modified the scope of the search of the parameter space conditions vary due to the severity of accidents or road conditions and time of the warning signs may be due to the incident, so the conditions set 30-100 meters were clustering results of comparison, and selection of 30 meters, 50 meters and 100 meters and other parameters were compared to investigate the difference. Because different set of search parameters, therefore need to explore events and eyewitness identification rate of false positive rate information, which should be seized nuclear situation is that when the search range is large (100 meters) would the two incidents linked merger incident; smaller search range or when the case (30 meters) whether due to uneven distribution of a single incident that loose, fragmented situation, resulting in a multi-incident identification of the incident.

(2).Clustering index

In analyzing the results of modified DBScan grouping in this study will be calculated according to the indicators of individual clustering results and to compare the results of the clustering analysis of indicators will be divided into two, contains the event recognition rate and false positive rate of two witnesses, a description as follows:

(a). Event recognition rate: The total number of clustering results of group identification number / simulated event

Event identification rate traffic incident identification of the most important aspect, in addition to instant access to the event location, time of occurrence, duration and termination time, the most important part of accurately determine whether the number of incidents, to identify automated advantage.

(b). Witnesses false positive rate: (Total noise misjudgment + Bulletin total Missing) / total number of witnesses

Through eyewitness identification error rate obtaining in the process, the outcome indicators when comparing different grouping parameters through eyewitness false positive rate indicators, to explore the impact of changes in search of practical results caused.

## 3. Results and analysis

(A) Clustering outcomes

In Table 1, for example, the total number of data for the 233 witnesses pen data, which includes 46 pen analog noise information can be found in accordance with the current experimental data to analyze the duration of the events are all the same, except the case of an event F, since the 100 meters and 50 meters clustering outcome, the final sum witnessed the event data values for the noise, all this noise value is added to the group pen, thus causing the error event time information.

As shown in Table 2, when the judge set the ellipsoid formula parameter (100 meters / 5 minutes) conditional case, you can get seven incidents of individual clustering results, before clustering, the raw data has to be marked number, and are known to contain true information and individual information, such as noise data classification, clustering results of the overall point of view, a total of 187 incident strokes witnesses informed information in this grouping conditions, the total number of sightings Missing data is a pen data, while false information on the number of noise, there are 30 pens, a high false positive rate in which event B, making the

overall outcome of the event clustering error rate increased to 18%, suggesting that the reason is that when the search scope to expand the information, the data can be searched opportunity will increase, so will reduce the number of false negatives, but will increase the chance of false positives relative, resulting to enhance the rate of false positives. When setting the ellipsoid formula to determine the parameters (50 meters / 5 minutes) conditional circumstances, through the known classification data clustering, eyewitness information includes the total number of 2 and 19 strokes of the pen Missing noise false information, in whole the average rate of 13% false positive events, compared to (100 meters / 5 minutes) grouping the results of a 5% reduction in its results for the noise there is a significant reduction in noise false false value.

Table1. The table of clustering results.

| Event | 100M | | | 50M | | | 30M | | | Simulate Time (minute) |
|---|---|---|---|---|---|---|---|---|---|---|
| | Announce | Stop | Duration (minute) | Announce | Stop | Duration (minute) | Announce | Stop | Duration (minute) | |
| A (Intersection) | 09:08 | 09:20+5 | 17 | 09:08 | 09:20+5 | 17 | 09:08 | 09:20+5 | 17 | 15 |
| B (Intersection) | 09:25 | 09:33+5 | 13 | 09:25 | 09:33+5 | 13 | 09:25 | 09:33+5 | 13 | 10 |
| C (Intersection) | 10:16 | 10:24+5 | 13 | 10:16 | 10:24+5 | 13 | 10:16 | 10:24+5 | 13 | 10 |
| D (Traffic Circle) | 10:56 | 11:05+5 | 14 | 10:56 | 11:05+5 | 14 | 10:56 | 11:05+5 | 14 | 10 |
| E (Linear) | 16:02 | 16:04+5 | 7 | 16:02 | 16:04+5 | 7 | 16:02 | 16:04+5 | 7 | 5 |
| F (Linear) | 16:30 | 16:38+5 | 13 | 16:30 | 16:38+5 | 13 | 16:30 | 16:35+5 | 10 | 10 |
| G (Intersection) | 11:57 | 12:30+5 | 38 | 11:57 | 12:30+5 | 38 | 11:57 | 12:30+5 | 38 | 35 |

Table 2. Experimental results of table grouping.

| Event | Data Amount | 100M | | | | 50M | | | | 30M | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Cluster | Omit | Misjudge | Error (%) | Cluster | Omit | Misjudge | Error (%) | Cluster | Omit | Misjudge | Error (%) |
| A (Intersection) | 21 | 27 | 0 | 6 | 0.22 | 24 | 0 | 3 | 0.13 | 22 | 0 | 1 | 0.05 |
| B (Intersection) | 10 | 15 | 0 | 5 | 0.33 | 12 | 0 | 2 | 0.17 | 10 | 0 | 0 | 0.00 |
| C (Intersection) | 21 | 23 | 0 | 2 | 0.09 | 21 | 1 | 1 | 0.09 | 17 | 4 | 0 | 0.16 |
| D (Traffic Circle) | 20 | 22 | 1 | 3 | 0.17 | 22 | 1 | 3 | 0.17 | 19 | 2 | 1 | 0.13 |
| E (Linear) | 13 | 15 | 0 | 2 | 0.13 | 15 | 0 | 2 | 0.13 | 15 | 0 | 2 | 0.13 |
| F (Linear) | 16 | 20 | 0 | 4 | 0.20 | 19 | 0 | 3 | 0.16 | 17 | 0 | 1 | 0.06 |
| G (Intersection) | 86 | 94 | 0 | 8 | 0.09 | 91 | 0 | 5 | 0.05 | 88 | 0 | 2 | 0.02 |
| Total | 187 | 216 | 1 | 30 | **0.18** | 204 | 2 | 19 | **0.13** | 188 | 6 | 7 | **0.08** |

## 4. Conclusions

Bulletin of traffic incidents in the past informed the architecture, there are already established methods and models, and many informed and more traditional research model to improve the notification strengthened as information integration, the study of natural language law, but also because of the popularity of mobile devices, making and receiving information conveyed will be more convenient. The study was to explore issues and problems VGI information on the use of the analysis, the use of DBScan law occurred dynamic identification and termination time in the event, to solve the timeliness of events, repetitive process information, and to consider the time and space to reach a clustering method, and retain DBScan original grouping ductility, noise filtering and other features to improve, mainly from the traditional two-dimensional space and time DBScan do combine to determine the event location, time, duration occurs, the termination time. Improved DBScan to determine the type most ellipsoid method of information search, to also consider the purpose of identification of time and space. Through measurement analysis was obtained in accordance with the parameters set by grouping cases Jieneng recognize success, but dynamically adjusting the interval of time in the event it can also identify, contain short and long influence and so, for the future of traffic information application to open another over one yuan information pipeline to provide more services and improve traffic information value.

## References


1. Chen Xu, 2010. Exploring volunteered geographic information(VGI) for emergency managementtoward a wiki gis framework, Texas A&M University.
2. Erskine, M. and Gregg, D., 2012. Utilizing Volunteered Geographic Information to Develop a Real-Time Disaster Mapping Tool: A Prototype and Research Framework. In CONFIRM 2012 Proceedings
3. Goodchild, M.F. & J. Alan Glennon., 2010. Crowdsourcing geographic information for disaster response: a research frontier, International Journal of Digital Earth, 3:3, 231-241
4. Goodchild, M.F., P. Fu, and P. Rich, 2007. Sharing geographic information: an assessment of the Geospatial One-Stop. Annals of the Association of American Geographers.
5. Goodchild, M.F, 2007. Citizens as Voluntary Sensors: Spatial Data Infrastructure in the World of Web 2.0,International Journal of Spatial Data Infrastructures Research, 2, 24-32.
6. John Steenbruggen, Maria Teresa Borzacchiello, Peter Nijkamp, Henk Scholten., 2012. Data from telecommunication networks for incident management: An exploratory review on transport safety and security.
7. Laura Spinsanti,Frank Ostermann, 2013. Automated geographic context analysis for volunteered information.
8. Martin Easter, Hans-Peter Kriegel, Jörg Sander, Xiaowei Xu, 1996. A density-based algorithm for discovering clusters in large spatial databases with noise. In Evangelos Simoudis, Jiawei Han, Usama M. Fayyad. Proceedings of the Second International Conference on Knowledge Discovery and Data Mining (KDD-96).pp. 226-231.
9. Michael A. Erskine,Dawn G. Gregg., 2012. Utilizing Volunteered Geographic Information to Develop a Real-Time Disaster Mapping Tool: A Prototype and Research Framework"
10. Renato Assunçãoa,Thais Correab., 2008. Surveillance to detect emerging space–time clusters.