18

Survey of Data Mining Methods for Crime Analysis and Visualization

Okwangale Fredrick R., Department of Networks, Faculty of Computing and Information technology, Makerere University and Dr Ogao Patrick, Head of Department, Department of Information Systems, Faculty of Computing and Information technology, Makerere University

Crime prevention is a primary concern of police as they perform their central role of protecting the lives and property of citizens. But the police force is usually relatively very small compared to the crime prone population they have to protect making them more of a reactive rather than preventive force. Police often have at their disposal vast amounts of least utilised crime data (such as crime incident reports) which if analysed could reveal some hidden information such as crime committing trends useful in crime prevention. Use of Information Systems techniques such as data mining and Geographic Information Systems for analysing these data is promising in boosting the police efforts. This paper reviews the applicability of various data mining methods and Geographic Information Systems in crime analysis and visualization in mainly poor planned settings characterised by missing electronic data a common phenomena in the developing countries like Uganda. The focus is on criminality of places rather than the tracing of individual criminals. The review tends to reveal that a combination of Geographic Information Systems and data mining techniques that can work under unclean data are best suited for use in the poorly planned settings.

1. Introduction

Security of citizens is the major concern of the police. Rather than focusing on enforcement and incarceration police can deter crime through the knowledge benefits that derive from information and its associated technologies The police force can employ information technology to turn police officers into effective problem solvers and to leverage their intellectual capital to pre-empt crime (Brown et al., 2003)[4]. However one challenge to law enforcement and intelligence agencies is the difficulty in analyzing large volumes of data involved in criminal and terrorist activities (Chen et al., 2003)[14]. A variety of techniques in data mining and Geographic Information Systems (GIS) are surveyed in their applicability in use in environments of less organized data.

2. Overview of crime, data mining and Geographic information system

2.1 Crime

Crime is referred to as a comprehensive concept that is defined in both legal and non-legal sense (Akpinar et al, 2005)[2]. From the legal point of view crime is the breaking or breaching of the criminal law (penal code) that governs a particular geographical area (jurisdiction) aimed at protecting the lives, property and the rights of citizens of belonging to that jurisdiction. Crime is an offence against a person (for example murder, and sexual assault), or his/her property (for example, theft and property damage) or the State regulation (for example traffic violations) (Akpinar et al, 2005)[2] (Oxford Dictionary of Current English). In non-legal terms crime is a set of acts that violate socially accepted rules of human ethical or moral behavior (Akpinar et al, 2005)[2]; for example acting against a ritual in some society.

= Crime occurs in a variety of forms which police informally categorizes as being either major or volume. Major crime consists of the high profile crimes such as murder, armed robbery and non-date rape. These crimes can either be one-offs or serial. In the case of serial crimes it is relatively easy to link crimes together due to clear similarities in terms of modus operandi or descriptions of offenders. This linking is possible due to the comparatively low volume of such crimes. Major crimes usually have a team of detectives allocated to conduct the investigation. In contrast volume crimes such as burglary and shoplifting are far more prevalent. They are usually serial in nature as offenders go on to commit many such crimes. Property crimes, such as domestic burglary offences, committed by different individuals are highly similar and it is rare to have a description of the offenders (Adderley and Musgrove, 2001) [1]. Table 1 shows the classification of crime (Chen et al., 2003)[14].

Table 1. Crime types at different levels. Source: (Chen et al., 2003)[14]

Crime Type	Description
Traffic Violations	Driving under the influence of alcohol, fatal/personal injury/ property damage traffic accident, road rage
Sex crime	Sexual offences
Fraud	Forgery and counterfeiting, frauds, embezzlement, identity deception
Arson	Arson on buildings
Gang / drug offences	Narcotic drug offences (sales or possession)
Violent crime	Criminal Homicide, armed robbery, aggravated assault, other assaults
Cyber crime	Internet frauds, illegal trading, network intrusion /hacking, virus spreading, hate crimes, cyber piracy, cyber pornography, cyber-terrorism, theft of confidential information.

Also a single detective officer may have a large number of different volume crimes to investigate at any point in time. With a view to satisfying this demand, police forces around the world employ specialist crime analysts, people who have specialist training in a variety of disciplines including investigation techniques, criminal psychology and information technology. It is their task to assist investigating officers by analyzing crime trends and patterns, identifying links between crimes and producing packages which target an individual or group of offenders linking them to a series of crimes (Adderley and Musgrove, 2001)[1].

Most, if not all, of current systems both manual and computerized revolve around the investigation of crimes already committed. They are, therefore, reactive. In the developed countries like the UK, a majority of crime prevention forces use different types of relational database management systems (RDBMS) for recording and subsequent analysis of crime. Standard or interactive queries are written to produce patterns of crime, offending and various statistics (Adderley and Musgrove, 2001)[1] but its is a common phenomena in the developing countries to find mainly manual criminal record books used alongside the pin-up maps for crime incidence location.

2.1 Data Mining

Data mining deals with the discovery of unexpected patterns and new rules that are "hidden" in large databases It serves as an automated tool that uses multiple advanced computational techniques, including artificial intelligence (the use of computers to perform logical functions), to fully explore and characterize large data sets involving one or more data sources, identifying significant, recognizable patterns, trends, and relationships not easily detected through traditional analytical techniques alone. This information then may help with various purposes, such as the prediction of future events or behaviors. (Reza et al, 2001)[14]

The development of new intelligent tools for automated data mining and knowledge discovery has led to the design and construction of successful systems that show early promise in their ability to scale up to the handling of voluminous data sets.

Theories of crime and delinquency tend to be discipline-specific and are dominated by psychological, sociological, and economic approaches (Reza et al, 2001)[14]

Data Visualization

Visual methods are powerful tools in data exploration because they utilize the power of the human eye/brain to detect structures. A number of data mining tools for visualization exist, a histogram and Kernel plots being the most basic used for displaying single variables. Scatter plots for the display two variables at a time and reveal correlation, if any, between them. And for more than two variables, scatter plot matrices are often used (David et al, 2001)[7]. GIS also provides a powerful visualization tool through display of maps that allow the exploration of spatial patterns in an interactive fashion (Pfeiffer, 1996) [12].

2.2 Geographic Information Systems

Over the past few years Geographic Information Systems (GIS) has become a standard tool for crime analysts in many police departments, regardless of their size (see for example McEwen and Taxman 1994; Rossmo, 1995). One of the inherent advantages of GIS is its ability to integrate information from a variety of sources into one user interface. In turn, this allows for spatial analyses that would either not have been possible or at a minimum far more difficult prior to the advent of GIS (Olligschlaeger, 1997)[11]

2.3 Crime Analysis

Crime being inherently spatial phenomena (Ratcliffe, 2004)[13], taking place within a given location and at a specific time, it is logical to analyse crime in terms of spatial crime analysis. The advent of geographic information systems (GIS) has given behavioural scientists the ability to map and track a wide range of social phenomena such as crime incidences. Much more information about what is happening "on the ground" is now available. This enhanced analytic capacity has presented police departments with the opportunity to better serve and protect the people and places in their care. Furthermore, as GIS assists one in seeing and understanding behaviour patterns, it also provides stakeholders with opportunities to join together in partnerships for the common good (Holzman et al, 2003))[8]. A data pattern is an expression in some language describing a subset of the data or a model applicable to that subset (Fayyad et al, 1996)[6] There are at least three types of police crime information (primary, secondary, and tertiary), intelligence (prospective, retrospective, and applied), and operational strategies (preventive, prospective, and reactive), each of which interacts in a complex fashion with technology (Manning, 1992) [9]. Crime data analysis aids police in transforming this information from one type to another.

Crime Spatial Data Analysis

Spatial data analysis utilizes statistical analysis methods that address specific issues relating to spatial data, including spatial dependence (autocorrelation) and spatial heterogeneity. These issues run counter to traditional statistical assumptions of heterogeneity and independence of sample data. If these issues are ignored, then analysis results might not be valid. In combining the powerful tools of GIS to integrate and manipulate spatial data with rigorous statistical methods, spatial data analysis shows great promise for criminology, criminal justice, and law enforcement research and practice. (NIJ, 2006)[10]

3. State Of Crime Data At The Uganda Police

Most of the crime data for the Uganda police is manually recorded in criminal record books (CRB) in form of statements made by arrested suspected criminals, reports by informers and information gathered by the police themselves. Pin-up maps are used to capture the locations of crime incidents (Uganda Police online, 2005)[15] The deployment areas tend to be specific.

4. The Data Mining Techniques for Crime Spatial Data Analysis and Visualisation

There are a number of data mining techniques for crime analysis and visualisation but the choice of which to use depends on the features of the problem, the data and the objectives (David et al, 2001)[7]. This paper is limited to the Exploratory Data Analysis (EDA) and specifically crime spatial data analysis. Typically, EDA techniques are interactive and visual, and there are many effective graphical display methods for relatively small low-dimensional data sets. The difficulty in visualisation of points increases with the number of variables involved (David et al, 2001)[7]. The methods used in crime spatial data analysis can be classified into those concerned with visualisation of data, those for exploratory data analysis and methods for the development of statistical models (Pfeiffer, 1996) [12]. The methods are surveyed categories of the data based on the categories of crime data to be analysed - point patterns.

Point Patterns

Spatial point patterns (SPP) are based on coordinates of events such as locations of crime incidences and may also include the time of occurrence. All or a sample of point pattern may be plotted on the map. The aim of SPP analysis is to detect whether the point pattern is distributed at random, clustered or regular. SPP is typically interpreted as analysis of clustering. A dot map is commonly used to represent SPP. The tool effectively used for analysis of clustering effects is the K function. This method assesses clustering of crime incidences in detection of hot spots (Kingham et al, 1995) where time and space relationship analysis is required, the methods used are Knox's method, Mantel's Method and K-nearest neighbour method. All the three methods require the production of distance matrices of the spatial as well as temporal relationship between crime incidences. Knox's method requires critical distance in time as well as space defining closeness has to be set but the determination of these critical distances requires subjective decision. Mantel approach does not however require use of critical distances but uses both time and space matrices. It is however insensitive to non-linear associations. The K-nearest neighbour is based on the approximate randomisation of the Mantel product statistic (Pfeiffer, 1996) [12].

5. Conclusion

Exploratory data analysis makes few assumptions about data and it is robust to extreme data values. It is possible to use simple analytical models with EDA The methods that are robust to missing data are useful in the data mining of crime data where data is not so precisely collected. The distances between crime locations are normally not easily available to the police in the areas that are not well planned. The poorly planned areas are best represented by dividing them into area clusters and the analysis is done based on the clusters. The methods that support clustering are therefore best suited for the crime analysis of the poorly planned settings. The manual pin maps are best replaced by the use of GIS.

References

1. Adderley R.William and Musgrove Peter, (2001), "Police crime recording and investigation systems: A user's view", An International Journal of Police Strategies and Management, Vol. 24 No. 1, pp. 100-114.

2. Akpinar E. and Usul N. (2004). "Geographic Information Systems Technologies in Crime Analysis and Crime Mapping".

3. Andy Turner, (2002), "State-of-the-Art Geographical Data Mining", available at http://www.geog.leeds.ac.uk/people/a.turner/publications

4. Brown Mary Maureen & Brudney Jeffrey L. (2003), "Learning Organizations in the Public Sector? A Study of Police Agencies Employing Information and Technology to Advance Knowledge". Public Administration Review 63 (1), 30-43.

5. Colleen MCcue, Emily S. Stone, M.S.W., And Teresa P. Gooch, M.S., 2003, "Data Mining and Value-Added Analysis" FBI Law Enforcement Bulletin

6. Fayyad U. M., Piatetsky-Shapiro G., and Smyth P. (1996), "Knowledge Discovery and Data Mining: Towards a unifying framework", Proc. 2nd Int. Conf. on Knowledge Discovery and Data Mining. .Portland, Orgeon, AAAI Press, Menlo Park, Carlifornia, pp. 82 - 88

7. Hand David, Mannila Heikki, Smyth Padhraic., (2001), "Principles of Data Mining" Prentice Hall.

8. Holzman, H.R., W.D. Wheaton and D.P. Chrest. (2003) "Partnering with the Police to Prevent Crime Using Geographic Information Systems".

9. Manning Peter K., (1992), "Information Technologies and the Police" Crime and Justice, Vol. 15, Modern Policing, pp. 349-398

10. National Institute of Justice "Spatial Data Analysis", (January 2006), USA http://www.ojp.usdoj.gov/nij/maps/research.html- Accessed February 2006

11. Olligschlaeger Andreas M (1997), "Weighted Spatial Adaptive Filtering and Chaotic Cellular Forecasting with Applications to Street Level Drug Markets" PhD dissertation on Spatial Analysis of Crime Using GIS-Based Data:, Carnegie Mellon University

12. Pfeiffer, D. U. (1996), "Issues related to handling of spatial data". Massey University, Palmerston North, New Zealand.

13. Ratcliffe Jerry H. (2004), "Geocoding crime and first estimate of the minimum acceptable hit rate". International Journal of Geographical Information Science, Vol. 18, No.. 1, pp 61–72

14. Reza Fadaei-Tehrani, Thomas M. Green, (2002) "Crime and society" International Journal of Social Economics Volume 29 Number 10 pp. 781-795

15 Uganda Police online available at http://www.ugandapolice.go.ug/past_publications.htm