

Analyzing and Displaying of Crime Hotspots using Fuzzy Mapping Method

Ranbir Kaur

PG Student, Department of Computer Science
Guru Nanak Dev Engineering College
Ludhiana, India

Sukhjot Singh Sehra

Asst. Professor, Department of Computer Science
Guru Nanak Dev Engineering College
Ludhiana, India

ABSTRACT

Pattern detection is one of the essential challenges in crime mapping and analysis. Data mining can be used to explore crime detection problems. A cluster technique is an effective method for determining areas with high concentrations of localized events. Conversely, it remains a particularly demanding task to detect hotspots with mapping methods in view of the vulnerability connected with the suitable number of groups to create and additionally securing significance of individual clusters identified. Fuzzy clustering means algorithm was used for identifying hotspots of Chicago police department's citizen law enforcement analysis and reporting system data. In fuzzy clustering, a membership value to each data is assigned, which indicate the strength of relationship between that data points and a specific cluster. In this study each cluster represented the group of global positioning system data points having latitude and longitude as their coordinates. The findings from this study were expected to aware the public about crime hotspots. Law enforcement agencies can take prior steps to prevent crime with the use of detected crime hotspots.

General Terms

Pattern Recognition, Spatial Analysis, Clustering

Keywords

Hotspots, crime hotspots, fuzzy clustering mean, cluster analysis, pattern recognition, and point pattern analysis.

1. INTRODUCTION

The clusters are groups of similar data and the degree of relationship is strong between data members of the same cluster and weak between data members of different clusters. There exist many fuzzy clustering methods used in various applications such as pattern recognition, medical diagnosis, civil engineering, and image processing and web retrieval. In spatial analysis many clustering algorithms are used to determine the hotspots for crime analysis, car crash analysis, disease tracking, etc. Hotspot analysis is the process of finding unusually dense event clusters through space and time. Many crime justice services agencies are discovering the benefits obtained by computer technologies to identify crime hotspots with a specific end goal to take preventive techniques such as deploying saturation patrols. For example, in crime analysis [1, 6–8, 11] the FCM method [2– 4] is used for determining areas with high concentrations of crime, the feature vectors are the crime event points geo-referenced on the geographic map [9, 10].

A spatial statistical tool, Crime-STAT [12] has developed by the National Institute of Justice at Washington DC (USA) for the GIS analysis of crime incident locations and this tool

includes the use of many clustering methods, among others, also hotspot analysis [13,14].

A hotspot is generally defined as an area containing dense clusters of localized events (e.g., criminal incidents). Identifications of hotspots on the geographic map are usually made by geo-referencing as points the events happened in a certain period. One of the most known approaches for the detection of hotspots is to use a cluster technique, which is an efficient method for deciding areas showing raised concentration of limited events.

2. THE FCM ALGORITHM

Fuzzy clustering means (FCM) is an information bunching strategy in which a dataset is gathered into n bunches with each information point in the dataset having a place with each group to a certain degree. Fuzzy clustering means (FCM), otherwise called fuzzy ISODATA, is an information grouping calculation in which every information point fits in with a bunch to a degree defined by a participation grade. Bezdek proposed this algorithm in 1973 as an improvement over earlier Hard C-means (HCM) clustering. FCM partition a collection of n vector $X_j, z = 1, \dots, n$ into c fuzzy groups, and finds a cluster center in each group such that a cost function of dissimilarity measure is minimized. The major difference between FCM and HCM is that FCM use fuzzy partitioning such that a given data point can belong to several groups with the degree of belongingness specified by membership grades between 0 and 1. To accommodate the introduction of fuzzy partitioning, the membership matrix U is allowed to have elements with values between 0 and 1. However, imposing normalization stipulates that the summation of degrees of belongingness for a data set always be equal to unity:

$$\sum_{i=1}^c U_{ij} = 1, \forall j = 1, \dots, n \quad \dots\dots\dots 2.1$$

The cost function (or objective function) for FCM is then a generalization of Equation 2.1:

$$J(U, c_1, \dots, c_c) = \sum_{i=1}^c J_i = \sum_{i=1}^c \sum_j^n U_{ij}^m d_{ij}^2 \quad \dots\dots\dots 2.2$$

Where U_{ij} is between 0 and 1; C_i is the cluster center of fuzzy group i; $d_{ij} = \|C_i - X_j\|$ is the Euclidean distance between ith cluster center and jth data point; and $m \in G [1, \infty)$ is a weighting exponent.

The necessary conditions for equation 2.2 to reach a minimum can be found by forming a new objective function J as follows:

$$\begin{aligned} \bar{J}(U, c_1, \dots, c_c, \lambda_1, \dots, \lambda_n) &= J(U, c_1, \dots, c_c) + \sum_{j=1}^n \lambda_j \left(\sum_{i=1}^c U_{ij} - 1 \right) \\ &= \sum_{i=1}^c \sum_{j=1}^n U_{ij}^m d_{ij}^2 + \sum_{j=1}^n \lambda_j \left(\sum_{i=1}^c U_{ij} - 1 \right) \end{aligned} \quad \dots\dots\dots 2.3$$

Where $X_j, j = 1$ to n , are the Lagrange multipliers for the n constraints in Equation 2.1. By differentiating $J(U, C_1, \dots, C_c, A_1, \dots, A_n)$ with respect to all its input arguments, the necessary conditions for Equation 2.3 to reach its minimum are

$$c_i = \frac{\sum_{j=1}^n U_{ij}^m X_j}{\sum_{j=1}^n U_{ij}^m}, \quad \dots\dots\dots 2.4$$

And,

$$u_{ij} = \frac{1}{\sum_{k=1}^c \left(\frac{d_{ij}}{d_{kj}} \right)^{2/(m-1)}} \quad \dots\dots\dots 2.5$$

The fuzzy C-means algorithm is simply an iterated procedure through the preceding two necessary conditions. In a batch-mode operation, FCM determines the cluster centers C_i and the membership matrix U using the following steps:

Step 1: Initiate with the membership matrix U with random values ranges from 0 to 1 such that the constraints in Equation 2.1 are satisfied.

Step 2: Find ‘ c ’ fuzzy cluster centers $C_i, i = 1, \dots, c$, using Equation 2.4.

Step 3: Compute the cost function using Equation 2.2. Stop if either it is below a certain tolerance value or its improvement over previous iteration is below a certain threshold.

Step 4: Compute a new U using Equation 2.5. Go to step 2.

The centers of cluster can also be first initialized and then the iterative procedure carried out.

3. TEST WITH CRIME DATA

3.1 Repository

The dataset taken in this study is reported incidents of crime that occurred in the city of Chicago. Data are extracted from the Chicago Police Department's CLEAR (Citizen Law Enforcement Analysis and Reporting) system. These crimes may be based upon preliminary information supplied to the Police Department by the reporting parties that have not been verified. The preliminary crime classifications may be changed at a later date based upon additional investigation and there is always the possibility of mechanical or human error. Every week this site is updated so this data is according to the crime held in actual. But method of collecting this data is bit complicate and the cases which are not registered in police stations are not included in this data. This website of Chicago has data points from 2001 to till date (except last

seven days). Every week data is updated with new entries. For this study we have taken data for year 2013 of month August.

The variables collected in this data set included ID, Primary type, ward, X coordinate and Y coordinate. Because of the very large size of the data set, we have decided to select a more limited subset of data for analysis. A subset of data was also chosen based on type of crime. This study only focuses on primary type of crime that which kind of crime is happened percentage wise. It divides crime into two major categories Sensitive, Materialistic types. As the purpose of the study was a comparison between analytical techniques, it made sense to focus on a limited subset of crime data. On the basis of this, crimes can be divided into following categories:

3.1.1 Sensitive Crimes

An expressive or sensitive crime has no purpose except to achieve the action that is desired by the criminal. Sexual assault, intra-family sexual assault, child abuse, murder, harassment, stalking etc. are the examples of sensitive crimes or expressive crimes. They really do nothing but allow the perpetrator to express his/her wishes/desires.

Table 1 Sensitive Crimes

| Primary type | Code given | Total count per 210 | Percentage |
|--------------------|------------|---------------------|------------|
| Assault | 1 | 22 | 10.4 |
| Deceptive Practice | 4 | 31 | 14.7 |
| Sex Offence | 7 | 0 | 0 |

3.1.2 Materialistic Crimes

Materialistic crimes, also termed instrumental crimes. An instrumental crime, are motivated to achieve a tangible goal, such as obtaining a physical good through theft. For example, a person may commit a theft, or sell drugs to get money to buy themselves a big house, or a fancy car, etc., that other people would get through their own efforts and hard work.

Table 2 Materialistic Crimes

| Primary Type | Code Given | Total count per 210 | Percentage |
|---------------------|------------|---------------------|------------|
| Burglary | 2 | 4 | 1.9 |
| Robbery | 8 | 4 | 1.9 |
| Theft | 9 | 106 | 50.4 |
| Motor Vehicle Theft | 10 | 6 | 2.8 |

3.1.3 Miscellaneous Crime

Three crime types, criminal Damage, Gambling and kidnapping offences do not exhibit traits of either sensitive or Materialistic crimes as they are not intended to express emotion or to obtain a tangible good. These are related to social issues of a city.

Table 3 Misc. Crimes

| Primary Type | Code Given | Total count per 210 | Percentage |
|-----------------|------------|---------------------|------------|
| Criminal Damage | 3 | 42 | 20 |
| Gambling | 5 | 0 | 0 |
| Kidnapping | 6 | 0 | 0 |

Materialistic and other crime type rates were calculated using the sum of residential and working population as the denominator because the sum of populations is the best indicator for the total number of at risk targets. We assume that the nature of sensitive crimes often takes offenders outside their residential or employment location to areas where there are many non mobile at risk targets, so working or residential populations alone are not appropriate. For mobile targets such as motor vehicles, the number of targets was not available so we assume that the sum of residential and working population provides a reasonable proxy indicator for the total number of motor vehicles in a census tract.

4. RESULTS

Fuzzy Clustering means (FCM) method is used to display crime hotspot with respect to given crime data. The dataset taken in this study is reported incidents of crime that occurred in the City of Chicago which is composed of 210 data points with Crime type, longitude, latitude, and ward numbers. In figure 1, Graph is shown where latitude is showing on X-Axis and longitude values on Y-Axis.

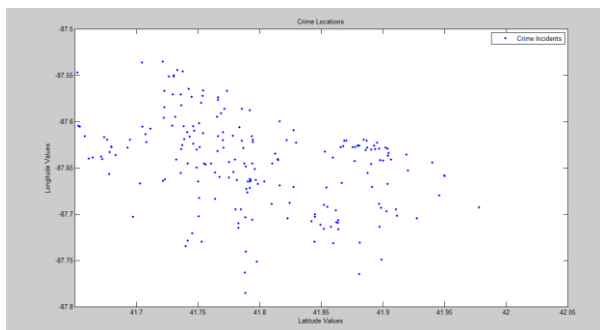


Figure 1 Data Points w. r. t. latitude and longitude

As the result of the above algorithm we have found the following hot spots. In figure 2, various crime incidents are shown in blue dots and crime hot spots are depicted in red circles with respect to longitude and latitude points.

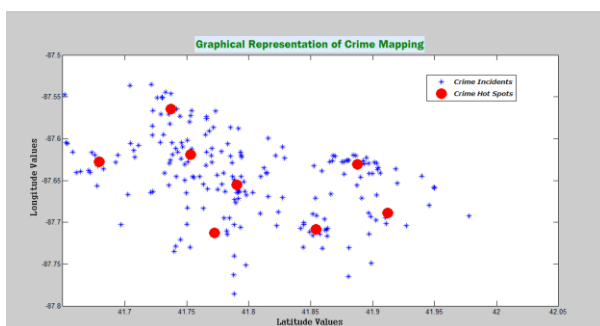


Figure 2 Crime hotspots w. r. t. latitude and longitude

Above algorithm can be used with different combination. In figure 3, Crime type and ward numbers are shown on X-axis and Y-axis respectively.

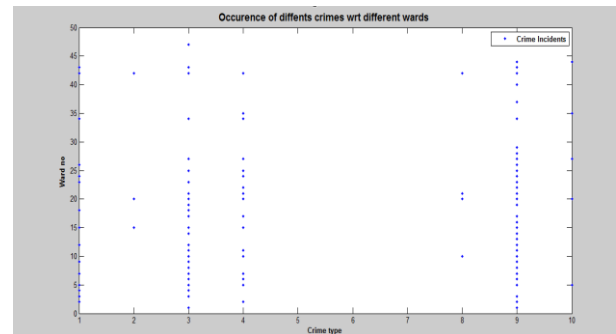


Figure 3 Data Points w. r. t. crime type and ward number

From the following figure 4, we come to conclusion that crime types 3- CRIMINAL DAMAGE and 9-THEFT are prominent crime type according to given data.

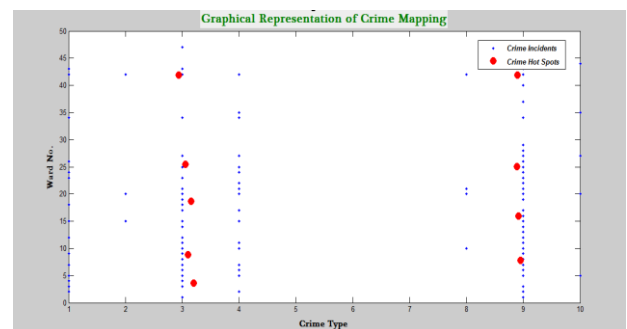


Figure 4 Crime hotspots w. r. t. crime type and ward number

5. CONCLUSION

There are several factors on which hotspots rely, varying from theory selection, to the nature of crime being analyzed, to the display of output results. Therefore it is important to note that carrying out analysis must have a logical and systematic approach. This work focused on the detection of the crime hot-spots where crime happened at the highest level. The findings from this study are expected to help police departments to lessen the crime rate. With this they will get the locations and can take prior steps to prevent crime.

6. REFERENCES

- [1] Brantingham, P. 1981, Environmental Criminology, Sage Publications, Los Angeles.
- [2] Bezdek, J. C. 1974, Cluster validity with fuzzy sets, IEEE Journal of Cybernetics, Vol. 8, 3, pp. 58–73.
- [3] Bezdek, J.C. 1981, Pattern Recognition with Fuzzy Objective Function Algorithms, Plenum Press, New York.
- [4] Bezdek, J. C., Ehrlich, R. and Full, W. 1984, The fuzzy C-means clustering algorithm, Computers and Geosciences Vol. 10, pp. 191–203.

- [5] CLEAR (Citizen Law Enforcement Analysis and Reporting, City of Chicago , Data Portal, 2011, (<https://data.cityofchicago.org/Public-Safety/Crimes-2001-to-present/ijzp-q8t2>)
- [6] Cohen, L. and Felson, M. 1979, Social change and crime rate trends: a routine activity approach, *American Sociological Review* Vol. 44, pp. 588–608.
- [7] CMRC (Crime Geographic Mapping Research Centre), Hotspot project, 1998, (www.ojp.usdoj.gov/cmrc/whatsnew/hotspot/toc.html).
- [8] Craglia, M., Haining R. and Wiles, P., 2000, A comparative evaluation of approaches to urban crime pattern analysis, *Urban Studies*, Vol. 37, 4, pp. 711–729.
- [9] Greenburg, S. and Rohe, W., 1984, Neighborhood design and crime, *Journal of the American Planning Association* Vol. 50, pp. 48–61.
- [10] Gordon, A. D., 1998, How many clusters? An investigation of five procedures for detecting nested cluster structure, in: *Data Science, Classification and Related Methods*, C. Hayashi, N. Ohsumi, K. Yajima, Y. Tanaka, H. Bock and Y. Baba, eds, Springer-Verlag, Tokyo, pp. 109–116.
- [11] Harries, K., 1999, *Geographic Mapping Crime: Principle and Practice*, National Institute of Justice, Washington DC.
- [12] Levine, N., 1999, *CrimeStat: A spatial statistics program for the analysis of crime incident locations*, version 1.1, Ned Levine and Associates, National Institute of Justice, Washington DC.
- [13] Sneath, P. H. A., 1957, The application of computers to taxonomy, *Journal of General Microbiology*, Vol. 17, pp. 201–226.
- [14] Wang, F., 2005, *Geographical Information System and Crime Analysis*, Idea Group Publishing, Hershey.