Spatio-Temporal Point Event Cluster Detection by Controlling Shape Complexity

R. Inoue and H. Kimoto

Graduate School of Information Sciences, Tohoku University, Sendai, Japan
Email:{rinoue@, hiroshi.kimoto.t2@dc}.tohoku.ac.jp

Abstract

In this paper, a method is proposed to detect spatio-temporal clusters of point events by controlling the shape complexity of detected clusters. The partitioning of the target region and period into voxels makes it easy to evaluate the complexity of the spatio-temporal shapes and to impose constraints on the shapes of clusters to be detected. Experiments using simulation data and actual crime data confirmed the applicability of the proposed method and the need for careful interpretation of results.

Keywords: Spatio-Temporal Cluster Detection, Point Event, Shape Complexity.

1. Introduction

It has recently become easier to apply detailed location and time data to regional analysis by following open data policies. One data type with detailed spatio-temporal information is point event data. It is often used to detect clustered areas and periods. The most renowned analysis method for spatio-temporal cluster detection is the spatial scan statistic (e.g., Kulldorf and Nagarwalla, 1995; Kulldorf, 1997). The scan statistic and its derivative can detect clusters that remain in the same position with the same size for a certain time. Clusters of point events may appear, expand, migrate, transform, shrink, or disappear through the passage of time; however, existing methods are not capable of detecting such clusters.

To address the above limitation, detection of flexible-shaped clusters in spatio-temporal space is needed. Shape flexibility enables extraction of clusters whose shapes differ over time. Nevertheless, the shapes of detected clusters may be too complex and difficult to interpret. This problem occurs in spatial cluster detection. A method to control shapes was thus proposed (Duczmal et al., 2006). In this paper, an index is proposed to quantify the shape complexity in the spatio-temporal space. In addition, a method is presented to detect the cluster with shape constraints.

2. Shape Complexity Control in Spatio-Temporal Cluster Detection

2.1 Shape Complexity in Euclidean and Network Spaces: Previous Studies

Existing methods that control shapes of detected clusters define the shape complexity index by a comparison with the compact shape. Duczmal et al. (2006) introduced a spatial shape complexity index. The shape complexity of region Z, $K(Z)$ is defined by the ratio of the area of Z over the area of a circle with the same perimeter of the convex hull of Z (Figure 1 (a)). It is a dimensionless index; its value ranges from zero to one, and approaches one when a region has a compact shape (Equation 1).

$$K_{space}(Z) = \text{area}(Z) \Big/ \pi\left(\frac{\text{perimeter}(\text{convex\_hull}(Z))}{2\pi}\right)^2 \qquad \text{Equation 1}$$

Tsukahara and Inoue (2014) and Inoue and Tsukahara (2016) introduced the shape complexity index of a region defined by a set of connected links in a network. Accordingly, the shape is evaluated by the length of a region, Z. A compact region of Z in the network is first defined. It has the same total length and same central node as Z. The central node is defined as the node whose network distance to the farthest node is the shortest. It is similar to the centre of the smallest enclosing circle in the Euclidean space. After the central node of Z is identified, the compact region of Z is generated by selecting the connected links in short-distance order from the central node until the total length exceeds that of Z. The network distance from the central node in Z to the farthest node is compared to the compact region that corresponds to Z (Figure 1 (b)). The shape complexity is thereby given by

$$K_{network}(Z) = \frac{\text{longest\_shortest\_path\_distance}(\text{compact\_network}(Z))}{\text{longest\_shortest\_path\_distance}(Z)} \qquad \text{Equation 2}$$

Cluster detection by the scan statistic first assumes a stochastic process for the point distribution, and then evaluates regions by a likelihood ratio over the null hypothesis in which there are no clusters. Duczmal et al. (2007), Tsukahara and Inoue (2014), and Inoue and Tsukahara (2016) applied a genetic algorithm to find a region that has the maximum likelihood ratio by imposing constraints on the shape complexity.

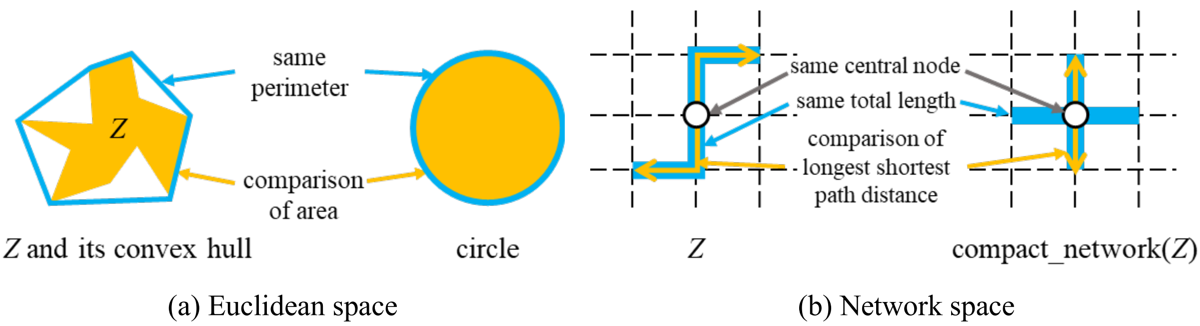


Figure 1: Shape complexity in Euclidean and network spaces proposed in previous studies.

2.2 Shape Complexity Index in Spatio-Temporal Space

We herein propose a shape complexity index in spatio-temporal space as a ratio of the volume of region Z over the volume of the convex hull of Z. When Z has compactness, an index value is close to one. However, the construction of the convex hull in three-dimensional space requires a time-consuming calculation. Thus, we propose an approach to discretizing the target region and period into voxels with consideration of a voxel as a minimum unit for analysis. The cluster region and period are represented as a set of contiguous voxels. The convex hull of the voxel set is also represented as a voxel set. Under this assumption, the convex hull of a given set of voxels is easy to calculate. The repetition of the procedure to create a convex hull in each plane that is parallel to x-y, y-t, or x-t planes can construct a convex hull in spatio-temporal space. We define the shape complexity index of spatio-temporal region Z by

$$K_{space-time}(Z) = \text{volume}(Z) / \text{volume}(\text{convex\_hull}(Z)). \qquad \text{Equation 3}$$

This approach searches the cluster region in the spatio-temporal space using a genetic algorithm under a shape complexity constraint.

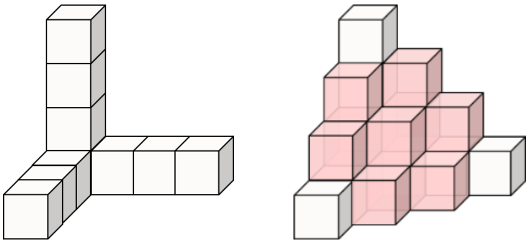


Figure 2: Sample set of voxels and its convex hull formed by a set of voxels.

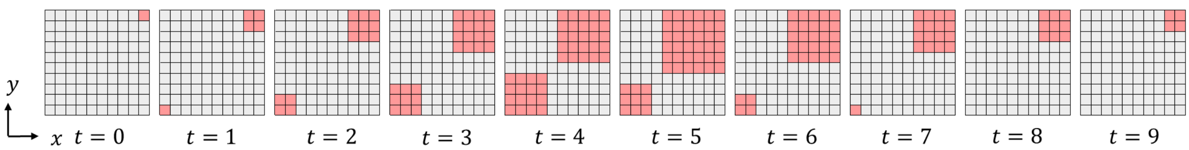
3. Application

3.1 Simulation Data

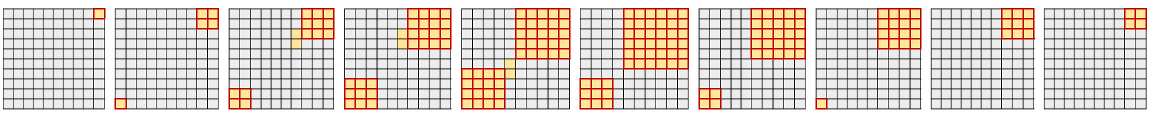
We tested the proposed method in terms of its feasibility, statistical power, and controllability of shapes using simulation data. We conducted an experiment that checked the shape differences using shape constraints. The target area consisted of 10×10×10 voxels. The two major cluster regions were set by 189 red voxels, as shown in Figure 1(a). The point distribution followed the Poisson point process; the parameter respectively inside and outside the cluster voxels was 50 and 10.

Figures 1(b) and (c) show the detected clusters under two different shape constraints. Yellow voxels represent the primary cluster in each analysis. The analysis without constraints extracts one large cluster by connecting the two regions with the inclusion of non-cluster voxels. The analysis with a strong constraint detects a small cluster that covers one region as a primary cluster, although the other region can be detected by further analysis.

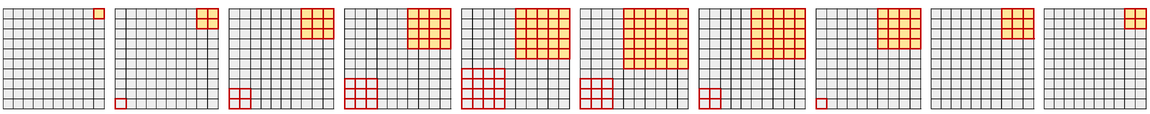
The results confirmed that the proposed method could identify a cluster that fit the shape constraint. However, the results could mislead analysts in their data interpretation. Analysts must conduct detection under several constraint conditions to obtain a probable interpretation from the point event dataset.



(a) Setting of clusters of simulation data.



(b) Detected cluster without the shape complexity constraint.



(c) Detected cluster with the constraint that the shape complexity index must be higher than 0.9.

Figure 3: Differences in detected clusters according to shape complexity constraint.

3.2 Crime report data

We analysed the locations and dates of burglaries in the London Borough of Barking and Dagenham in 2014 and 2015, as recorded by the Metropolitan Police Service. There were 3,362 burglaries in the area and period. The voxel size was 1 km × 1 km × 1 month. The number of voxels was 41 per month (Figure 2) and 984 in total. The results were obtained under two shape constraint conditions: no constraint and a minimum shape complexity index of 0.8.

Table 1 lists the detected clusters. It is natural that no constraint case outputs a cluster with a higher likelihood ratio value. However, the point density in the cluster is higher when a shape constraint is imposed. The result with no constraint could include voxels where burglaries are not concentrated.

Figure 4 shows the distribution of voxels in the primary clusters. Because the shape constraint tries to find compact clusters, the cluster ends early in June 2015. It does not mean there are no clusters in the latter half of 2015; further analysis may enable identification of different clusters.

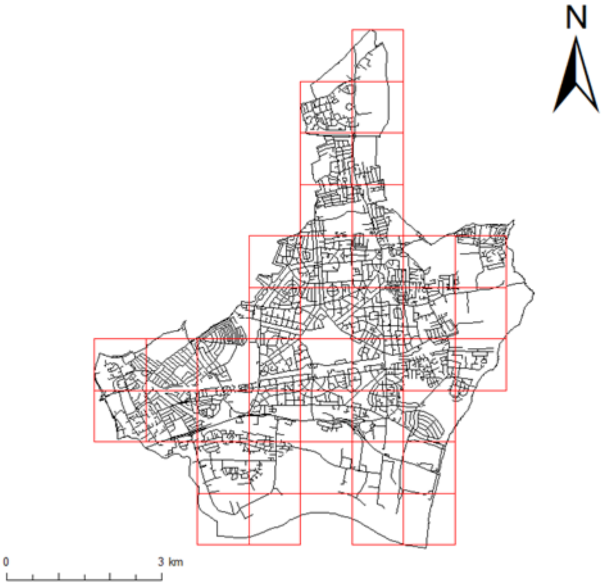
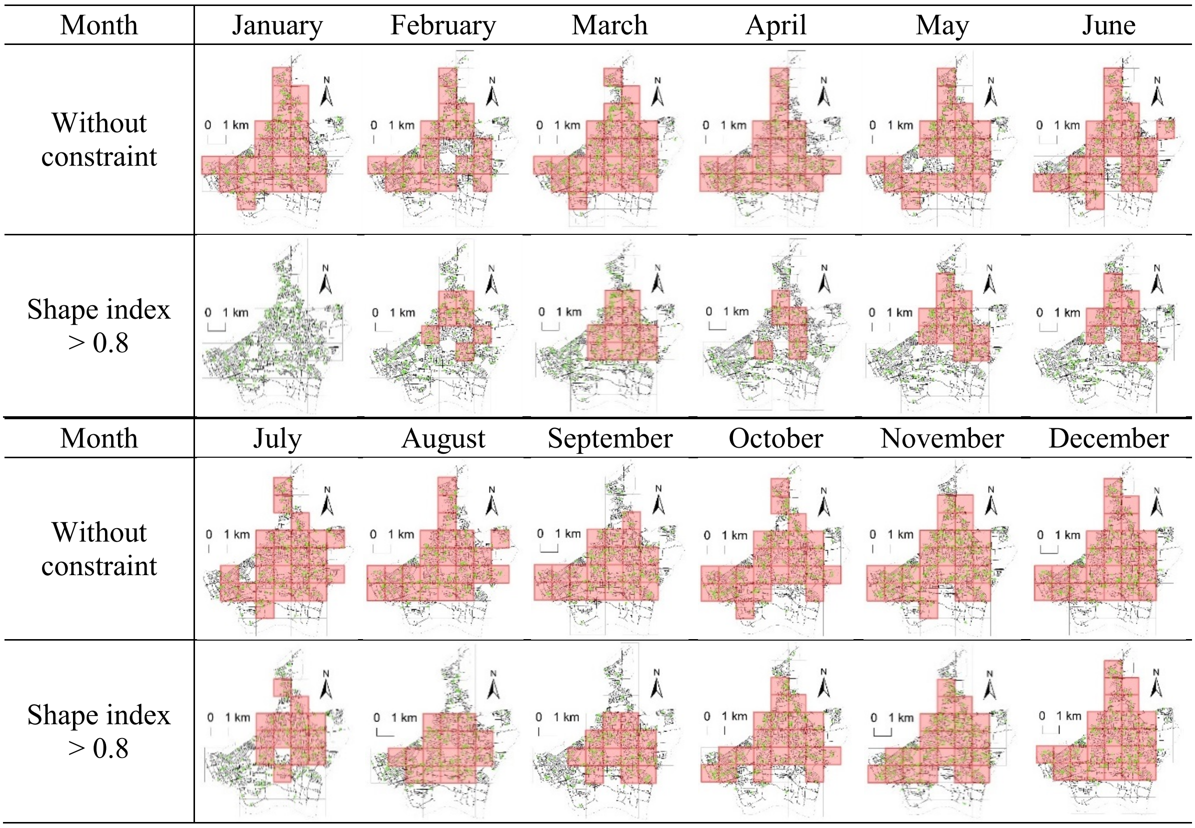


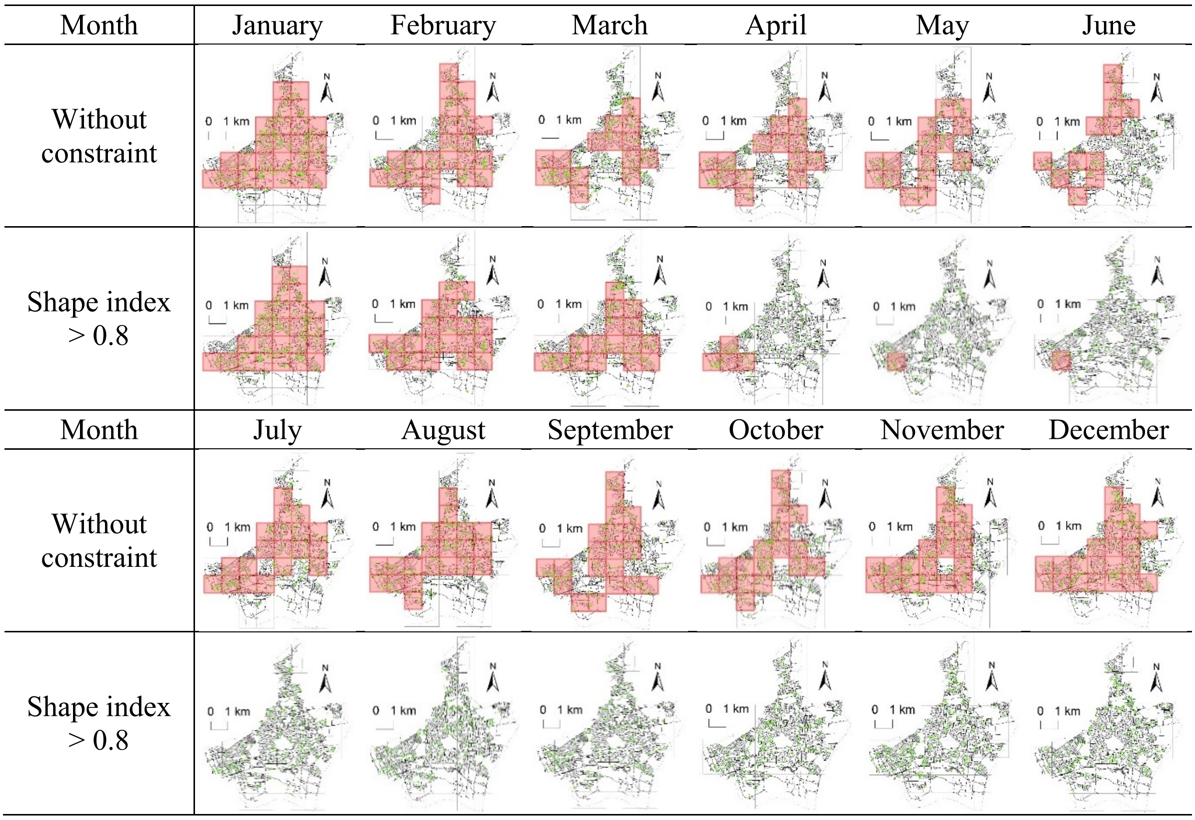
Figure 4: Target area and spatial distribution of voxels.

Constraints on shape index	Shape index	Log-likelihood ratio	Number of points in cluster	Number of voxels in cluster	Number of points per voxel	
					Inside cluster	Outside cluster
Without constraint	0.67	987.7	2,985	524	5.70	0.82
Over 0.8	0.81	403.2	1,534	229	6.69	2.42

Table 1: Summary of detected clusters.



(a) 2014



(b) 2015

Figure 5: Detected burglary clusters.

4. Conclusion

We proposed a cluster detection method of spatio-temporal point events. By splitting the target region and period into voxels, it is easy to evaluate the complexity of spatio-temporal shapes and to impose constraints on the shapes of clusters to be detected. Experiments using simulation data and actual crime data confirmed the applicability of the proposed method and the necessity of careful results interpretation.

5. Acknowledgements

We thank the Metropolitan Police Service for providing the crime report data. This work was supported by JSPS KAKENHI (Grant No. JP26630233).

6. References

Duczmal, L., Kulldorff, M. and Huang, M. 2006. Evaluation of spatial scan statistic for irregularly shaped clusters. *Journal of Computational and Graphical Statistics*. **15**(2), pp.428-442.

Duczmal, L., Cançado, A., Takahashi, R., and Bessegato, L. 2007. A genetic algorithm for irregularly shaped spatial scan statistics. *Computational Statistics and Data Analysis*. **52**, pp.43–52.

Inoue, R. and Tsukahara, M. 2016. Spatial cluster detection imposing constraint on shape complexity in networks. *Theory and Applications of GIS*. **24**(1), pp.39-48 (in Japanese).

Kulldorff, M., and Nagarwalla, N. 1995. Spatial disease clusters: detection and influence. *Statistics in Medicine*. **14**, pp.799–810.

Kulldorff, M. 1997. A spatial scan statistic. *Communications in Statistics: Theory and Methods.* **26**(6), pp.1481-1496.

Tsukahara, M. and Inoue, R. Cluster detection in networks by controlling shape flexibility level. *Extended Abstract Proceedings of the GIScience 2014*. pp.441-445.