On the Use of Data Mining Techniques for Crime Profiling

Fagbaibi, S.O
Computer Service Department
The West African Examinations Council
Ibadan Zonal Office
Ijokodo, Ibadan, Nigeria
sherifolumide@gmail.com

Ibrahim, Y. L
Department of Mathematical Sciences
Kogi State University
Ayungba
Kogi State, Nigeria
enemakwu@gmail.com

Longe, O.B (PhD)
Fulbright Fellow
International Centre for Information Technology & Development
Southern University System
Baton Rouge, Louisiana, USA.
longeolumide@fulbrightmail.org

On the Use of Data Mining Techniques for Crime Profiling

Fagbaibi, S.O, Yahya, Y. I & Longe, O.B.

ABSTRACT

Crime is today a salient fact, an integral part of the risks we face in everyday life. The concern about national and international security has increased significantly since the incident of September 11th, 2001 attacks. However, information overload thwarts the effective and efficient analysis of criminal activities. Application of data mining in the context of law enforcement and intelligence analysis holds the promise of solving such problems. The benefit of data mining for police seems tremendous, yet only a few limited applications are documented. Data mining can be used to model crime detection problems. Any research that can help in solving crimes faster will pay for itself. This paper gives reviews current trends in profiling crime using data mining techniques. We proposed the use of clustering algorithm as a data mining approach to help detect the crimes patterns and speed up the process of solving crime.

Key words: Crime, profiling, data mining, criminals, attacks and detection

1. INTRODUCTION

Human investigators with many years of experience can often analyze crime trends precisely, but as the incidence and complexity of crime increases, human errors occur, analysis time increases, and criminals have more time to destroy evidence and escape arrest. By increasing efficiency and reducing errors, crime data mining techniques can facilitate police work and enable investigators to allocate their time to other valuable tasks. Without a common understanding of crime patterns, security detectives and law enforcement agents will not be able to develop a common picture of crime in an area, incorporate other information that may help explain crime patterns that make the analysis easier, hence solving crime faster. Historically solving crimes has been the prerogative of the criminal justice and law enforcement specialists (Chang et al, 2003).

1.1 Crime

Crime is defined as “an act or the commission of an act that is forbidden, or the omission of a duty that is commanded by a public law and that makes the offender liable to punishment by that law” (Webster Dictionary, 2007). An act of crime encompasses a wide range of activities, ranging from simple violation of civic duties (e.g., illegal parking) to internationally organized crimes (e.g., the 9/11 attacks). The concern about national security has increased significantly since the terrorist attacks on September 11, 2001. Intelligence agencies such as the CIA and FBI are actively collecting and analyzing information to investigate criminal activities. Local law enforcement agencies have also become more alert to criminal activities in their own jurisdictions. One major challenge to law enforcement and intelligence agencies is the difficulty of analyzing effectively and efficiently large volumes of data involved in criminal activities. For example, complex conspiracies are often difficult to unravel because information on suspects can be geographically diffuse and span long periods of time.

Detecting cybercrime can likewise be difficult because busy network traffic and frequent online transactions generate large amounts of data, only a small portion of which relates to illegal activities. Crime cases can be classified broadly into the following types: Traffic Violations, Sex Crimes, Theft, Frauds, Arson, Gang / drug offenses, Violent Crime and Cyber Crime. Some types of crime, such as traffic violations and arson, primarily concern police at the local and state levels. Other crime types are investigated by local law-enforcement units as well as by national and international agencies. For example, a state police department’s sex crimes unit may track local pedophiles and prostitutes, while the FBI and the International Criminal Police Organization focus on transnational trafficking in children and women for sexual exploitation. Many crimes, such as the theft of nuclear weapons data, can have profound implications for both national and global security.

Transnational fraud and trafficking in stolen property or contraband can severely impact trade, business, and government revenue. Local gangs as well as foreign-based drug cartels and criminal organizations exact a large financial cost as well as threaten public health and safety. Although most types of violent crime—such as murder, robbery, forcible rape, and aggravated assault—are local police matters, terrorism is a global problem that relies on cooperation at all levels of government. The Internet’s pervasiveness likewise makes identity theft, network intrusion, cyber piracy, and other illicit computer-mediated activities a challenge for many law-enforcement bodies.

1.2 Data Mining

According to the Gartner Group, “Data mining is the process of discovering meaningful new correlations, patterns and trends by sifting through large amounts of data stored in repositories, using pattern recognition technologies as well as statistical and mathematical techniques.”(The Gartner Group, www.gartner.com.). It can also be defined in the following ways:

- Data mining is the analysis of (often large) observational data sets to find unsuspected relationships and to summarize the data in novel
ways that are both understandable and useful to the data owner” (Hand et al., 2001).

- Data mining is an interdisciplinary field bringing together techniques from machine learning, pattern recognition, statistics, databases, and visualization to address the issue of information extraction from large data bases” (Evangelos Simoudis in Cabena et al, 1998).
- Data mining is defined as the identification of interesting structure in data, where structure designates patterns, statistical or predictive models of the data, and relationships among parts of the data (Fayyad & Utharussamy, 2002).

Cluster (of crime) has a special meaning and refers to a group of crime in geographical settings, i.e. a lot of crimes in a given geographical region. Such clusters can be visually represented using a geo-spatial plot of the crime overlaid on the map of the police jurisdiction. The densely populated group of crime is used to visually locate the ‘hot-spots’ of crime. However, when we talk of clustering from a data-mining standpoint, we refer to similar kinds of crime in the given geography of interest. Such clusters are useful in identifying a crime spree. There is a need to develop a data mining paradigm that can help solve crimes faster, more specifically, using clustering based models to help detect the crimes patterns and speed up the process of solving crime.

2. CRIME AND CRIMINAL OPERATIONS

Crime is today a salient fact, an integral part of the risks we face in everyday life. In both scholarly and public opinion crime is associated with harm and violence; harm to individuals, destruction of property, and the denial of respect to people and institutions. On the surface, to ask “what is a crime?” seems to warrant a straightforward answer in that one can simply suggest that “crime is something that is against the law.” For those who adopt such a strict definition or a legal-consensus approach to crime, studying the law as it is written is sufficient for understanding what society considers harmful behavior.

To ask “what is a crime?” is certainly not a novel endeavour. For decades academics from numerous disciplines (such as law, sociology, and criminology) have struggled to understand various aspects of this question. From studies that examine the factors contributing to the enactment of certain prohibitions or the impact of law and its enforcement, to studies that focus on the events that precede the decriminalization of certain behavior, there are countless examples of scholarly work dedicated to exploring the nature of crime. In the last half of the twentieth century, various scholars noted that crime is not an objective phenomenon and that the way in which certain behavior is understood and responded to is more a reflection of how society is structured than an indication of any inherent problems with those individuals regarded as criminals. In the 1940s, for example, Edward Sutherland introduced the concept of “white-collar crime” to draw attention to crimes committed by the upper class and corporate elite, thereby challenging the common perception that crime was committed primarily by those in the lower class.

Other critical scholars built upon Sutherland’s work by continuing to explore definitions of crime and its enforcement. The 1960s produced a considerable body of literature within this tradition. Howard Becker’s (1963) pioneering work on moral entrepreneurship highlighted the process by which the perceptions and claims of certain groups (often criminal justice officials) were shaped by law and law enforcement, emphasizing that, in many instances, there was an absence of empirical evidence to substantiate the level of concern associated with a given problem. Using the emergence of anti-marijuana legislation in the 1930s as an example, Becker illustrated how criminal justice officials effectively manufactured a crisis over the nature and extent of drug use, when in actuality drug consumption was rare. According to Becker (1963), “from this point of view, deviance is not the quality of the act the person commits, but rather a consequence of the application by others of rules and sanctions to an ‘offender’. Stanley Cohen’s work on the sociology of deviance also brought to light the idea that definitions of crime were not objective, by revealing how crime can be produced through “moral panics.” In essence, Cohen revealed that if we deem a group and its behavior to be a threat, and if this belief is supported through media accounts and professional claims, then we are likely to respond as such even though the nature and extent of the concern may be more perceived than real (Cohen, 1980).

In more recent decades, a new generation of critical criminological, legal, and socio-legal scholars has discussed and debated the nature of crime and law in society. Various feminists, for example, have brought to light different forms of gender bias within notions of law and its enforcement. As Comack et al. (1999) suggest, over the last twenty years, feminism has made considerable inroads in challenging to reconsider the traditional approaches to understanding the law–society relation as well as the claims that law itself makes in its Official Version.” Critical race scholars have also drawn attention to racial biases within criminal justice processes, noting that conceptions of crime and its control are rooted primarily in the experiences of white men (Neugebauer, 2000).

Also in Canada, scholars (both Aboriginal and not) have pointed to the importance of understanding the colonization of Aboriginal peoples when examining their overrepresentation in the criminal justice system (Williams, 2001). In addition to reminding us that crime and law are not objective phenomena, these critical feminist, race, and Aboriginal literatures reveal the troubling differences between how the law is written and how it is enforced. The late 1970s and early 1980s witnessed a recognition by government agencies that a reflex application of criminal law had become commonplace and that this might not be the most appropriate approach for dealing with complex social issues.


2.1 Criminality
A theme that has been and continues to be a considerable area of concern for critical legal scholars is the differences between the law on the books and the law in action, or the differences between laws as they are written and laws as they are enforced. A law is not only something that is brought to bear through (subjective) social processes, but also, once on the books, something that continues to be subject to many of the same societal and institutional pressures that helped define the nature of the criminalized behavior and the associated response. In particular these pressures help determine the extent to which a certain law is enforced and in what manner. Previous studies have revealed the impact of the law in action in various ways. Some examples include:

- how the law is unequally enforced to the detriment of certain groups, as is the case with the overpolicing of Aboriginal peoples and other racial minorities in Canada (Manitoba 1991; Williams 2001).
- how domestic violence has been underpoliced (Bonnycastle & Rigakos 1998; Tutty & Godd 2002) as well as the unanticipated consequences of mandatory arrest policies, including more women arrested in domestic-violence cases (McMahon and Pence 2003).
- gender bias in the law’s unequal application to women – for example, only women being charged with soliciting although the legislation theoretically includes men (Layton 1979; Lowman 1986).

Reflecting upon the processes involved in claiming certain actions or events as crimes serves as a route into analyzing differences and similarities in actual events and their role within the sets of beliefs, understandings, and reactions to others that enable our societies to cohere.

Table 2: Summarizes the different types of crimes sorted by the degree of their respective public influence.

<table>
<thead>
<tr>
<th>Crime type</th>
<th>Local Law Enforcement</th>
<th>National Security Level</th>
</tr>
</thead>
<tbody>
<tr>
<td>Increase</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Traffic Violations</td>
<td>Driving under influence (DUI), fatal/personal</td>
<td></td>
</tr>
<tr>
<td></td>
<td>injury and traffic accident.</td>
<td></td>
</tr>
<tr>
<td>Sex Crime</td>
<td>Sexual offenses, sexual assaults, child</td>
<td>Organized prostitution, people smuggling.</td>
</tr>
<tr>
<td></td>
<td>molesting</td>
<td></td>
</tr>
<tr>
<td>Theft</td>
<td>Robbery, burglary, larceny, motor vehicle</td>
<td>Theft of national secrets or weapon information</td>
</tr>
<tr>
<td></td>
<td>theft, stolen property</td>
<td></td>
</tr>
<tr>
<td>Fraud</td>
<td>Forgery and counterfeiting, fraud, embezzlement</td>
<td>Transnational money laundering, transnational financial</td>
</tr>
<tr>
<td></td>
<td>and identity deception</td>
<td>fraud</td>
</tr>
<tr>
<td>Arson</td>
<td>Arson on buildings, apartments</td>
<td></td>
</tr>
<tr>
<td>Organized Crime</td>
<td>Narcotic drug offenses (sales or possession),</td>
<td>Transnational drug trafficking, terrorism (bioterrorism,</td>
</tr>
<tr>
<td></td>
<td>gang-related offenses</td>
<td>bombing, hijacking, etc.)</td>
</tr>
<tr>
<td>Violent Crime</td>
<td>Criminal homicide, armed robbery,</td>
<td>Terrorism</td>
</tr>
<tr>
<td></td>
<td>aggravated assault, other assaults</td>
<td></td>
</tr>
<tr>
<td>Cyber Crime</td>
<td>Internet fraud (e.g., credit card fraud,</td>
<td></td>
</tr>
<tr>
<td></td>
<td>advance fee fraud), illegal trading, network</td>
<td></td>
</tr>
<tr>
<td></td>
<td>intrusion/hacking, virus spreading, cyber-</td>
<td></td>
</tr>
<tr>
<td></td>
<td>piracy, cyber-pornography, cyber-</td>
<td></td>
</tr>
<tr>
<td></td>
<td>terrorism.</td>
<td></td>
</tr>
</tbody>
</table>

Crime remains elusive and ever strives to hide itself in the face of development. As measures and techniques for detecting crimes and criminals advance, criminals also look for means of hiding from these measures. (Longe & chiemeke, 2008). Criminals may escape arrest by using false identities; drug smugglers may enter any country by holding counterfeited passports or visas. Organized crimes, such as terrorism and narcotics trafficking, are often diffuse geographically, resulting in common security concerns across cities, states, and countries. Cybercrimes can pose threats to public safety across multiple jurisdictional areas due to the widespread nature of computer networks. International and domestic terrorism, in particular, often involves multiple crime types (e.g., identity theft, money laundering, arson and bombing, organized and violent activities, and cyber-terrorism) and causes great damage.

Table above summarizes the different types of crimes sorted by the degree of their respective public influence (Chen et al., 2004).

3. THEORETICAL FRAMEWORK
Analysis of police and policing should begin with careful delineation of the two interrelated concepts and phenomena. Police refers to a socio-political and quasi-legal institution – state agencies charged primarily with the enforcement of criminal law and the maintenance of order. Many quasi-police agencies such as the Custom and Immigration organizations and economic regulatory agencies are also involved in public policing.
Analytically, policing refers to measures and actions taken by a variety of institutions and groups (both formal and informal in society) to regulate social relations and practices in order to secure the safety of members of community as well as conformity to the norms and values of society. It is therefore a “sub-set of control processes” which involves “the creation of systems of surveillance coupled with the threat of sanctions for discovered deviance – either immediately or in terms of the initiation of penal process or both” (Reiner, 2000). State agencies designated as police as well as community groups are involved in policing. But community policing groups who carry out activities aimed at safety and social order do not constitute police.

No society can do without policing. However, historical evidence indicates that societies have existed without formal police forces. The danger of ‘police fetishism’ should be avoided so that the capacity of society for evolving variety of policing organization and strategies is not undermined. According to Reiner (2000):

Modern societies are characterized by what can be termed ‘police fetishism’, the ideological assumption that the police are a functional prerequisite of social order so that without a police force chaos would ensure. In fact, many societies have existed without a formal police force of any kind, and certainly without the present model. …It is important to distinguish between the ideas of ‘police’ and ‘policing’. ‘Police’ refers to a particular kind of social institution, while ‘policing’ implies a set of processes with specific social functions. Police are not found in every society, and police organizations and personnel can have a variety of shifting forms. ‘Policing’, however, is arguably a necessity in any social order, which may be carried out by a number of different processes and institutional arrangements. A state-organized specialist police organization of the modern is only one example.

3.1 Data Mining for Intelligence-Led Policing

In this digital era, police forces have access to a rapidly growing amount of data. Combined with the dynamic nature and complexity of criminal behavior, this sets the stage for successful data mining applications. Still, examples of consistently used police data mining implementations are scarce. There are numerous reasons why interest in intelligence-led policing has grown in recent years. Certainly a central factor in the desire to explore new approaches to crime control (including problem-oriented policing and community policing) stems from the recognized ineffectiveness of the standard model of policing. The paucity of evidence that a reactive and investigative approach to policing has any impact on the level of crime has led to considerable exploration of alternative models of policing (Weisburd and Eck, 2004).

The financial constraints imposed on police departments during the rapid increases in recorded crime in the 1970s and 1980s added further impetus to the search for alternative models (Ratcliffe, 2002). With demands on the police far outstripping the resources available, operational commanders looked to new policing methods to reduce crime in lieu of increased personnel and resources.

New technologies increased the volume of information and capacity of inform analysts, and the amount of data available to policy chiefs, and these developments helped spur interest in analytical approaches to problem identification and definition commonly known problem-oriented policing (Eck and Spelman, 1987; Goldstein, 1979, 1990; Skogan, 1996) and Compstat (Bratton, 1998; Maple and Mitchell, 1999; Moore, 2003; Silverman, 2006; Walsh, 2001). It has been argued that problem-oriented policing lacks the evidentiary base for widespread adoption (Sherman, 1998), and while a body of knowledge is certainly growing in this area, the adoption of problem-oriented policing has been at best one of slow emergence rather than one of rapid and enthusiastic embracing by law enforcement (Scott, 2000; Townsley et al., 2003). By contrast, intelligence-led policing has been slowly and enthusiastically adopted (Cope, 2004; Gill, 2000; Heaton, 2000; IACP, 2002; Loyka et al., 2005; Maguire and John, 2006; Ratcliffe, 2003; Smith, 1994).

Intelligence-led policing is a conceptual framework for carry out the business of policing. It is not a tactic in the way saturation patrolling is, nor is it a crime reduction strategy in the way situational crime prevention is. Rather, it is a business model (John and Maguire, 2003) and an information-organizing process that allows police agencies to better understand their crime problems and take a measure of the resources available to be able to decide on an enforcement tactic or prevention strategy best designed to control crime.

Intelligence-led policing has its origins in the UK where two influential government reports identified many of the problems associated with traditional policing (Audit Commission, 1993; HMIC, 1997). In suggesting a remedy, the Audit Commission recommended a focus on the offender rather than a focus on reported crime. As such, they recommended more resources be made available to support intelligence-gathering and the proactive targeting of prolific offenders (Hale et al., 2004). A central tenet of a proactive approach to crime management is therefore the identification and targeting of the ‘criminally active’ subpopulation as part of a broad crime reduction model (Amey et al., 1996; Maguire and John, 2006; Oakeisen et al., 2002; Ratcliffe, 2003; Maguire and John, 2006).

3.2 Police Need for Data Mining

The increasing adoption of the Intelligence-Led Policing model by most of the developed countries puts analysis at the heart of operational, tactical and strategic decision-making. In the model, intelligence serves as a guide to operations, rather than the reverse. Therefore, it is now more important than ever to find out how data mining can help create better understanding and predictions in the course of their policing. Traditionally, police systems focus on small parts of the available data (e.g. year, month, type of crime) for a specific purpose (e.g. monitoring crime rates for strategy). Without data mining, the amount of data used in analysis is limited by the time that analysts have to go through it, step by step. It is simply unfeasible to analyze all the potentially useful data by hand. However, for many policing problems it is important to use as much data as possible, to be able to explain, understand, link and predict.
The explanation of a phenomenon (e.g. the sudden increase of pickpocket activity) typically lies in small details, for example in the fact that in the recent period street festivals took place, with many potential pickpocket victims on the streets. This shows that by using more data, patterns offer more contextual information and help analysts reach the right conclusions. So, to understand crime, data is needed that goes beyond simple aspects of an incident or person, e.g. the type of neighborhood, the Modus Operandi (MO or manner of working), witness descriptions, stolen goods, vehicles involved and the background of the people involved (history, crime profile, socio-demographic profile). Linking crimes by similarity also benefits from rich data for the same reasons. Finding links can help to detect crime series, connect cases and solve them. This is where data mining comes in. Automated pattern recognition is necessary to turn the data overload into a manageable flow of information that matters.

4. SHORTCOMINGS OF CRIME ANALYSIS SYSTEMS

Traditional method of crime analysis tool sets suffers from the following issues:

- **Based on selecting variables**
  Traditional analytical tools require the analyst to look at variables one by one. This way of working is not viable for rich data sets containing many variables.

- **Static results**
  Existing systems usually generate static reports that do not allow interaction. They cannot be used to find the explanations behind the numbers they present.

- **Based on simple patterns**
  When an analyst focuses on one or two variables, the traditional tools allow only the analysis of those individual variables. In rare cases, interaction between two chosen variables is analyzed, but all other combinations are not used which is why useful interactions between variables can be overlooked.

- **Difficult extraction**
  It is typically hard to extract data from police source systems because of old and diverse database systems with data models based on transactions instead of analysis. There are several standard analytical applications working on small extractions, but when analysis needs to go a step further, the analyst faces a challenge in getting the right data from the systems. The fact that there are many different systems in the organization adds to the challenge and so does poor data quality. Typically, extraction, linking, correction and preparation of data need to be carried out for each analysis.

In addition, police data typically challenges the user because of quality issues. We believe that the high demands for users of standard data mining tools are the main reasons why there seem to be just a few successful police mining applications and most of these applications are either academic endeavors or small applied projects - not continuous activities. Furthermore, when these applications do appear to be ongoing activities they seem to be limited to a single police force.

5. EMERGING ISSUES

The concern about national security has increased significantly since the 9/11 attacks in the United States of America. However, information overload hinders the effective analysis of criminal activities. A major challenge facing all detectives and law-enforcement is accurately and efficiently analyzing the growing volumes of crime data. For example, complex conspiracies are often difficult to unravel because information on suspects can be geographically diffuse and span long periods of time. Without a suspected crime pattern, the detective is less likely to build the complete picture from bits of information from different crime incidents. Today most of it is manually done with the help of multiple spreadsheet reports with their own crime logs.
5.1 Research Direction

We propose data mining as a technique that can help solve crimes faster, more specifically, by using clustering based models to help in identification of crime patterns. The specific objective is to use clustering technique since crimes vary in nature widely and crime database often contains several unsolved crimes. Therefore, classification technique that will rely on the existing and known solved crimes, will not give good predictive quality for future crimes. Also nature of crimes change over time, such as Internet based cyber crimes or crimes using cell-phones were uncommon not too long ago. Thus, in order to be able to detect newer and unknown patterns in future, clustering techniques will work better.

We propose using data mining to detect crime patterns since there are many attributes or factors for crime and often there is partial information available about the crime. In a general case it will not be easy for a detective to identify these patterns by simple manual querying. Thus clustering technique using data mining comes in handy to deal with enormous amounts of data and dealing with noisy or missing data about the crime incidents, hence the usage of data mining approach to support security detectives and law enforcement agents in solving crimes.

Employing data mining for criminal profiling activities will involve a review of criminal cases, academic writings, and press publications in the subject area. Personal interviews, including security specialists, law enforcement agents and legal counsels. Survey of the crime database of a Police criminal Department, Nigeria Police Force will be done. Clustering based models can then be applied to detect the crimes patterns so as to aid the detectives and intelligence analysis in speeding up the process of solving crime. Datasets for multiple years will have to be formatted and the analysis carried out using data mining software such as “WAIKATO Environment for Knowledge Analysis (WEKA)” in order to detect the crime patterns.

5.2 Concluding Remarks

A research directed at crime profiling will assist in developing pattern detection as a machine learning task thereby using data mining to support security detectives and law enforcement agents in solving crimes. The proposed data Mining technique will be able to identify the crime patterns from a large number of crimes databases and improve the productivity of detectives and other law enforcement officers.

REFERENCES


Authors Bio

Fagbaibi Sheriff Olumide holds a Master’s Degree in Computer Science from the University of Ibadan. He is currently an Examination Officer (CSD), at The West African Examinations Council (WAEC), Ibadan Zonal Office, Ibadan, Nigeria. His research interest is on Data Mining and Network Security. He can be reached by phone on +2348068932999 and through E-mail sherifolumide@gmail.com

Mr. Yahya Lawal holds a Master’s degree in Computer Systems from the University of Ibadan, Nigeria. He is a system administrator at the ICT Unit, Kogi State University, Ayungba and a graduate of the Department of Mathematical Sciences of the same institution. He can be reached at He can be reached through e-mail at enemakwu@gmail.com. Phone on +2348033572346.

Dr. Longe Olumide is on Faculty at the Department of Computer Science, University of Ibadan, Nigeria. His research has focused on using social theories and computer security theories to explain causation and apprehension of cyber crimes. He is currently a Fulbright Fellow at the College of Business, Southern University, Baton Rouge, LA, USA. He can be reached by phone on +18572078409 and through E-mail longeolumide@fulbrightmail.org