

How Data Mining Can Help Curb City Crime

Shakir Khan*

Abstract : Data mining has been used in different fields to help improve on the efficiency and effectiveness in extracting patterns and predicting certain events using existing statistical data. This paper sets out to apply data mining techniques and approaches in curbing city crime. By obtaining access to crime records held in various city police departments and classifying the data into fields such as address, location, type of crime, and time, it become possible to extract some useful indicators that would help manage crime. Neural networks and other learning algorithms were employed to analyze crime narratives to extract important information that would help the police to identify criminals, crime scenes, times of crime prevalence, and criminal activities in a city.

Keywords : Data mining, crime, crime records, Geographic Information Systems (GIS), Curb City Crime, Database Management Systems (DBMS)

1. INTRODUCTION

Conventionally, it has been the entitlement of law enforcement agencies to handle crimes. However, with the proliferation of automated or computerized information systems such as Geographic Information Systems (GIS) and Database Management Systems (DBMS) in cities, governments and private security bodies can now easily trail, unravel and therefore general manage crimes [1][9]. This is because specialized data analysis can aid the law enforcement agencies in this area to speed up the process of identifying, arresting, and prosecuting criminals. To achieve this, though, it requires a collaborative effort between data mining and criminal justice paradigms where the process of clustering models are used to identify crime patterns and if possible, crime hot-spots.

Clustering ensures grouping of crimes or hot spots in a particular geographical location such as a city. Through data mining, the clusters help in the identification of the patterns of crime in the city to help deal with common criminals including serial rapists, deadly gangs, and serial killers. Unfortunately, data mining experts have to balance between public and private data that can be extracted and made available for public consumption. This is because of privacy and confidentiality of certain information about crimes, criminals, or victims of crime [2].

In some developed nations such as Singapore, Britain and US, police officers use digital systems to handle crime reporting. Often, the reports have crime-related information grouped into location, crime type, date and time, and crime description among others. While obtaining data based on these categories is possible even through database queries, it is often difficult to extract useful data from crime descriptions because it is stored in form of text. This calls for the application of specialized data mining techniques and models such as Decision Trees and Neural Networks [3][4]. This paper engages in creating a framework for managing crimes in cities using data mining techniques.

* Assistant Professor, Information Management Department, College of Computer and Information Sciences Al-Imam Muhammad Ibn Saud Islamic University Riyadh, Saudi Arabia shakirkhan2006@gmail.com



Fig. 1. Categories of crime data.

2. RELATED WORKS

Since the 11/9 attack, state security has been enhanced in almost every country in the world. While it is useful to collect much needed information about criminals such as terrorists, an overload of such information proves detrimental in fighting crime. Several researches have been conducted in the area of crime management. However, the concept of using data mining in crime management is a relative new one. Data mining is the process of identifying interesting patterns or relations in data to help create predictive or statistical models [5]. Nonetheless, this section of the paper focuses on useful information relating to this topic, as published in related works.

When the security agencies (government or private) synthesize collected crime data, they stand a better position to obtain certain useful crime patters in cities and thus be able to prevent recurrence of the crimes or help in general crime management afterwards [6]. The benefits of using data mining in managing city crimes include better rates of crime resolution, fair and optimal allocation of crime management resources, updated information, lower officer training costs and time, and improve crime detection and prevention mechanisms [10]. This requires availability of robust crime detection and management tools that would benefit from the data mining techniques and tools that aid security agencies in managing crime.

- In most of the studies on crime management in urban areas, the following aspects of crime management are important:
- Identification of crime patterns in different geographical locations (cities, rural areas, etc.)
- Exploration of the characteristics of criminals in each geographical location
- Examination of the nature of crimes committed in different geographical locations
- Extraction of characteristics of criminals that commit crimes in society according to certain categories (serious, medium, and low crimes)
- To determine which criminals engage in different categories of crimes.

The following table highlights the various categories and types (sub-categories) of crimes committed in cities. These crimes are also committed in rural areas but the scale is low.

Putting wrong doing information into a GIS and make a guide with a reasonable visualization and reason for existing is the beginning stage of numerous wrongdoing examinations. To concentrate data from information and to change that data into learning is the essential destination of a choice making methodology. Anyhow to draw a decent guide is dependably a testing assignment. The imperativeness of information quality is vital to get powerful results however the exactness of spatially referenced information is additionally pivotal. [11]

Table 1. Crime categories and types.

<i>Type</i>	<i>Description of crimes committed in a city</i>
Traffic offenses	Personal injuries, damage of property, road rage, and driving under drug/alcohol influence
Theft	Burglary, carjacking, property theft, robbery, larceny, motor vehicle theft
Fraud	Embezzlement of funds, forgery, identity theft, and counterfeiting
Arson	Torching of apartments, offices, or buildings
Drug/Gang crimes	Drug trafficking: Possession of or sale of narcotics
Violent Crimes	Armed robberies, homicides, and motivated assaults among others
Cybercrime	Network hacking/intrusion, hate crimes, online pornography, theft of classified information, internet frauds, internet piracy, and illegal trades

At the point when confronted with new authoritative limits, a region is faced with another spatial modification. What's more the social and conservative flows are quick and surprising. With regards to adjust a mix of variables like policemen in the city, migration of regions and a ton of different variables, choices are troublesome.

The last consequences of this examination can be indispensable to choose a powerful police technique. The choice to continue with this new regulatory situation is so troublesome as to choose about better approaches to manage this new regional representation. To choose what bases to consider and what advances to receive are likewise troublesome decisions.

Lately, scientists and professionals have made gigantic advances in tackling the expository abilities of GIS to track wrongdoing examples over the long run and afterward utilize this data to make prescient models. These advances turned GIS in a significant device to aid and help choice making methods for the police powers and security administrations.

Since numerous law authorization organizations have embraced wrongdoing mapping and have started to create the sorts of instruments said, they need more.

The requests for more complex spatial scientific systems lead to the exploration on prescient models to help the anticipation of the "following wrongdoing".

The wrongdoing must be inspected in the setting of dangers and considering evaluation of the demographic, monetary, social and ecological variables that could influence the criminal demonstration. Simply seeing the genuine dangers and creating arrangements utilizing Geography, GIS and multivariate insights examination, we can dissect the current guide of censurability.

The progressed spatial examination came to distinguish wrongdoing examples and defenseless ranges (as far as frailty). This sort of exploration encourages learning keeping in mind the end goal to take key choices to battle the criminal phenomena. Nonetheless, just at micro scale, the criminal information can be considered as helpful to a vital arranging against wrongdoing.

Supplementing spatial examination of wrongdoing with the exact information of the recorded and social parts of a given region, one may recognize the purpose behind the event of a specific kind of wrongdoing in a given land zone and accordingly arranging and assets to avert and lessening it.

[12] Discovers and recognizes a few restrictions in Crime Document Clustering. For one thing, it addresses the shortcoming recognition and distinguishing proof in the k-implies calculation, then, it looks at the shortcoming of extraction terms from archives as it is expressed previously. In this way, this study means to improve the dependability of Document Clustering of wrongdoing report by proficient k-implies and additionally the extraction peculiarities of wrongdoing archive. Moreover, it is utilized for wrongdoing report bunching, and its results are the best confirmation for its productivity as it means to improve the K-means calculation for Document Clustering and additionally the concentrating of data which assemble themes/occasions of criminal acts can beat the first Document Clustering and other Document Clustering focused around two criteria of time and execution. Plus, we look forward that our recommendation of wrongdoing record bunching improves the execution and the viability.

In the late decade, a lot of logical inquiries about and studies have been performed on wrongdoing information mining. The results are typically developed in the part of new programming applications for recognizing and examining wrongdoing information. In [13] the creators present a general diagram on applying clever wrongdoing examination techniques including neural systems, Bayesian systems, and hereditary calculations in foreseeing and matching wrongdoing episodes. In [14], neural systems have been requisitioned wrongdoing information grouping and wrongdoing information characterization through utilizing both managed furthermore unsupervised learning strategies. The COPLINK national venture [15-16] which was initially created by the College of Arizona Artificial Intelligence Lab with subsidizing from the National Institute of Justice speaks to a conspicuous system for content mining, order and bunching of wrongdoing information intending to achieve moderately complex wrongdoing examination. The venture contains two basic segments: 1) COPLINK CONNECT and 2) COPLINK DETECT. The previous handles information preprocessing and information social affair loads and the later manage concentrating examples out of substantial volumes of wrongdoing information by utilizing information mining and counterfeit consciousness.

3. APPROACHES FOR MINING CITY CRIME DATA

This research entails use of the following techniques in mining crime data so that above five aspects of crime management.

- (a) **Entity extraction :** This approach will be used in producing automated identification of people by providing information such as one's address, properties owned, and vehicle details as captured in police report descriptions. This study suggests the use of neural network to identify valuable entities from narrative reports held by police departments. By a combination of lexical lookup and finite state machine to find matches or mismatches of neighboring with extracted words, noun phrasing, and a neural network, it becomes easy to identify useful entities in crime records held by police.

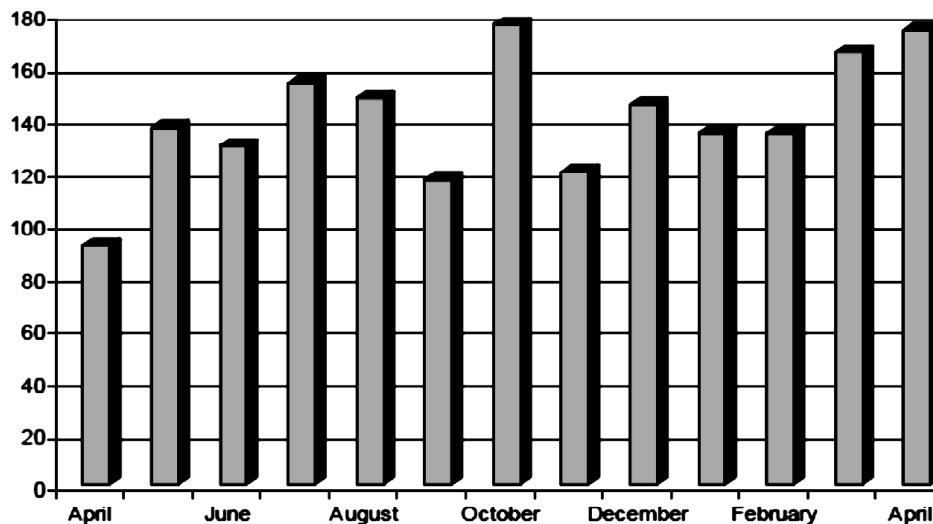


Fig. 2. Number of carjacking incidents across several months of a year.

- (b) **Clustering :** This technique helps to link dissimilar objects like people and vehicles or organizations in crime-related records).
- (c) **Deviation recognition :** This approach helps in detecting crimes such as network intrusion, fraud, and the analysis that requires tracing of strange happenings.
- (d) **String comparator :** This technique helps identify misleading information recorded in various criminal records. This is achieved by using an algorithm for detecting deceptions in the name, social security number, date of birth, and address of an individual.
- (e) **Social network analysis :** The technique is useful in managing crimes where criminals have social sites (*e.g.* Facebook and Twitter) for their communication and where they post some information relating to their activities. Using the proven Social Network Analysis techniques, this research was able to identify

groups of criminals using social networks to advance criminal activities. The process involves four components: network extraction, subgroup detection, interaction pattern recognition, and central member identification.

- (f) **Classification:** This is useful in identifying criminals engaged in crimes such as email spamming. To manage cybercrime, the research proposes a framework that uses an inductive learning algorithm to extract, trace, and identify cybercriminals based on the content they post on the Internet.

4. DISCUSSION

Investigating crime is a stimulating area of application for data mining as it plays a vital role in police systems of any city. Criminals can engage in dangerous criminal activities in cities, using sophisticated gadgets and means to attack their victims. Police departments on the hand can rely on the robustness of data mining tools and database management systems to curb city crime by providing a useful analysis of crime patterns, criminals, and crime sites for preparedness, alerts, prevention of crime, and general management of city crimes [7]. Data mining tools such as Oracle Data Miner can be applied against oracle databases, for instance, to extract crime related data that can be used to establish clusters visually showing crime hotspots and concentration of crime in cities [8].

From the findings above, it is evident that crime management using data mining techniques to support existing tools will go a notch higher to help curb crime in cities. For instance, basing on the number of crimes reported according to categories, one could predict when the highest number of incidents is expected so that proper security measures are put in place to counter the criminals. Fig 2 above, for instance, shows that the number of carjacking over several months. If from a data set of three consecutive years the trend appears like the one shown, then it is proper to predict that October and April are the worst months of the year in terms of carjacking incidents in a given city. A combination of specialized computer algorithms proves very useful in processing text and numeric data obtained from crime records held in police departments.

5. CONCLUSION AND FURTHER WORKS

This paper has presented a synopsis of how to use data mining techniques in managing crime in cities. From the related works and approaches used, we have confidence in the use of data mining to improve efficiency and effectiveness in the management of crimes in cities because the techniques promise support to the security experts and intelligence analysts in combating crimes in cities. However, this study area being relative young, this research would recommend several future directions for further study. In this paper, we propose two of them. First, that a study on network visualization using more visual intelligence investigation and intuitive criminal. Second, a combination of CCTV technologies should be combined with data mining to help achieve instant crime detection and management. Finally, using GPS and GIS technologies [9], data extracted from data.

6. REFERENCES

1. B. Devesh. "Emerging Trends in Utilization of Data Mining in Criminal Investigation: An Overview". *Journal of Environmental Science, Computer Science and Engineering & Technology*, 2012.
2. A. Malathi, S. Santhosh B. and A. Anbarasi. "An Intelligent Analysis of a City Crime Data Using Data Mining". *Internation Conference on Information and Electronics Engineering*, 2011.
3. K. Thearling. *An introduction to data mining*. 2003. Accessed on December 28, 2014 <http://www3.shore.net/~kht/text/dmwhite.htm>.
4. Hsinchun et al. "Crime Data Mining: A General Framework and Some Examples". *IEEE Computer Society*, 2004
5. U.M. Fayyad and R. Uthurusamy. "Evolving data mining into solutions for insights". *Communications of the ACM*, vol. 45, Iss. 8, 2002, pp.28-31.
6. Megaputer Intelligence, Inc. *Crime Pattern Analysis: Megaputer Case Study*. 2002. Accessed on December 28, 2014 at http://www.elon.edu/facstaff/mconklin/cis230/cases/crime_pattern_case.pdf

7. R. G. Uthra. "Data Mining Techniques to Analyze Crime Data". International Journal for Technological Research in Engineering, 1 (9), 2014
8. Whitepaper, "Oracle's Integration Hub For Justice and Public Safety", Oracle Corp. 2004, available at: http://www.oracle.com/industries/government/IntegrationHub_Justice.pdf
9. F. Jorge, Joao, P. and Martins, J. "GIS for Crime Analysis - Geography for Predictive Models" *The Electronic Journal Information Systems Evaluation*, 15 (1), pp.36-19, 2012
10. Asegehn, L. W. "The Application of Data Mining In Crime Prevention: The Case of Oromia Police Commission". Faculty of Informatics: Addis Ababa University, 2013
11. Ferreira, J, João, P and Martins, J. "GIS for Crime Analysis - Geography for Predictive Models" *The Electronic Journal Information Systems Evaluation* Volume 15 Issue 1 2012, (pp36 -49), available online at www.ejise.com
12. Qusay Bsoul, Juhana Salim*, Lailatul Qadri Zakaria "An Intelligent Document Clustering Approach to Detect Crime Patterns" Knowledge Technology Research Group, Universiti Kebangsaan Malaysia, 43600 Bangi Selangor, Malaysia The 4th International Conference on Electrical Engineering and Informatics (ICEEI 2013)
13. G.C. Oatley, J. Zeleznikow, B.W. Ewart, "Matching and Predicting Crimes," In Applications and Innovations in Intelligent Systems XII in Proceedings of AI2004, The Twenty-fourth SGA International Conference on Knowledge Based Systems and Applications of Artificial Intelligence. Ann Macintosh, Richard Ellis and Tony Allen Ed. London: Springer, , pp. 19-32, 2004.
14. R. William Adderley, "The use of data mining techniques in crime trend analysis and offender profiling," Ph.D. thesis, University of Wolverhampton, Wolverhampton, England , 2007.
15. Y. Xiang, M. Chau, H. Atabakhsh, H. Chen, "Visualizing criminal relationships: comparison of a hyperbolic tree and a hierarchical list," *Decision support systems*, Elsevier Science Publishers, vol. 41 no.1, pp: 69-83, Nov. 2005.
16. R.V. Hauck, H. Atabakhsh, P. Ongvasith, H. Gupta, H. Chen, "Using Coplink to Analyze Criminal-Justice Data," *Computer*, vol. 35, no. 3, pp. 30-37, Mar. 2002.