ISSN No. 0976-5697
Volume 1, No. 3, Sept-Oct 2010
International Journal of Advanced Research in Computer Science
RESEARCH PAPER
Available Online at www.ijarcs.info

# Crime Hotspot detection using spatial Clustering: A literature review of related work

M. Vijay Kumar*
Assistant Professor
Department of Computer Applications
KSR College of Engineering
Tiruchengode, India
vij370@gmail.com

Dr. C. Chandrasekar
Reader
Department of Computer Science
Periyar University
Salem, India

***Abstract*:** Hotspot is a special kind of clustered pattern. Like general clustered patterns, objects in the hotspot regions have high similarity in comparison to one another, and are quite dissimilar to all the objects outside the hotspot. One important feature that distinguish hotspot from general cluster is that the objects in the hotspot area are more active compared to all the others (density, appearance, etc). Hotspot discovery/detection in spatial data mining is a process of finding certain spatial regions where more events are likely to happen, or more objects are likely to appear, in comparison to the other areas.

***Keywords***: hotspots, spatial clustering, Scanning, Classification

## INTRODUCTION

Hotspot detection is mainly used in crime data analysis [2, 5] and disease data analysis. Crime data analysis aims at finding areas that have greater than average numbers of criminal or disorder events, or areas where people have a higher than average risk of victimization. Fig. 1 shows an example of crime hotspot pattern. In cancer/disease data analysis, hotspots of locations where disease are reported intensively are detected, which may indicate a potential breakout of this disease, or imply the underlying cause of the disease. Other domains of application include: transportation (to identify unusual rates of accidents along highways), and ecological science (to conduct geo-informatics surveillance for geospatial hot-spot detection)

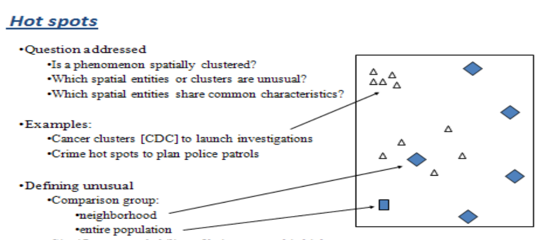


Fig. 1 An example of spatial hotspots

The pattern of spatial hotspot varies. Usually it is defined based on the level of the granularity and the shape of the region about which the user is concerned. For example, in the crime hotspot detection, the results may be given in the form of crime hotspot blocks, hotspot areas or hotspot cities. Also, the hotspots region may be location spots, street segments or block areas [1].

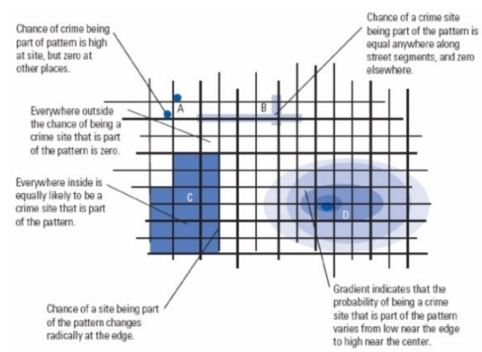


Fig. 2. An example of hotspots with different shapes

### *A. Classification of Work: By Hotspot Detection*

There are several types of widely used hotspot detection methods. They fall in three main catalogs based on different scales of hotspots they are looking for. Some methods in the classification are computational models that represent one or more applications with different settings. Those applications are not listed.

Table I: Classification of methods on the scale of hotspot they can find

| Method vs. Hotspot Scale | Global | Hierarchical | Local |
|---|---|---|---|
| K-means [6] | x | | |
| Moran's I [3] | x | | |
| NNH Clustering[4] | | x | |
| RNNH Clustering[4] | | x | |
| STAC [4] | | x | |
| LISA [7] | | | x |

*a) Global Hotspot Detection:* Global Hotspots are regions where activities are higher than other regions in the entire spatial area. It is the top level of hotspot scale. The following methods are able to detect such kind of pattern. The K-means algorithm creates k (user-defined number) clusters by partitioning the crime point data into groups. The routine finds the best positioning of the K centers and then assigns each point to the center that is nearest. Moran's I statistic works by comparing the value at any one location with the value at all other locations. Moran's I requires an intensity value for the crime point, which is often represented as the centroid of the geographic boundary area. This point is then assigned an intensity value. For crime applications, this most often is the count of crimes within that geographic area. The Moran's I result varies between –1.0 and +1.0. Where points that are close together have similar values, the Moran's I result is high.

*b) Hierarchical Hotspot Detection:* Hierarchical hotspot is a mixed pattern of different scales. It shows a correlation between local hotspots. The following figure shows that several hotspots near the mean road form a larger scale hotspot. The following methods are used to detect hierarchical hotspots.

Fig.3 An example of hierarchical hotspot pattern

The Nearest Neighbor Hierarchical (NNH) Clustering: Based on a nearest neighbor analysis technique analysis technique, crime incident locations are first grouped into nearest neighbor clusters containing a minimum number of point locations specified by the user. These first-order clusters are further grouped into larger, second-order clusters, and this process continues until no more clustering is possible and its variation.

The Risk-adjusted Nearest Neighbor Hierarchical Clustering (RNNH): It is a variation of NNH. Its goal is to find the concentration of incidents relative to a baseline. It achieves this goal by adjusting the threshold distance dynamically. The background data is represented as a fine grid using kernel density estimation, and this is used to adjust the threshold distance for clustering the original point set, on a cell-by-cell basis. Note that with both NNh and RNNh not all events are assigned to clusters, and each point is assigned to either one cluster at a given hierarchical level or none at all.

STAC: The STAC Hot Spot Area routine in *CrimeStat* searches for and identifies the densest cluster s of incidents based on the scatter of points on the map. The STAC Hot Spot Area routine creates a real unit from point data. It identifies the major concentrations of points for a given distribution. It then represents each dense area by the STAC is a scan-type clustering algorithm in which a circle is repeatedly laid over a grid and the number of points within the circle are counted.

*c) Local Hotspot Detection:*This is a small scale pattern. The hotspot itself may not be part of the global hotspot but it has a higher activity than surrounding areas. The following methods can be used for detecting this kind of pattern.

*d) LISA:* Local Indicators of Spatial Association. LISA statistics assess the local association between data by comparing local averages to global averages. For this reason they are useful in adding definition to crime hot spots and placing a spatial limit on those areas of highest crime event concentration It, thus, shares with those other scan routines the property of multiple tests, but it differ s in that the overlapping clusters are combined into larger cluster until there are no longer any over lapping circles. Thus, STAC clusters can be of differing sizes. The routine, therefore, combines some element s of partitioning clustering (the search circles) with hierarchical clustering (the aggregating of smaller clusters into larger clusters).

### *B. Another Classification of Related Work: By detection method*

A few literature reviews [2] studied the classification of related work based on how they generate hotspots. This is a quite different classification since it focuses on applications of hotspot detection. These applications may use computational models listed in the previous section.

*a) Kernel Estimation Method:* The Hot Spot Detective developed by Jerry Ratcliffe [8] program uses the raster grid scan approach to calculate a regular grid using the kernel estimation approach of Bailey and Gatrell *et al* (1996) that was developed specifically to address the point analysis hotspot problem in epidemiology. It generates an isodensity surface using the technique where a count is made on a regular grid of all points lying within a given search radius. LISA was used in computing the hotspot boundaries.

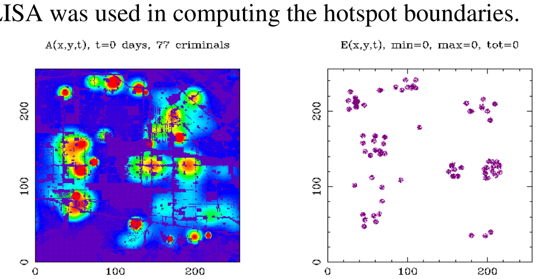


Fig.4. An example of Hot Spot Detective

*b) Tract or Grid Scanning:* Statistical Scan defines a scan window, and moves this window over the spatial region to calculate the density of events in the regions. By setting different window size, this method is able to find different scales of hotspots.

An example of this kind is SatScan [9]. This program is widely used to detect and evaluate disease outbreaks in time and space, and was developed by Martin Kulldorff.

It is available to download free of charge (www.satscan.org, 2006). An important factor considered by spatial scan statistical analysis is the time baseline. Recently the program has been modified to use variable ellipsoid areas in addition to circles (Kulldorff, 2006). SatScan generates circles of variable size and location to mark areas within which hotspots might be said to occur. The time baseline is included in the analysis. The underlying tract patterns are not circular, or even ellipsoidal; they are at bestamoeboid. SaTScan uses circles or ellipsoids based at the centroids of the tracts to calculate its statistics but cannot reflect the tract shape itself, or variable areal size of the tracts.

*c) Ellipsoid Scanning:* STAC and CrimeStat [4] used this method to generate hotspot. The patterns generated are ellipsoids of different scales. The first level clusters are ellipses centered, sized and oriented according to the hotspot calculation. These have then been aggregated into the much larger, but possibly weaker, hotspots shown at the second level. Computation models listed in the previous classification can be used to generate ellipsoids.

Fig.5.Ellipsoids generated using CrimeStat

\

Table II. Summarization of applications for hotspot detection

| App. Vs. How hotspots generated | Kernel Estimation | Tract grid scanning | Ellipsoid |
|---|---|---|---|
| Hot Spot Detective[8] | x | | |
| SatScan[9] | | x | |
| STAC/CrimeStat[4] | | | x |

## III CONCLUSION

The classification methods may be based on input data, neighborhood definition (i.e. spatial network rather than point sets), etc. In future work give further summarization in Spatial patterns may be discovered using Techniques like classification, associations, clustering and outlier detection New techniques are needed for SDM due to Spatial Auto-correlation, Continuity of space, Regional knowledge and also establishes a need for scoping Separation between spatial and non-spatial subspace in traditional approaches clusters are usually defined over the complete attribute space.

## IV. REFERENCES


[1] J.E.Eck, S.Chainey, J.G. Cameron, M.Leitner, R.E. Wilsons. Mapping Crime: Understanding Hotspots. NIJ Special Report, 2005.

[2] Michael J. McCullagh. Detecting Hotspots in Time and Space, University of Nottingham, 2006.

[3] N.A. Cressie. Statistics for Spatial Data (Revised Edition). Wiley, New York, 1993.

[4] Ratcliffe JH, The Hotspot Matrix: A framework for crime reduction, Police Practice and Research, Volume 5, Issue 1: 5-23.

[5] http://www.icjia.state.il.us/public/ index.cfm?metaSection=Data&metaPage=StacFacts, accessed 6-8-2006.

[6] J. Han, M. Kamber, and A. K. H. Tung. Spatial Clustering Methods in Data Mining: A Survey. In Geographic Data Mining and Knowledge Discovery. Taylor and Francis, 2001.

[7] Anselin, L. 1995. "Local Indicators of Spatial Association—LISA." Geographical Analysis, 27(2), 93-115.

[8] Ratcliffe J, Hot Spot Detective (HSD), http://jratcliffe.net/hsd/index.htm, *accessed 6-8-2006*.

[9] Kulldorff. M, Huang L, Pickle L and Duczmal L, An elliptic spatial scan statistic. Statistics in Medicine. 2006.