Identifying Optimal Scales for Spatio-temporal Crime Clusters

Tongxin Chen[*1], Tao Cheng[†1]

[1] SpaceTimeLab For Big Data Analytics, Department of Civil, Environmental and Geomatic Engineering, University College London, UK

January 16, 2020

Summary

The spatial and temporal scales are not only two essential parameters for the spatio-temporal clustering algorithm to generate the crime clusters but are significantly helpful for determining the interventive distance at space and time in place-based crime prevention. This study presents the issue of identifying the optimal spatial-temporal scale when examining the micro-level crime clusters approached by density-based spatio-temporal clustering methods. The approach comprises adopting a clustering evaluation index to examine the performance of different clustering results from a range of space and time values iteration. For this purpose, two types of density-based clustering algorithms called ST-DBSCAN and ST-OPTICS are compared to determine the optimal scales for space-time crime clusters. A case study is demonstrated using individual crime records of burglary from Vancouver, Canada in 2010. Several derived results are significant. First, appropriate scales – 500m and 3 days can be distinctively determined by clustering algorithm ST-OPTICS from our tested parameters. Second, the narrowed scales were found in this study significantly for spatio-temporal crime clusters, which can help to develop a more focused and specific policing tactics.

KEYWORDS: Spatio-temporal clustering, crime hotspot, near-repeat victimisation, density-based clustering, ST-OPTICS

1. Introduction

The spatio-temporal scales are significant for detecting appropriate crime concentrations in space and time known as spatio-temporal crime clusters following the real-word phenomena and can help to build a specific crime intervention tactics. In previous works, the spatial and temporal scales has been widely discussed in criminological researches, which denotes the spatio-temporal crime clusters are distributed in small units of space and time (Bowers and Johnson 2005, Johnson et al. 2007). However, previous works about detecting such crime patterns are limited by the statistical processes, so as the selection of spatio-temporal scale is either arbitrary defined or ignored. In this study, density-based clustering algorithms called ST-DBSCAN and ST-OPTICS are used to extract crime clusters and select the optimal scales by the comparison of clustering evaluation index across the iteration of different parameters.

2. Methodology

2.1 Data

Dataset concerning reported incidents of Vancouver, Canada were obtained from the Vancouver Police Department published website (https://geodash.vpd.ca/opendata/). It compromises spatial and

[*] tongxin.chen.18@ucl.ac.uk

[†] tao.cheng@ucl.ac.uk

temporal information (i.e., the datetime and coordinates), neighbourhood information and categories of crime incidents. In this study, we choose the burglary-related category -- 'Breaking and entering – residential' (BER) of Vancouver in 2010 as the test data. In sum, it shows 3270 burglary incidents were reported in Vancouver city (115 km^2) in 2010.

2.2 Density-based Spatio-temporal Clustering Methods

Previous studies about spatio-temporal crime patterns mainly utilised the statistics-based method. For example, the space-time interaction method represented by *Knox's Test* is popularly used to detect the space-time crime clusters (Townsley, Homel and Chaseling 2003, Bowers and Johnson 2004, Johnson and Bowers 2004, Johnson et al. 2007, Ratcliffe and Rengert 2008). Further, recent studies employed the *space-time scan statistics* to discuss such crime clusters in space and time (Uittenbogaard and Ceccato 2012, Malleson and Andresen 2015). However, statistical processes are lack of detecting the clusters of arbitrary shapes which can be considered a key attribute in the context of crime pattern distribution. In density-based clustering methods, clusters are separated by a sparse region comprising the data are "relatively few" so as to the clusters with regular or arbitrary shapes can be appropriately extracted (Bhuyan and Borah 2013).

Herein, we choose two density-based spatio-temporal clustering algorithms called ST-DBSCAN and ST-OPTICS to detect the spatio-temporal crime clusters in the study data. First, Spatial Temporal Density-based Spatial Clustering of Applications with Noise (ST-DBSCAN) proposed by Birant and Kut (2007). The ST-DBSCAN algorithm was developed from DBSCAN built by Ester et al. (1996). And it needs three pre-set parameters: spatial maximum reachable distance (SMRD), temporal maximum reachable distance (TMRD) and minimum number of points (MinPts). In the case of space-time crime clusters, the SMRD and TMRD refer to the spatial scale and temporal scale, respectively. In addition, MinPts refers to the minimum numbers of crime incidents in a crime cluster. Second, ST-OPTICS, Spatial Temporal Ordering Points to Identify the Clustering Structure. The basic idea and the parameters of ST-OPTICS and ST-DBSCAN are the same, however, OPTICS can still extract meaningful clusters from a density-varied dataset approached by creating an augmented *ordering* of the data to optimize the choice of cluster structure (Ankerst et al. 1999).

2.3 The Evaluation Index for Clustering Results

Evaluation Index, i.e., the performance metric is used to assess different results from various parameters in an algorithm or different algorithms. In the case of clustering issue, abundant indices were developed to evaluate the quality of clusters extracted from a dataset by considering similarities across clusters, such as the Dunn index developed by Dunn (1973) or DB index proposed by (Davies and Bouldin 1979). In this study, DB index is selected for the evaluation of crime clustering results as it is not only easily understandable but less computationally time-consuming than Dunn index. The DB index is defined as:

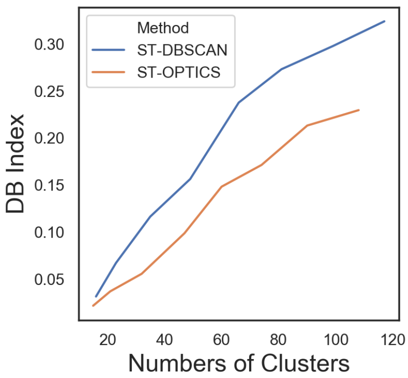
$$\text{DB index} = \frac{1}{k}\sum_{i=1}^{k} \max_{i \neq j}\left\{\frac{\Delta(X_i) + \Delta(X_j)}{\delta(X_i, X_j)}\right\} \tag{1}$$

where, k is the numbers of clusters. $\delta(X_i, X_j)$ is the distance between cluster X_i and X_j known as the inter-cluster distance and $\Delta(X_i)$ is the intra-cluster distance for cluster X_i. Following the definition, a lower DB index denotes a better clustering result.

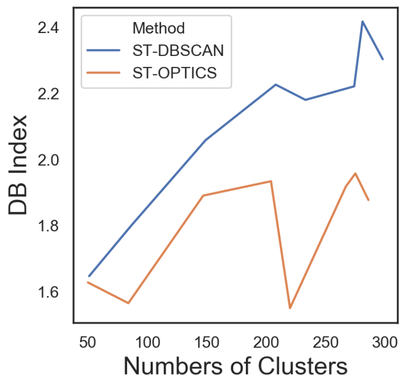
In our spatio-temporal clustering algorithms, the spatial distance SMRD iterates over the range from 100m to 1000m with steps of 500 meters, and the temporal distance TMRD iterates over the range [3, 30] with steps of about 5 days (1 day for TMRD is excluded as crime occurs within one day are not considered in this study). Then, the MinPts is defined as 3 for this experiment. The comparisons of DB index across clustering results are separated into four groups (100m, 500m, 1000m and 1500m) under the control of spatial scales.

3. Results and Discussion

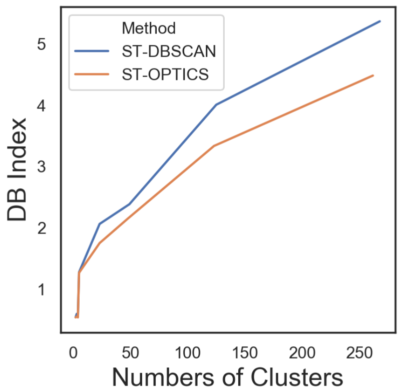
Figure 1 illustrates the evaluation of clustering results based on ST-DBSCAN and ST-OPTICS. With four groups shown in subfigures, the minimum values of DB index could be also found at the minimum numbers of clusters based on ST-OPTICS with 100m and 3 days in group A, 1000m and 20 days in group C and 1500m and 7 days in group D. Particularly, for group B in subfigure b), the minimum value of DB Index in the results is expected at the cluster number equals to 233 based on ST-OPTICS with 500m and 3 days.



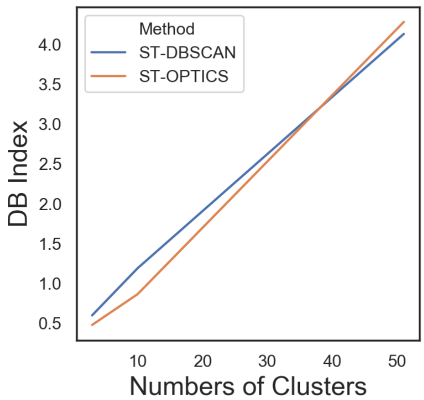
a) Group A: 100m (SMRD)



b) Group B: 500m (SMRD)



c) Group C: 1000m (SMRD)



d) Group D: 1500m (SMRD)

Figure 1 Evaluation of clustering results from ST-DBSCAN and ST-OPTICS.

To compare the clustering performance across inter groups, the scales of 500m and 3 days for crime clusters is more appropriate than the other three groups as the numbers of clusters are extremely limited from this study area (16 clusters based on 100m and 3 days; 2 clusters detected based on 1000m and 20 days; 3 clusters detected based on 1500m and 7 days). In addition, all the index values of clustering results from ST-OPTICS is normally lower than the ones from ST-DBSCAN, which denotes the performance of ST-OPTICS is better than ST-DBSCAN in retrieving space-time crime clusters in the study area.

Figure 2 denotes the spatio-temporal crime hotspots of burglary in 2010 based on ST-OPTICS (500m and 3days) in Vancouver city, in which 233 crime clusters including 948 incidents are concentrated on the northern area.

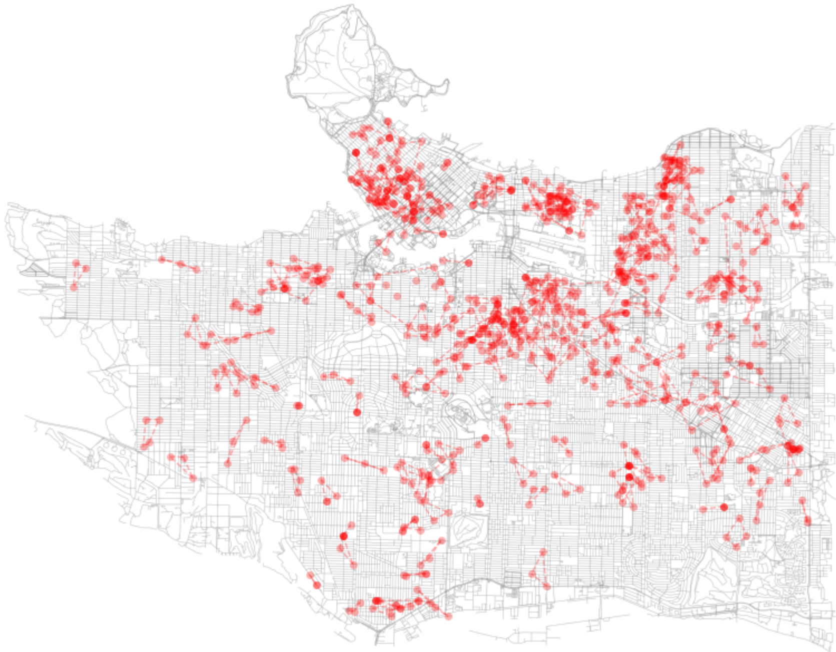


Figure 2 Spatio-temporal crime hotspots of Vancouver in 2010.

The characteristic of a spatio-temporal crime hotspot incorporates the crime incidents number, day periods, and the hotspots are defined by the convex hull. Table 1 shows the description of crime cluster characteristics, which shows the spatio-temporal crime clusters of burglary are restricted in small areas and short-terms.

Table 1 The characteristics of spatio-temporal crime hotspots.

	Mean	SD	Median	Minimum	Maximum
Crime incidents number	4.1	2.0	3	3	15
Day periods (day)	4.8	2.6	4	1	19
Crime hotspots area (m^2)	1259	727	1228	0.1	4428

4. Conclusion

This study has presented an approach to identify optimal scales for space-time crime clusters by comparison of the evaluation index. In summary, 500m and 3 days are considered as the appropriated scales from our tested parameters based on ST-OPTICS in this study area. Further, ST-OPTICS reveals a better performance than ST-DBSCAN in detecting spatio-temporal crime clusters. However, it is naïve that evaluation of crime clustering only considering the inter- and intra-distance across clusters, but some application-related factors could be involved in the future work.

5. Acknowledgements

Valuable suggestions from anonymous reviewers are gratefully acknowledged.

6. Biography

Tongxin Chen is a PhD student at SpaceTimeLab, Department of Civil, Environmental and Geomatic Engineering, University College London. He has the educational background of criminal investigation and criminology. His research interests include crime mapping, spatio-temporal crime pattern analysis and applied machine learning for crime analysis.

Tao Cheng, is a professor in Geoinformatics at Department of Civil, Environmental and Geomatic Engineering, University College London. She is the Founder and Director of SpaceTimeLab for Big Data Analytics. Her research interests span network complexity, Geocomputation, space-time analytics and Big data mining (modelling, prediction, clustering, visualisation and simulation) with applications in transport, crime, business, health, social media, and natural hazards.

References

Ankerst, M., M. M. Breunig, H.-P. Kriegel & J. Sander. 1999. OPTICS: ordering points to identify the clustering structure. In *ACM Sigmod record*, 49-60. ACM.

Bhuyan, R. & S. Borah. 2013. A survey of some density based clustering techniques. In *Proceedings of the 2013 conference on advancements in information, computer and communication*.

Birant, D. & A. Kut (2007) ST-DBSCAN: An algorithm for clustering spatial–temporal data. *Data & Knowledge Engineering,* 60**,** 208-221.

Bowers, K. J. & S. D. Johnson (2004) Who Commits Near Repeats? A Test of the Boost Explanation. *Western Criminology Review,* 5.

--- (2005) Domestic burglary repeats and space-time clusters: The dimensions of risk. *European Journal of Criminology,* 2**,** 67-92.

Davies, D. L. & D. W. Bouldin (1979) A cluster separation measure. *IEEE transactions on pattern analysis and machine intelligence***,** 224-227.

Dunn, J. C. (1973) A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters.

Ester, M., H.-P. Kriegel, J. Sander & X. Xu. 1996. A density-based algorithm for discovering clusters in large spatial databases with noise. In *Kdd*, 226-231.

Johnson, S. D., W. Bernasco, K. J. Bowers, H. Elffers, J. Ratcliffe, G. Rengert & M. Townsley (2007) Space–time patterns of risk: A cross national assessment of residential burglary victimization. *Journal of Quantitative Criminology,* 23**,** 201-219.

Johnson, S. D. & K. J. Bowers (2004) The stability of space-time clusters of burglary. *British Journal of Criminology,* 44**,** 55-65.

Malleson, N. & M. A. Andresen (2015) Spatio-temporal crime hotspots and the ambient population. *Crime science,* 4**,** 10.

Ratcliffe, J. H. & G. F. Rengert (2008) Near-repeat patterns in Philadelphia shootings. *Security Journal,* 21**,** 58-76.

Townsley, M., R. Homel & J. Chaseling (2003) Infectious burglaries. A test of the near repeat hypothesis. *British Journal of Criminology,* 43**,** 615-633.

Uittenbogaard, A. C. & V. Ceccato. 2012. Space-time clusters of crime in Stockholm, Sweden. In *Review of European studies*, 148-156.