

Crime Profiling and Pattern Detection Using MLA Tools: A Review

Aslina Mat Asli, Nurazean Maarop, Nilam Nur Amir Sjarif

Faculty of Technology and Informatics Razak
Universiti Teknologi Malaysia
Jalan Sultan Yahaya Petra 54100, Kuala
Lumpur .
aslina3@live.utm.my, nurazean.kl@utm.my,
nilamnur@utm.my

Article history

Received:
2 July 2019 Received in

Received in revised form:
25 August 2019

Accepted:
30 September 2019

Published online:
3 October 2019

Aslina Mat Asli,
aslina3@live.utm.my

Abstract

The application of Machine Learning Algorithm (MLA) in performing statistics and modeling to reveal pattern of current and future behavior is significantly useful as to cope up with the recent evolution of Big Data environment. In crime investigation, data mining is an important tool used to uncover hidden information from a large amount of data. There are several approaches and techniques of data mining discussed in this paper in regard to managing crime dataset and analytic in criminology world. This study aims to describe the recent method and the most commonly model used in performing the crime-matching process by highlighting a number of important directions for future research to assist in the process of identifying patterns of crime. This paper explores the application of MLA to find the crime pattern by comparing all types of crime with different profiling such as date, time, and location of the crime and these are commonly known as an attributes, variables or parameters. As a result, this study observed that crime location and time are the most important profiling variables in criminological based analysis. This study also suggested that the combination of both clustering and classification techniques is best applied in such context.

Keywords: *Crime, Crime Pattern, Criminal Profile, Machine Learning, Data Mining*

1. Introduction

Crime is defined as a behavior disorder that is an integrated result of social nuisance, economic and environmental factors that caused our society in several ways [1]. Whereas, criminal profiling refers to any means of inferring the traits of individuals responsible for committing criminal acts [2]. This profile can give police departments a basic idea on who they are looking for based on some understanding and criminal behavior [3]. Profiling also allows investigators to precisely predict the characteristics of current and future offenders, allowing them to be caught before they can continue committing to other crimes based on the patterns and motives of

previous crime [4]. Besides, criminal profiling also known as a biographical sketch usually used to discover behavioral patterns, trends and tendencies in committing crimes [5]. The system usually processes the inputs over four integrated components: geographic profiling, social network analysis, crime profile, and physical matching to detect suspects [6].

* *Corresponding author. aslina3@live.utm.my*

De Bruin et al. [7] introduced a crime trends framework by comparing individuals based on their profiles using a new distance measure before clustering them accordingly. This method also provides a visual clustering of criminal career identification of classes of criminals. According to Keyvanpour et al. [8], criminal analysis entails systematic approach to identify commonly used techniques to predict criminal incidents. Significance of crime-related data, hidden existence and the possibility connection between them has made data mining techniques become popular among criminologists, criminal investigators and criminal analysts [8]. However, Tong Wang et al. [9] stated that to find a specific crime patterns can be much difficult in particular involving the review of current crime reports and the comparison to past crimes. Somehow, the combination between human crime analyst's capabilities and the technology of machine learning tools may be able to complement the work [9]. In addition, an analyst usually processes the information instinctively while the algorithm searches through the database looking for the similarity to discovery some important of crime patterns [10].

Preetha et al. in year 2017 [11] indicates that data mining is the computational process of discovering patterns in large datasets involving artificial intelligence, machine learning, statistics, and database systems and has been used nowadays in worldwide research. Knowledge derived from data mining approaches can help and support law enforcement teams to resolve crime more quickly [8]. This paper explores the data mining approach in manipulating crime information to find a possible crime pattern based on different profiling and attributes that can be used to predict the trends of crimes.

2. Existing Works

The traditional system in managing criminal records has failed to live up the requirements of existing crime cases [12]. This traditional system is also time consuming as Crime Department officers have to manually analyze these big databases of hundreds and thousands of entries before making a firm conclusion [12]. Several approaches based on current technology have been adopted in detecting and predicting crime pattern to overcome the limitation of manual inspection. One of the most common tools used in criminology analysis is by using Machine Learning Algorithm (MLA). Bindiya in year 2010 [13] indicates that data mining is a potential technique to help criminal investigators focus on the most important information in crime data and fasten the process of discovering crime's patterns for solving crime.

Therefore, data mining plays an important role in analyzing and predicting crime [14]. Most common data mining techniques used in the present research work are clustering, classification and deviation detection [1]. Most of researcher in

criminology field use data mining as the process of finding patterns in a set of crime dataset to solve problems through data analysis. Additionally, data mining can also be used to convert raw data into a meaningful information to investigate the causes of crime, the social impact of crime, the actors involved, and to identify suspected persons before they commit crimes.

Keyvanpour et al. [8] indicates that data mining techniques are needed as an approach to manage the large volumes of crime-related data in police departments for crime investigation. The author [8] have stated that crime-matching process is depends on the relationship between crime or criminals with the previous resolved or unresolved crime cases. Therefore, artificial neural network (SOM and MLP) and binary encoding was chosen to overcome the limitation on K-Means algorithm for higher dimensions of crime-related to improve the accuracy and reliability of criminal detection process [8]. Furthermore, his [8] study have defined that crime characteristic and crime-matching are two main elements that are important in crime analysis as crime variables for the algorithm to works on them.

Three different groups of crime variables based on their uniqueness have been categorized to help crime analyst in analyzing the crime [8]:

- Crime location coordinates or the time of occurrence – spatio-temporal crime variables;
- Crime scene characteristics, offender's behavioral pattern – crime natural specifications;
- Age, sex, and race – offenders' profile.

Finally, to make the crime matching engine work, [8] have applied a backpropagation training method for classification process using MLP neural network with the ability to tolerate noisy data instances to be used in parallel data processing.

In addition, Rishabh et al. [12] proposed a swift response system to discover crime pattern among criminal network to identify the probability of local suspects involved in a crime based on relevant case histories analysis. Thus, prediction techniques were used to filter and identify the different profile of people present at the crime scene [12]. According to the author [12], K-means clustering algorithm helps to find cases that are related to each other to be gather in one cluster. As a result, only particular cases will be focused to find out possible suspect. Based on the scenario, the criminals make use of cell phones for communication, analyzing of these cell phone records involving both suspects and the victims to track clues in crimes to track the real perpetrator [12]. In order to resolve the case concurrently, [12] have mapped the suspect's current location by using mobile GIS to analyses crime database in certain areas where crime incidents occurred to identify the suspect as per illustrated in figure 1 below:

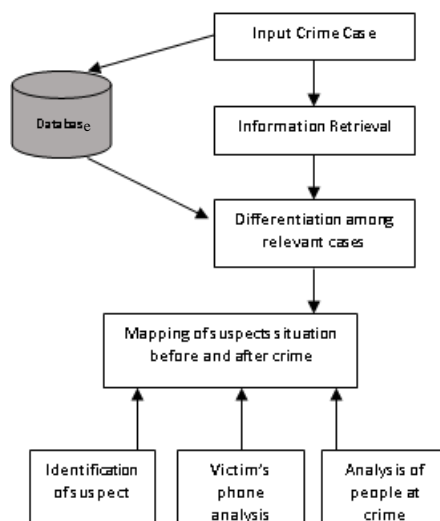


Figure 1. System Block Diagram [12]

With the help of K-means clustering algorithm, the suspect to the crime are visualized. Thus, Apriori algorithm also been used to find frequent pattern analysis to generate frequent item sets among callers and callees [12].

Zakaria et al. [15] in year 2014 stated that data mining techniques can help the analysts to speed up the process of discovering crimes for a better decision making. Based on the case study conducted in Libya Police Department, [15] proposed a model for a more strategically decision in preventing high crime rate in Libya. In this regard, clustering technique was applied to identify a crime spree within the same group of suspects or committed by individual to find the variables in providing the best clustering. Clustering technique was used to identify a crime spree in the same group of suspects or done by individuals in preparing the best cluster of variables [15]. Therefore, the automatic detection of crime patterns will help detectives in solving one crime based on spree in advance rather than whole “spree” [15]. For some cases, if the groups of incidents are suspected, a complete description of the different pieces of information were developed from each criminal incident [15]. The author [15] also used a combination of both data mining algorithms including association rules mining and clustering to classify crime records based on the values of attributes crime to illustrate the outcome with a better performance. Besides, the rules generated by association rule mining could be easily presented in human language that can be understood by the officers in determining a crime prevention strategy [15]. In his study, there were different attributes and profiles were used in searching for crime pattern as stated below [15]:

- Crime ID : individual crimes are designated by unique crime IDs;
- Crime Type : indicates crime type;
- Date : indicate when a crime happened;
- Gender : male or female;
- Age : age of individual criminal;
- Crime Address : location of the crime;
- Marital status : status of the criminal.

However, Kalaiselvan et al. [16] applied clustering technique to help identify identical group records among them with multiple spreadsheet reports from their own crime data logs. The author [16] implements K-means algorithm with a combination of clustering and geo-spatial algorithm to convert the data crime report into data mining report which contains some unresolved crime as illustrated in figure 2.

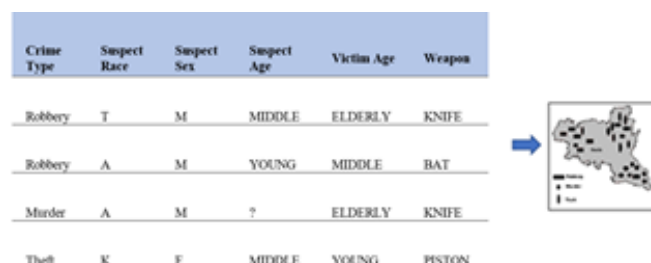


Figure 2. Report Mapping [16]

Based on figure 2 above, author [16] categorized the crime profile into six main attributes:

- Crime type
- Suspect race
- Suspect age
- Victim age
- Weapon used

Thus, mapping the real data to data mining attributes is not an easy task which requires a data miner and crime data analyst with good domain knowledge and working closely with the detectives from the beginning of crime investigation [16]. A research conducted by Almanie et al. [17] using an existing real-world crime dataset in two cities of the US aims to find spatial and temporal criminal hotspot has implements Apriori algorithm by having formula of three item set (Location, Day and Time). Two different supervised learning methods has been applied to the models – Naïve Bayesian and Decision Tree classifier to predict potential crime types in a specific location within a time. However, the author [17] indicates that more classification models are needed in future to increase crime prediction accuracy and to enhance the overall performance. Thus, [17] also believes that by extracting all possible patterns from crime variable could help in predicting potential crime hotspots at a certain time in the future. Three main variables of crimes data as per stated below [17]:

- Crime type
- Crime occurrence time

- Crime location

Table 1. Summary of Existing Work on Machine Learning Algorithms for Crime Pattern Detection Based on Crime's Profiling

Author	Profiling (Crime	Method	Attributes)	Findings
[8]	<ul style="list-style-type: none"> • Crime location coordinates or the time of occurrence. • Crime scene characteristics and offender's behavioural pattern. • Age, sex, and race – offenders' profile. 	□	Artificial Neural Network (SOM and MLP) Binary Encoding	Accuracy and reliability of intelligent crime detection process improvement.
[12]	<ul style="list-style-type: none"> □ Phone Records • Suspect Records by mapped the • Current Location 	□	K-means Algorithm current location of the	Generates frequent item Records Apriori callees <u>suspect using mobile GIS.</u>
[15]	<ul style="list-style-type: none"> □ Crime ID's • Crime Type • Date • Age • Crime Address • Marital Status 	□	Clustering Technique Association □	Crime spree identification. Gender Rules Mining
[16]	<ul style="list-style-type: none"> □ Crime Type • Suspect Race • Suspect Age • Victim Age • Weapon Used 	□	K-means Algorithm Geo-Spatial Algorithm	Identifies identical group of records from multiple spreadsheets document which contain unresolved previous crime.
[17]	<ul style="list-style-type: none"> □ Crime Type • Crime Occurrence Time • Crime Location 	□	Apriori Algorithm Naïve Bayesian Decision Tree □	Spatial and temporal criminal hotspot based on specific location.

Based on the approaches discussed above, most of researchers applied data mining technique as commonly used for crime pattern detection in criminology analysis. However, the enhancement of the approach must be fully undertaken in future by considering the combination of both clustering and classification techniques to

improve the quality result of crime related matching-process. In addition, crime with different profiling such as location and time of crime occurrence are one of the important criterions in order to discovering the crime pattern detection.

3. Conclusion and Future Work

This study has presented the important related works of profiling in providing the law enforcement to generate psychological profile based on criminal's personality to assist the investigation process. All attributes discussed above can be considered by other researchers to be aided in their study for future crimematching process. In addition, the presentation of information from this study can be used by those working in social network crime analytics in discovering the relationship between criminal and related crime's profiling. However, this study has limited information regarding other environment where future research may need to consider other attributes such as religion belief, contextual practices and culture aspect among criminals.

Acknowledgments

We would like to thank Universiti Teknologi Malaysia, Ministry of Education Malaysia.

References

- [1] A. Malathi and Dr. S. Santhosh Baboo, "Evolving Data Mining Algorithms on the Prevailing Crime Trend – An Intelligent Crime Prediction Model ", *International Journal of Scientific & Engineering Research* Vol. 2, Issue 6, 2011.
- [2] BE Turvey, "Criminal Profiling: An Introduction to Behavioral Evidence Analysis", *The Encyclopedia of Clinical Psychology*, Pp.1-6, 2014.
- [3] Gibson, Dirk C., "Clues from Killers: Serial Murder and Crime Scene Messages", Westport CT: Praeger, Pp.13, 2010.
- [4] Ibe, Patrick, Charles Ochie and Evaristus Obiyan. "Racial Misuse of "Criminal Profiling" by Law Enforcement: Intentions and Implications", *African Journal of Criminology & Justice Studies* 6, No. 1/2, 2012.
- [5] Vorpagel, R.E, "Painting Psychological Profiles: Charlatanism, Coincidence, Charisma or New Science", *Police Chief*, 3(8), Pp.156–159, 1982.
- [6] Li Ding, Dana Steil, Matthew Hudnall, Brandon Dixon, Randy Smith, David Brown, Allen Parrish, "PerpSearch: An Integrated Crime Detection System", *Proceedings of the 2009 IEEE International Conference on Intelligence and Security Informatics*, Pp.161-163, 2009.
- [7] J.S. de Bruin, T.K. Cocx, W.A. Kusters, J. Laros and J.N. Kok, "Data Mining Approaches to Criminal Career Analysis", in *Proceedings of the Sixth International Conference on Data Mining (ICDM'06)*, Pp.171-177, 2006.
- [8] Keyvanpour, M.R., Javideh, M. and Ebrahimi, M.R, "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, 2010.
- [9] Wang, Tong, Cynthia Rudin, Dan Wagner, and Rich Sevieri. "Learning to Detect Patterns of Crime, Discovery in Databases", *ECML PKDD 2013, Prague*, Pp.23-27, 2013.
- [10] Kumara Pathirana, Premaratne, "Survey on Data Mining Methods Which Can be Used in Crime Analysis and Visualization for Sri Lanka", 2017.
- [11] Dr. S. Preetha, S. Preethikha, S. Premiya Jasmine, "Criminal Investigation", Vol. 9, No. 1, 2017.
- [12] Rishabh Singh, Akash Agivale, Mahesh Mane, Bhavin Oza, Akshay Naik, "CDR and TD Analysis Using Data Mining", Vol. 3, Issue 6, 2017.
- [13] Bindiya M Varghese, Unni Rishnan A, Paulose Jacob, Justin Jacob, "Correlation Clustering Model for Crime Pattern Detection", *International Journal of Advancements in Computing Technology* 2:125128, Vol. 2, Issue 5.14, 2010.

- [14] Revatthy Krishnamurthy,J. Satheesh Kumar, "Survey of Data Mining Techniques on Crime Data Analysis", International Journal of Data Mining Techniques and Applications (IJDMTA), Vol. 1, Issue 2, Pp.47-49, 2012.
- [15] Dr Zakaria Suliman Zubi, Ayman Altaher Mahmud, "Crime Data Analysis Using Data Mining Techniques to Improve Crimes Prevention", International Journal of Computers, Vol. 8, 2014. [16] P. Dhakshinamoorthy, T.Kalaiselvan, "Crime Pattern Detection Using Data Mining", International Journal of Advanced Research in Computer Science and Applications, Vol.1, Issue 1, 2013.
- [17] Tahani Almanie, Rsha Mirza and Elizabeth Lor, "Crime Prediction Based on Crime Types and Using Spatial and Temporal Criminal Hotspots", International Journal of Data Mining & Knowledge Management Process (IJDKP), Vol.5, No.4, 2015.