

Crime Hot Spot Analysis and Prediction Using Spatial Data Mining Approaches - A review

S.Sivaranjani ¹ and Dr.S.Sivakumari ²

*Dept. of Computer Science & Engg.,
Faculty of Engineering,
Avinashilingam Institute for Home Science and Higher Education for Women,
Coimbatore-641 108, India.
¹sivaranjanicse@gmail.com
²hodcseau@gmail.com*

Abstract

Criminology is a process that targets to find crime characteristics. Crime analysis is a part of criminology which includes discovering and identifying crimes and their relationships with criminals. The difficulty in analyzing huge volume of crime data and finding their relationships have made criminology an appropriate field for applying data mining techniques. The knowledge obtained by applying data mining technique find very useful in supporting police force to identify the location of crime incidents. The aim of this paper is to study various applications of data mining techniques for the finding and prediction of future crime and in turn helps the police force by providing accurate prediction for the location of the criminal.

Keywords Crime analysis, Hot spots, Clustering, Classification, Geographic Information Systems (GIS).

Introduction

Spatial data mining refers to the abstraction of knowledge, spatial relationships, or other remarkable patterns not explicitly stored in spatial databases [1,2]. It can be used for understanding spatial data, discovering spatial relationships and relationships between spatial and non-spatial data, constructing spatial knowledge bases, restructuring spatial databases, and optimizing spatial queries. It is probable to have wide applications in geographic information systems (GIS), geo-marketing, remote sensing, image database exploration, medical imaging, navigation, traffic control, environmental studies, and numerous other areas where spatial data are used. However, extracting interesting and valuable patterns from spatial databases is much

more difficult than extracting corresponding patterns from traditional numeric and characterized data due to the complexity of spatial data types, spatial relationships, and spatial autocorrelation. Spatial outlier detection, Co-location pattern discovery, spatial classification and regression modeling, spatial clustering, spatial hotspot analysis are the significant tasks in data mining. Among this Hotspot analysis plays an important role in how law enforcement agencies address crime in problematic areas which capture us to take up this task for our study.

Hotspots are a distinct type of clustered pattern. As in clustered patterns, objects present in hotspot regions have high similarity when compared with one another and are relatively dissimilar to all the objects present outside the hotspot. The objects present in the hotspot areas are more active and dense when compared with the other area which makes the difference between ordinary cluster and hotspot. Hotspot analysis is a spatial data mining process in which we discover the regions where the crime frequently happens. This analysis is broadly used in the detecting crime and disease spots.

Van Eck N J and Waltman L [3] has presented a study on crime data analysis which targets at identifying areas that have crime incidents more than average, or finding out areas where people tend to have higher risk of victimization than average count. The design of hotspot maps is mainly oriented toward aiding law enforcement to make appropriate assignment of their resources for crime investigation.

Levine N [4] made use of clustering algorithms for spatial hotspot analysis. The algorithms which include *K*-Means, hierarchical clustering, etc. Anselin L, Getis A [5] used STAC (spatio-temporal analysis of crime) and LISA (local indicators of spatial association) method to help law enforcement agencies for crime justification.

Spatial analysis tools and techniques

The SDM techniques, such as spatial clustering, spatial hotspot analysis, and Co-location mining, are most famous approach which found their way into commercial products such as Murray Cea used Oracle [6], Scott L, Janikas M used ArcGIS [7] for analysis and mapping the hot spot areas. Further than these commercial products, there are many open source tools such as GeoDA used by Anselin L [8] in his paper for the geographical analysis; Levine N [4] used CrimeStat spatial statistics program for the analysis of crime incident locations.

Table 1: Spatial analysis software

Technique employed in spatial analysis	Software used for the spatial analysis
Co-location mining	Oracle 10g
Spatial clustering	ArcGIS 9.3, Oracle 10g, CrimeStat, Terra Seer
Spatial hotspots	ArcGIS 9.3, CrimeStat, GeoDa,
Spatial outliers	ArcGIS 9.3, GeoDA
Spatial network hotspots	CrimeStat, SANET
Kriging	ArGIS 9.3, S+ Spatial Stats, fields package and geoR in R
Spatial autoregression	S+ Spatial stats, GeoDa
Conditional autoregression	CrimeStat
Geographically weighted regression	ArcGIS 9.3

Keyvanpour M R, Javideh M and Ebrahimi M R [9] proposed an approach for crime analysis and crime matching process by extracting important entities from police narratives reports by applying Self-Organizing Map(SOM) clustering method.

Shyam Varan Nath S [10] implemented data mining frame with the geospatial design of crime which helps detectives in increasing the security for the homeland by improving the productivity of Law enforcement officers. Author made use of K-means clustering to identify the crime patterns very effectively and increased the speed in the process of resolving the crimes.

Manish Gupta, Chandra B and Gupta M P [11] highlighted the current systems used by Indian police as e-governance initiatives and also suggested an interactive query based interface as crime analysis tool to support police in their activities. The proposed system is used to extract useful information from the massive crime database which was maintained by National Crime Record Bureau (NCRB) and clustering was employed to find crime hot spots. The recommended crime analysis tool proposed by the author for Crime Criminal Information System (CCIS) is very easier to implement and use. The latest visualization techniques can be assimilated with crime analysis tool for enhancing the better understanding of the results and patterns which in turn provides an active tool for crime detection and crime prevention. Geographical Information System (GIS) tool plays a major role in crime analysis and found a wider application in Law of enforcement.

Jonathan J. Corcoran, Ian D. Wilson and Andrew Ware J [12] introduced the approaches for crime forecasting around geographical areas of concern that excel the policing boundaries. The author developed a practical solution for use in policing environment. Geographical crime incidence- scanning algorithm was used to identify the clusters. The clusters can be analysed using Gamma Test (GT). Using the output from GT two techniques were implemented such as ANN and Regression methods.

Anna L. Buczak, Christopher M and Gifford [13] applied fuzzy association rule mining for community crime pattern discovery. This was the first experimental study

done by applying fuzzy association rule mining to a crime data set for crime detection. The author suggested that further analysis is required to gain knowledge about crime spot at both the community and national levels. It aids the law enforcement people to find out significant crime patterns easier rather than wasting their time in uncommon patterns.

Tony H. Grubestic and Alan T.Murray [14] suggested that the cluster analysis plays an important role in the detection of hotspots by the usage of software packages such as SPSS, SPLUS, SAS, CrimeStat. The author discussed the problematic aspects of optimization based cluster analysis in crime hotspot detection.

Aniruddha Kshirsagar and Lalit Dole[15] accomplished crime uncovering in both transaction domain and in application domain by using data mining methods. The author made use of outlier detection in the domain of credit card application and transaction to overcome fraud identity.

Hsinchun Chen *et al.*, [16] presented an outline about crime data mining along with four case studies. The case studies being done here was, Entity Extraction for Police Narrative Reports, Detecting Criminal Identity Deceptions: An Algorithmic Approach, Authorship Analysis in Cybercrime, and Criminal Network Analysis. The efficiency and effectiveness of the crime and intelligent analysis can be improved by using data mining methods.

Divya Bansal and Lekha Bhambhu [17] applied association rule mining in mining pattern that occur frequently within a dataset. Association algorithm can be applied to the transformation of quantitative data into qualitative data. The author compares the result of two algorithms namely Apriori and Predictive Apriori in mining association rules for the datasets of crime against women collected from session court by using WEKA.

Devesh Bajpai [18] discussed about the crime network analysis to extract the useful information by using the recent techniques like Social Network Analysis, Entity Extraction, String Comparator, etc. The data mining techniques can be applied in the area of crime investigation for decision support, prediction, forecasting etc., The author stated that the usage of artificial intelligence tools for enhanced crime analysis.

Jitendra Kumar, Sripati Mishra, Neeraj Tiwari [19] presented a study in identifying the area which is consisting of major crime named hotspots and the area with minimum crimes. The author applied clustering techniques for the identification of hotspots and safe zone of crime for the data collected from State Crime Record Bureau, Uttar Pradesh.

John David Elijah Sandig *et al.*, [20] offered an online Graphical Information System (GIS) for crime rate and models using frequent pattern analysis. It was a web-based system that contains GIS for robbery, homicide and physical injury incidents within a (Iloilo) City. The frequent happening of the crime and its occurrence time have been identified by using this system. The frequent victims in the community can be identified. The author made use of Google Heat Maps for plotting the crimes.

Hangman *et al.*, [21] offered the prominence of data mining technology to design active services to decrease crime rates. For the police system by using the criminal information it's hard for crime prevention and forecasting. Data mining

techniques can be employed for efficient prediction and investigation.

TongWang *et al.*, [22] suggested a pattern detection algorithm called Series Finder to find the common features of all crime patterns and the distinctive aspects of each specific crime pattern which are committed by the same criminal or groups of criminals. A comparison was done between the series finder algorithm and clustering methods for crime analysis.

Dawei Wang *et al.*, [23] introduced a spatial data mining framework to study crime hotspots using Geospatial Discriminative Patterns (GDPatterns) to find the significant difference between hotspots and normal areas in a geo-spatial dataset. By utilizing the GD patterns the author developed a unique model called Hotspot Optimization Tool (HOT) for the improved efficiency in identification process of crime hotspots. The real world crime dataset collected from a northeast city in the United States were used for the study. GDPattern clusters were grouped and visualized based on the similarity measure.

Kadhim B. Swadi Al-Janabi [24] proposed a framework for crime data analysis and detection using Decision tree Algorithms for data classification and data clustering was done by using K-Means algorithm. The author tends to help the data specialists in forecasting, discovering patterns, mapping criminal networks and identifying possible suspects. The results obtained from both clustering and classification is used for prediction of crimes and criminals.

Donald E. Brown [25] applied two methods called data fusion and data mining. Since the law enforcement agencies are handling huge amount of crime data, we have to convert those data into useful information. In order to combine, interpret, organizing and to remove noisy backgrounds of crime data from several resources the author made use of data fusion technique.

Data mining technique was employed to discover patterns and their relationships with large databases. The author builds ReCAP (Regional Crime Analysis Program), by combining data fusion and data mining to provide crime analyses in large databases.

Thongtae P and Srisuk S [26] did a survey about the data mining methods for crime data analysis for its efficiency and effectiveness. The author discussed about the difficulty in detecting crime from huge amount of criminal data sets which comprises of daily activities of the criminals. They proposed some guidelines to overcome the problem in identifying the suitable algorithm for knowledge discovery.

Jerry H. Ratcliffe and Michael J. McCullagh [27] present a theoretical framework for the temporal analysis of aoristic crime data. This framework was being used for observing the change in crime patterns over time and applied to area which is out of police boundaries. The author depicts that the crime data frequently lacks temporal definition, so two temporal methods called rigid temporal search and aoristic methods were used to predict the temporal possibilities.

Jerry H. Ratcliffe [28] established a technique that uses police START and END crime times to create a crime occurrence probability which can be mapped or visualized graphically. A study in the eastern suburbs of Sydney, Australia, exposes that crime hotspots with a geographical proximity can have specifically dissimilar temporal patterns.

Chen, H *et al.*, [29] applied traditional data mining techniques such as classification, association analysis, prediction, outlier analysis and cluster analysis to find the crime patterns from both structured and unstructured crime data. Crime data mining increases privacy concerns when compared with other forms of data mining, however, many researchers have developed various automated data mining techniques for both local law enforcement and national security applications.

Jerry H. Ratcliffe [30] summarized the data mining techniques used to identify the spatial and temporal components of crime hotspots. Temporal hotspots and spatial hotspots were identified by using data mining techniques and they were described in hotspot matrix format. The author concluded with a suitable crime prevention or detection policy by using the hotspot matrix method applied in real examples to show how effectively we can combine the spatial and temporal characteristics.

Kate J. Bowers, Shane D. Johnson and Ken Pease [31] explored the development of a mapping technique that seeks to produce 'prospective' hot-spot maps. Novel map evaluation method was proposed and some standardized metrics that can be derived from maps are compared for its efficiency. The proposed predictive mapping technique was found more advantage than other traditional methods and might ascertain predominantly useful in the shift-by-shift positioning of police personnel.

Anthony A. Braga [32] made a review that attentive police force can prevent crime in great way.

The study also suggests that focused police actions at exact locations do not necessarily result in crime displacement. Although these estimations disclose that even though these actions works well in preventing crime ,further more research have to be made to unravel other important policy-relevant issues to attentive police enforcement efforts.

Ahmed M and Salihu R S [33] studied the use of Geographic Information Systems (GIS) and spatial database of crime which supports in the determination of hotspots in Dala LGA of Kano State and also it classifies the challenges faced by the police people in implementing computerized crime mapping system. The author concluded that the crime doesn't occur nearer to police stations, but seldom occur all over the place outpost to the police stations and the author had mentioned the places where the police stations must be established.

Timothy C. Hart and Paul A. Zandbergen[34] determined the implications of data quality on predictive hotspot mapping for the data obtained from 6 large law enforcement jurisdictions in the United States for about 400,000 crime incidents. The authors developed the following set of recommendations for use by researchers and analysts engaging in hotspot analysis, the overall quality of the information must be assessed in prior to geocoding, crime incidents and geocode like crime events must be disaggregate separately, adapt geocoding techniques to fit for specific needs, positional accuracy must be characterized.

Jorge Ferreira, Paulo João and José Martins[35] pools cluster analysis and spatial models created along with GIS, based on police crime reports. The author developed a structure for short-term tactical positioning of police resources in which the objective is that the high level crime areas were identified to enable accurate

predictive models and to produce rigorous thematic maps. This work provides the essential tools for Decision Support System (DSS) of police departments in order to predict the hotspots.

Arbind Kumar Singh and Manimannan G [36] introduced a novel method of mapping the top level crime in various district/cities of Tamilnadu on the basis of crime factors. Narrative information and crime records are warehoused digitally across individual police departments, allowing the collection of this data to compile a district wise database of crimes they committed in Tamilnadu. The author applied k-mean clustering and Geographical Information System (GIS) analyses as data mining tools to develop the hidden structure present in the data for each year. Finally, the groups were identified as crimes belonging to High Crime Activity (HCA), Intermediate Crime Activity (ICA) and Low Crime Activity (LCA) in that order, which show the behaviour of High Crime Activity cities, Intermediate Crime Activity cities and Low Crime Activity cities. The results of this paper by applying data mining and GIS could potentially be used to identify the hot spot and even prevent the crime happening in future.

Xiang Zhang *et al.*, [37] proposed improved attribute oriented induce method and algorithm for crime hotspot detecting. Since the attributes for crime data is enormous the accurate prediction is not possible with the attribute oriented induce method as shown by the experimental study so the improved method was proposed here. Also a simple mapping method was employed for mapping the crime spots to show the clear map of hotspot areas.

Conclusion

This paper presents a broad review on various research papers relating to the applications of data mining techniques in crime hot spot analysis and crime forecasting. It explains about various spatial data mining techniques and tools that can be used for effective crime analysis. We believe that this paper will help researchers to obtain the overview of data mining application in crime analysis. This work is best suitable for real time applications intended for social welfare.

References

- [1] Nikhil Dubey, Setu kumar chaturvedi., 2014 “A survey paper on crime prediction technique using data mining”, International journal of Engineering and applications, Vol.4, Issue 3 (Version 1), PP.396-400.
- [2] Widmann, N., Baumann, P., 1997, “Towards Comprehensive Database Support for Geoscientific Raster Data,” Proceedings of ACM-GIS'97, Las Vegas/USA.
- [3] Van Eck, N.J., Waltman, L., “ Bibliometric mapping of the computational intelligence field” International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems 2007, 15:625–645.

- [4] Levine, N., 2010, "CrimeStat: A spatial statistics program for the analysis of crime incident locations" (v 3.3). Ned Levine & Associates, Houston, TX, and the National Institute of Justice. Washington, DC.
- [5] Anselin, L., Getis, A., 1995 "Local indicators of spatial association-lisa. Geographical Analysis," 27:93-155.
- [6] Murray, C., 2003 "OracleR_ Spatial Users Guide and Reference 10g Release 1 (10.1)" Redwood City, Oracle Corporation, 602.
- [7] Scott, L., Janikas, M., 2010 "Spatial statistics in ArcGIS. In Handbook of Applied Spatial Analysis," Fischer M., Getis A. eds., London, UK: Springer, 27-41. ISBN: 978-3642036460.
- [8] Anselin, L., 2003, "GeoDa 0.9 users guide," Urbana, 51:61801.
- [9] Keyvanpour, M.R., Javideh, M., and Ebrahimi, M.R., 2010, "Detecting and investigating crime by means of data mining: a general crime matching framework," Procedia Computer Science, World Conference on Information Technology, Elsevier B.V., Vol. 3, PP. 872-830.
- [10] Shyam Varan Nath, S., 2007, "Crime data mining, Advances and innovations in systems, K. Elleithy (ed.)," Computing Sciences and Software Engineering, PP. 405-409.
- [11] Manish Gupta, Chandra B and Gupta M. P., 2006, "Crime Data Mining for Indian Police Information System" in the 'Journal of Crime' Vol.2 Issue. 6, PP.43-54
- [12] Jonathan, J., Corcoran, Ian D. Wilson, J., 2003 "Andrew Ware Predicting the geo-temporal variations of crime and disorder," International Journal of Forecasting, Vol.19, PP.623-634.
- [13] Anna Buczak, L., Christopher Gifford, M., 2010. "Fuzzy Association Rule Mining for Community Crime Pattern Discovery," ISI-KDD, ACM ISBN 978-1-4503-0223-4/10/07 Washington, July 25.
- [14] Grubestic, Tony H., and Alan T. Murray. "Detecting hot spots using cluster analysis and GIS." Proceedings from the Fifth Annual International Crime Mapping Research Conference. Vol. 26. 2001.
- [15] Aniruddha Kshirsagar, Lalit Dole, 2014, "A Review On Data Mining Methods For Identity Crime Detection," International Journal of Electrical, Electronics and Computer Systems, Vol.2, Issue.1, PP. 51-55.
- [16] Hsinchun Chen, *et al.*, "Crime Data Mining: An Overview and Case Studies".
- [17] Divya Bansal, Lekha Bhambhu, 2013, "Usage of Apriori Algorithm of Data Mining as an Application to Grievous Crimes against Women", International Journal of Computer Trends and Technology, Vol.4, Issue.9, PP.3194-3199.
- [18] Devesh Bajpai, 2012, "Emerging Trends in Utilization of Data Mining in Criminal Investigation: An Overview", Journal of Environmental Science, Computer Science and Engineering & Technology, Vol.1, Issue.2, PP.124-131
- [19] Jitendra Kumar, Sripati Mishra and Neeraj Tiwari, 2012, "Identification of Hotspots and Safe Zones of Crime in Uttar Pradesh, India: Geo-spatial Analysis Approach", Int J Remote Sensing Applications, Vol.2, Issue.1, PP.15-19.

- [20] John David Elijah Sandig, Ruby Mae Somoba, Ma. Beth Concepcion and Bobby D. Gerardo, 2013, "Mining Online GIS for Crime Rate and Models based on Frequent Pattern Analysis", Proceedings of the World Congress on Engineering and Computer Science, Vol.2, PP.23-27
- [21] Hanmant ,N., Renushe, Prasanna ,R., Rasal and Abhijit ,S., Desai,2012, " Data Mining Practices for Effective Investigation of Crime", Int J Comp Tech & Applns,Vol.3, Issue.3, PP.865-870
- [22] TongWang , Cynthia Rudin, Daniel Wagner and Rich Sevieri, 2013, "Detecting Patterns of Crime with Series Finder", Twenty-Seventh AAAI Conference on Artificial Intelligence, PP.140-142.
- [23] Dawei Wang et. al. 2013, "Understanding the spatial distribution of crime based on its related variables using geospatial discriminative patterns", Comput Environ Urban Syst., 39, PP.93–106.
- [24] Kadhim ,B and Swadi Al-Janabi, 2011, "A Proposed Framework for Analyzing Crime Data Set Using Decision Tree and Simple K-Means Mining Algorithms", Journal of Kufa for Mathematics and Computer, Vol. 1, Issue.3, PP.8-24.
- [25] Donald Brown, E., 1998, "The Regional Crime Analysis Program (RECAP): A Framework for to Catch Criminals", PP.2848-2853.
- [26] P.Thongtae, S. Srisuk, 2008, "An Analysis of Data Mining Applications in Crime Domain", IEEE Conference, PP.122-126.
- [27] Jerry Ratcliffe , H.,Michael McCullagh,J., 1998,"Aoristic crime Analysis", Int J Geogr Inf Sci.,Vol.12,No.7,PP751-764.
- [28] Jerry Ratcliffe ,H.,"Aoristic Signatures and the Spatio-Temporal Analysis of High Volume Crime Patterns" J Quant Criminol ., Vol. 18, No. 1, March 2002
- [29] Chen, H.,Chung, W.,Xu, JJ., Wang, G., Qin, Y., and Chau, M,2004 "Crime Data Mining:A General Framework and Some Examples", v. 37 n. 4, PP. 50-56 IEEE.
- [30] Jerry Ratcliffe, H., 2004, "The Hotspot Matrix: A Framework for the Spatio-Temporal Targeting of Crime Reduction", Police Practice and Research,Vol. 5, No. 1, PP. 05–23.
- [31] Kate Bowers,J., Shane Johnson,D., and Ken Pease ,2004,"Prospective Hot-Spotting The Future of Crime Mapping?" Brit J Criminol, vol: 44, PP.641-658
- [32] Anthony Braga,A., 2005 ,"Hot spots policing and crime prevention: A systematic review of randomized controlled trials", J Exp Criminol.,springer 1:PP. 317–342.
- [33] M. Ahmed ,M., Salihu ,R.S., 2012, "Spatiotemporal Pattern of Crime Using Geographic Information System (GIS) Approach in Dala L.G.A of Kano State, Nigeria" ,Department of Geography Bayero University, Kano, Am J Eng Res., (AJER)e-ISSN: 2320-0847, ISSN : 2320-0936,Volume-2, Issue-3, PP.51-58.
- [34] Timothy, C., Hart Paul ,A., and Zandbergen,2012,"Effects of Data Quality on Predictive Hotspot Mapping" ,National Institute of Justice ,US Department of Justice,Office of Justice Programs,United States of America.

- [35] Jorge Ferreira, Paulo João and José Martins,2012, “GIS for Crime Analysis: Geography for Predictive Models”, The Electronic Journal Information Systems Evaluation Volume 15, Issue 1 , PP36 -49.
- [36] Arbind Kumar Singh., Manimannan, G., 2013,“Detecting Hot Spots on Crime Data Using Data Mining and Geographical Information System” ,Int J of Statistika and Matematika, ISSN: 2277- 2790 E-ISSN: 2249-8605, Volume 8, Issue 1, PP 05-09.
- [37] Xiang Zhang *et al.*, 2010, “Detecting and mapping crime hot spots based on improved attribute oriented induce clustering,” IEEE International conference on Geoinformatics, PP.1-5.