CRIME PATTERN ANALYSIS USING SPATIAL DATA MINING

R.Ragunath[1], P.Sathishkumar[2]
[1]PG Scholar, K.S.Rangasamy College of Technology, Tiruchengode, India.
Email: raghu251990@gmail.com, Mobile No: +91 9790426067.
[2]Assistant professor (Academic), K S Rangasamy College of Technology, Tiruchengode, India.
Email: psvsathishkumar@gmail.com

***Abstract*-- Crime pattern analysis (CPA) is the process of analytical reasoning facilitated by an understanding about the nature of an underlying spatial framework that generates crime. For example, law enforcement agencies may seek to identify regions of sudden increase in crime activity, namely, crime outbreaks. Many analytical tools facilitate this reasoning process by providing support for techniques such as hotspot analysis. However, in practice, police departments are desirous of scalable tools for existing techniques and new insights including, interaction between different crime types. Identifying new insights using scalable tools may help reduce the human effort that may be required in CPA. Formally, given a spatial crime dataset and other information familiar to law enforcement agencies, the CPA process identifies interesting, potentially useful and previously unknown crime patterns. For example, analysis of an urban crime dataset might reveal that downtown bars frequently lead to assaults just after bar closing. However, CPA is challenging due to: (a) the large size of crime datasets, and (b) a potentially large collection of interesting crime patterns. The chapter explores, spatial frequent pattern mining (SFPM), which is a spatial data driven approach for CPA and describes SFPM in the context of one type of CPA, outbreak detection.**

Keywords – Crime pattern, spatial crime dataset, spatial frequent pattern mining

I INTRODUCTION

Crime pattern analysis (CPA) is a key step employed by law enforcement and criminal justice agencies towards understanding the spatial environment that generates crime patterns. For example, the analysis of crime datasets with multiple crime types may reveal sudden increase in the activity of a subset of crime types in certain areas. This understanding provides insight into predicting future crime incidents and mitigates existing crimes.

The importance of CPA is clearly evident in the growth of spatial crime reports and other spatial information known to law enforcement. Rapid collection and archival of crime reports coupled with the growing analytical needs of law enforcement has given rise to a variety of tools including CrimeStat, ArcGIS 10 Spatial Statistics Toolbox, GeoDa, Rigel, SANET, SatScan etc.

However, the growing needs of law enforcement stresses scalable ways to generate meaningful crime patterns that may lead to hypotheses regarding the nature of crime as opposed to human driven enumeration of all possible hypotheses. For example, in a typical crime dataset containing 40 different crime types, there may be over 240 different patterns of association between different types. Enumerating all these patterns manually would be an arduous task even for trained analysts. Many police departments aim to accomplish crime mitigation and crime prevention with very few resources. However, the growth in the size and volume of crime datasets poses serious challenges. Hence, there is a growing need for scalable tools that can assist trained analysts and accomplish law enforcement goals with minimal resource allocation. CPA helps law enforcement planners accomplish this goal by identifying interesting, potentially useful, and non-trivial spatial patterns including, regions of sudden increase in crime activity, frequent co-occurrence of crime types around features such as bars and crime prone streets.

II BASIC CONCEPTS

Spatial frequent pattern mining (SFPM) is the process of discovering interesting, useful and non-trivial patterns from spatial datasets. Figure 2 shows a typical SFPM process that is based on the crime datasets collected and archived by law enforcement as the basis. The SFPM process usually begins with knowledge of criminological theories from environmental criminology. Based on these theories, analysts pose certain questions on the data.

A Spatial neighbourhood is a collection of related spatial entities such as crime reports.

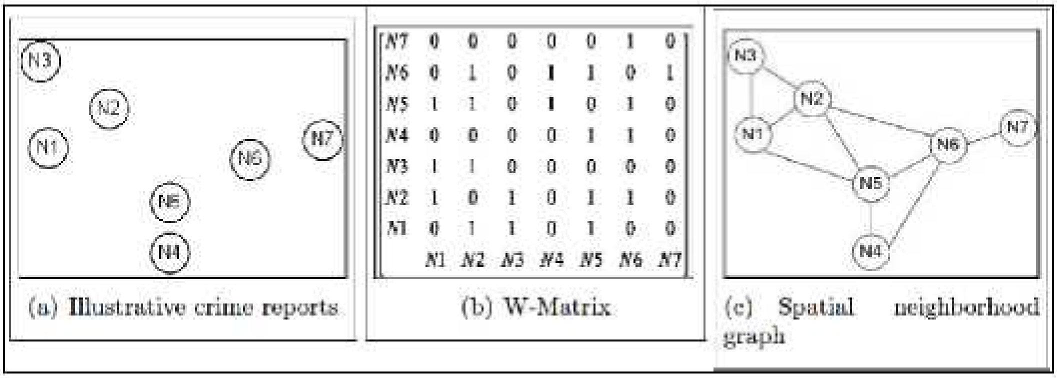

Figure 1 Spatial neighbourhood matrix and spatial neighbourhood graph

In most crime analysis applications involving crime reports, the most common type of neighbor relation is a distance-based relation. The application of a spatial neighbor relation on a collection of crime reports produces **a spatial neighborhood matrix**, commonly referred to as the **W-Matrix**. For example, Figure 1(a) shows an illustrative crime report dataset showing different crime reports represented as circles with labels **N**1, **N**2 etc. Application of a neighbor relation based on a distance threshold (e.g., 1 mile, 2 mile etc.) produces a spatial neighborhood matrix as shown in Figure 1(b). The matrix in this figure consists of 0s or 1s to represent the absence or presence of a neighbor relationship. Figure 1(c) shows an alternative representation of the W-Matrix called **the neighborhood graph**, where the edges represent the presence of a neighbour relation and the nodes represent the crime reports.

Given spatial crime data, neighbor relationships, environmental criminology theories and other inputs known to law enforcement, SFPM employs several techniques to identify interesting, useful and non-trivial crime patterns. One such technique is **Regionally frequent crime pattern (RFCP) discovery**.

RFCPs represent collections of spatial features and crime types frequently associated with each other at certain localities. For example, the RFCP, <*(Bar, Assaults), Downtown* > indicates that a frequent pattern involving assaults and bars is often localized in downtown regions. Given feature types2 (e.g., Bars), crime types and their geo-located instances, along with a spatial neighborhood size and a likelihood threshold, the RFCP discovery process finds all interesting RFCPs. The local fraction of instances of any crime type participating an RFCP is measured using a **Regional conditional probability (RCP).**

III CRIME OUTBREAK DETECTION

Figure 2 shows an overview of the proposed approach to detect crime outbreaks. The Lincoln crime dataset from the year 2007 contains crime reports with multiple crime types and several feature types such as bars. In addition, the user also provides a sense of spatial neighborhood (in terms of the maximum neighborhood size, typically 0.5 miles, 1 mile etc.). Based on these inputs, the first stage, computes crime type level counts around each spatial feature. Once these counts are obtained, we make use of the Multinomial scan statistic test implemented in the SatScan program. The multinomial scan statistic routine in SatScan computes several interesting measures including, (a) spatial features that may be a part of significant outbreaks, (b) Radius of a possible crime outbreak cluster around each feature and (c) The risk of occurrence for each crime type within each outbreak cluster. The risk value for each crime type corresponds to the ratio of, the number of cases the crime type within the radius of the feature to the number of cases of the crime type expected to occur based on a standard multinomial distribution. In our analysis we consider only crime types that have a high risk of occurring within the neighborhood of the spatial feature(i.e. risk > 1). A detailed description of the notion of risk and its interpretation can be found in the SatScan manual. To ensure that only crime types participating in statistically significant outbreak are highlighted, SatScan performs Monte Carlo simulation and computes p-values. Assuming a standard significance level of 0.05, only significant spatial features and crime types that have a high-risk level within their neighborhood are retained after the multinomial scan statistic test.

The results of the multinomial scan statistic test include, crime types that may be involved in one or more interesting crime outbreaks. However, it is still important to extract the actual regions of these outbreaks. This goal accomplished via RFCP discovery. The RFCP discovery process requires other inputs including interestingness thresholds, and spatial neighborhood information. Based on this the RFCP process reports all interesting RFCPs of crime types that may participate in one or more outbreaks. Since RFCPs also include the actual location of the crime report, they provide an enhanced spatial view (e.g. convex polygon) of the actual outbreak as opposed to a simple circular neighborhood.

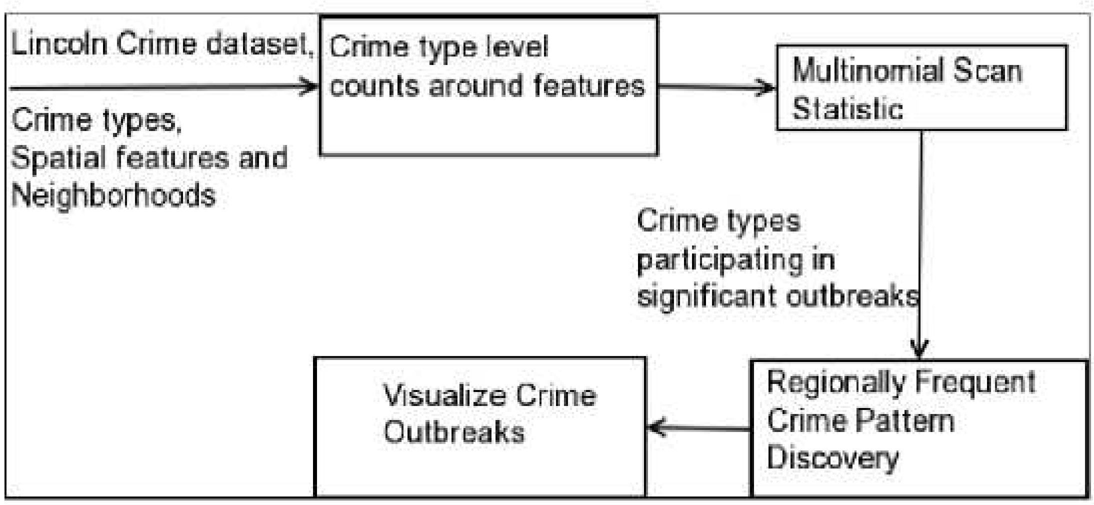

Figure 2 Overview of proposed approach

IV OUTBREAK DETECTION AT MULTIPLE ANALYSIS SCALES

Handling spatial scale has been an open research challenge in many GIScience applications. In crime outbreak detection one of they key inputs specified by analysts is the spatial neighborhood size. The results of the any analysis are sensitive to the neighborhood size specified by the user. Particularly, the crime type level counts are sensitive to the spatial neighborhood size. Also, the RFCP process requires a spatial neighborhood size as an input. This makes any spatial analysis technique sensitive to spatial scale. For example, Figure 3 shows a simple scenario for detecting outlier buildings from a subset of the Lincoln, NE dataset using their area attribute.

Figure 3(a) and (b) are the analysis performed at different spatial scales, namely, two nearest neighbors and eight nearest neighbors respectively. Outlier analysis using the first scale, two nearest neighbors, highlights the building ***B1*** as anomalous (Figure 3(a)), whereas, when we increase the scale of analysis, building ***B2*** is flagged as anomalous (see Figure 3(b)). New techniques are needed that can perform SFPM at multiple analysis scales.

New research is also needed to explore the use of frequent patterns such as CSTPs to drive models that can predict future crime. With the recent interest in predictive policing, pursuing this direction may help in enhancing intervention strategies via effective preparedness.

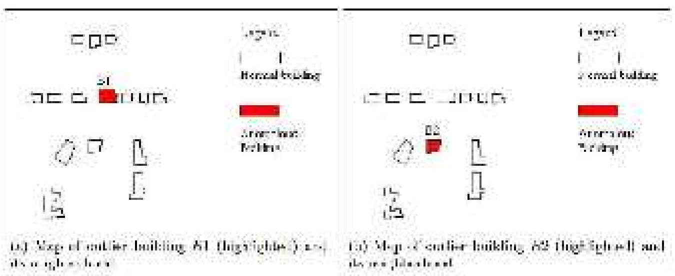

Figure 3 Map of two outlier buildings in Lincoln crime dataset

V CONCLUSION

The chapter explored Spatial Frequent Pattern Mining (SFPM), which is a data driven approach to crime pattern analysis. It identified the benefits of data driven approaches in the face of large spatio-temporal crime datasets and highlighted that they are useful in reducing human effort. Hypotheses regarding real world phenomena pertaining to crime can be generated only after analysts have evaluated the results of the SFPM process. Hence, SFPM simply reduces the effort an analyst might have to undertake to formulate a meaningful hypothesis regarding the nature of crime patterns. Pressing research needs, including new SFPM methods for outbreak detection that account for crime distributions along street networks and analysis across multiple scales were identified with specific examples.

VI REFERENCES

[1] Lincoln City Crime Dataset. (2007) Lincoln City police department, Lancaster County, NE, USA.

[2] Arcgis desktop edition 10. (2010) Environmental Systems Research Institute(ESRI) Redlands, CA, USA.

[3] Anselin. L. (2008) Spatial econometrics. A companion to theoretical econometrics, pages 310-330, .

[4] Anselin. L, Cohen. J, Cook. D, Gorr. W and Tita. G. (2000) Spatial analyses of crime. Criminal Justice : Measurement and analysis of crime and justice, 4.

[5] Anselin. L, Syabri. I, and Kho. Geoda. Y: (2006) An introduction to spatial data analysis. Geographical Analysis, 38(1):5-22.

[6] Beavon. D, Brantingham. P, and Brantingham. P. (1994) The influence of street networks on the patterning of property offenses. Crime prevention studies, 2:115-148.

[7] Beck.C and McCue. C. (2009) Predictive Policing: What Can We Learn From Wal-Mart and Amazon About Fighting Crime in a Recession? The Police Chief, 76(11).

[8] Brantingham. P and Brantingham. P. (1981) Environmental criminology. Sage Publications Beverly Hills, CA.

[9] Brantingham. P and Brantingham. P. (1984) Patterns in crime. Macmillan.

[10] Brantingham. P and Brantingham. P. (1993) Nodes, paths and edges: Considerations on the complexity of crime and the physical environment. Journal of Environmental Psychology, 13(1):3-28.

[11] Brantingham. P and Brantingham. P. (1995) Criminality of place: Crime generators and crime attractors. European Journal on Criminal Policy and Research, 3(3):5-26.

[12] Brantingham. P, Brantingham. P, and Glasser. U. (2005) Computer simulation as a research tool in criminology and criminal justice. Criminal Justice Matters, 58:19-20.

[13] Brantingham. P, Brantingham. P, Vajihollahi. M, and Wuschke. K. (2009) Crime analysis at multiple scales of aggregation: a topological approach. Putting Crime in its Place, pages 87-107.

[14] Brantingham .P, Glasser .U, Jackson. P, and Vajihollahi. M (2009) Modeling criminal activity in urban landscapes. Mathematical Methods in Counterterrorism, pages 9-31.

[15] Celik. M, Shekhar. S, George. B, Rogers. J, and Shine. J. (2007) Discovering and quantifying mean streets: A summary of results.

[16] Chainey. S and Ratcliffe. J. (2005) GIS and crime mapping, volume 6. John Wiley & Sons Inc.