

Survey: Tools and Techniques implemented in Crime Data Sets

M.Ramzan Begam¹

Research Scholar, Bharathiar University,
Coimbatore, Tamilnadu

Dr. P.Sengottuvelan²

Associate Professor, Bannari Amman Institute of Technology,
Sathyamangalam, India

T. Ramani³

Research Scholar, Bharathiar University,
Coimbatore, Tamilnadu

Abstract

Crime is a major issue where the top priority has been given by our government. We analyze the various data mining techniques implemented for crime data base. The present research paper proposes the use of optimization data mining techniques for developing such a crime analysis tool. This survey paper gives an overview of the Tools and Techniques that are implemented in the Crime data mining. The developed tool has promising value in the current changing crime scenario and can be used as an effective tool by Indian police and enforcement of law organizations for crime detection and prevention.

Keywords: *criminals, techniques, tool , data mining*

1. Introduction

Crime incidents are reported per day is increasing significantly. The Department of police is the major organization of preventing crimes; it is very difficult to find a country without crime free society. The present society has also packed with various kinds of crimes. The Police department is responsible for enhancing security, the public became feel safety, maintaining order and keeping the silence.

2. Police Structure

Superintendence over the police force in the State is exercised by the State Government. The head of the police force in the State is the Director General of Police (DGP). States are divided territorially into administrative units known as districts. A group of districts form a range, which is looked after by an officer of the rank of Deputy Inspector General of Police (DIGP). Some States have zones comprising two or more ranges, under the charge of

an officer of the rank of an Inspector General of Police (IGP). A Senior Superintendent of Police (SSP)/Superintendent of Police (SP) is the head of the district police administration and is assisted by an Assistant Superintendent of Police (ASP) and few Deputy Superintendents of Police (DSP).

A district may have many Police Stations that are manned by Inspectors, Sub Inspectors, Assistant Sub Inspectors, Head Constables and Constables. Police Station is the basic unit of police administration through which both crime and non crime duties are discharged. Police Stations are the places where complaints and First Information Reports (FIRs) are lodged. Police Stations also serve as the window of 'citizen interface' for the police. Common people approach Police Stations for assistance. Therefore, public expectations from police stations are more direct, pressing and at times extremely demanding. Operationally, police stations are at the nucleus of all policing activities. All important operational duties – be it duties to the State or services to other government departments or citizens- are executed and coordinated through police stations

3. Challenges and Approach

The most incredible challenge for the police department is investigating crime with the current technologies. Police organization everywhere has been handling large amount of information and huge volume of records. They need advanced technologies to handle crime. An ideal crime recognition system should be used to identify crime pattern quickly for future crime pattern detection and action.

Data mining technique can help discovery and exploitation of knowledge. In the process of data mining, the most important step is the data preprocessing.

The main focus is to develop a crime recognition system that assists the police in

1. Detecting crime patterns and performs crime analysis.
2. Provide information to formulate strategies for crime prevention and reduction.
3. Identify and analyze common crime patterns to reduce further occurrences of similar incidence.

4.Data source

National Crime Record Bureau (NCRB) was created with the aim to help the investigating agencies by providing with updated information on crime and criminal data. Crime Criminal Information System (CCIS) providing information technology training to senior police officers and law enforcement agencies.

Data Used

The daily recorded crime data is used for the study. The collected data set provides the following:

- (i) Location of Crime
- (ii) Street address
- (iii) Time of the crime
- (iv) Type of Crime
- (v) Geographical information (x, y) coordinates

Dataset

The techniques are applied on real dataset on non-disclosure agreement from the crime archive system. Non spatial data / Attribute information from Police records:

- CrimeID :Individual Crimes are designated by unique Crime IDs
- CrimeName :Disguised crime's name
- Gender:Belongs to which gender.
- Age:Age of Individual criminal.
- Height:Hight of Individual Criminal.
- Location:Location of Individual criminal .
- CrimeType :Type of Crime committed by the criminal
- WeaponUsed : Type of weapon a criminal used.
- Investigator ID: Key identifying the Investigator
- Area: Location or spatial data about the incident area
- Date of reporting : Time period of the incident
- Affected person : details about the victim
- Witness ID: if any witness, Key identifying the witness

- Criminal name : identified criminal of an existing crime
- Motivation: motivation of crime
- Weapons used : types of weapons used for the crime
- Evidence : available evidences recorded

5. Related Work

It is important to study the previous related works to both learn from the experience of others and to add something to our existing body of knowledge. Existing literature has been reviewed in three different areas: crime data mining, data extraction and data focus.

Recent developments in crime control applications aim at adopting data mining techniques to aid the process of crime investigation. COPLINK is one of the earlier projects which is collaborated with Arizona University and the police department to extract entities from police narrative records. Bruin, Cocx and Koster et al. presented a tool for changing in offender behavior. Extracted factors including frequency, seriousness, duration and nature have been used to compare the similarity between pairs of criminals by a new distance measure and cluster the data accordingly.

Brown proposed a framework for regional crime analysis program (ReCAP). The data mining was adopted as an algorithm for crime data analysis. J.S.de Bruin et.al compared all individuals based on their profiles to analyze and identify criminals and criminal behaviors . Nath et.al used K-means clustering to detect crime pattern to speed up the process of solving crimes. Adderley and Musgrove applied Self Organizing Map (SOM) to link the offenders of serious sexual attacks. Recently, Ozgul et.al proposed a novel prediction model CPM (Crime Prediction Model) to predict perpetrators of unsolved terrorist events on attributes of crime information that are location, date and modus operandi attributes. LianhangMa, Yefang Chen, and Hao Huang et.al presented a two phase clustering algorithm called AK-modes to automatically find similar case subsets from large datasets. In the attribute-weighting phase, the weight of each attribute related to an offender's behavior trait using the concept Information Gain Ratio (IGR) in classification domain is computed. The result of attribute weighing phase is utilized in the clustering process to find similar case subsets.

Crime Classification:

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.

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.

6. DATA MINING TECHNIQUES FOR ANALYSE OF CRIME DATA

Data mining is basically used to find out unknown patterns from a large amount of data. There are popular tools of data mining to rub data mining algorithms. There are two approaches for the implementation of data mining, first is to copy data from data warehouse or source and mine it. Other approach is to mine the data within a data warehouse. There are various data mining techniques available as follows:

Classification Maps a data item into one of several pre-defined categories. These algorithms normally output "classifiers", for example, in the form of decision trees or rules. An ideal application in intrusion detection will be to gather sufficient "normal" and "abnormal" audit data for a user or a program, then apply a classification algorithm to learn a classifier that will determine (future) audit data as belonging to the normal class or the abnormal class.

Clustering and segmentation is used to segment a database into subsets, or clusters based on set of attributes.

It is a method to group data into classes with identical characteristics in which the similarity of intra-class is maximized or minimized.

Association identifies affinities/ associations among the collection of data as reflected in the examined records. A result is patterns describing rules of association in data.

Decision Tree is predictive model that can be viewed as tree, each branch is a classification question and leaves of the tree are partitions of data set with their classification. It divides data on each branch point without losing any of the data. The number of churners and non churners is conserved as we move up or down the tree. ID 3, C4.5, CART and CHAID are some algorithms used in this technique.

Neural Networks are biological systems that detect patterns, make predictions and learn. The artificial neural networks are computer programs implementing sophisticated pattern detection and machine learning algorithms on a computer to build predictive models for historical databases

Link analysis it determines relations between fields in the database. Finding out the correlations in audit data will provide insight for selecting the right set of system features for intrusion detection.

Sequence analysis it models sequential patterns. These algorithms can help us understand what (time-based) sequence of audit events are frequently encountered together. These frequent event patterns are important elements of the behavior profile of a user or program

Table2: Analysis of existing tools & techniques

Digital Forensic Techniques	Data Mining Techniques	Tool
Data Recovery, data generation and preprocessing	Statistical Test of sphericity Bartlett's test of sphericity Kaiser-Meyer-Olkin (KMO)	Recuva FTK Encase Sleuth kit/Autopsy ProDiscover
Data Analysis	Clustering – K-means, EM, Hierarchical Clustering	Weka
	Classification – Supervised learning -	Weka

	Decision Tree, Neural Networks, SVM, Naïve Bayesian	
	Unsupervised learning – PCA, Karnohuen Map	-
	Frequent Pattern Mining/Association rule Mining - Apriori, Eclat	Weka
	Named Entity recognition	LingPipe
	Visualization	CyberForensicTimeLab
	Statistical Analysis and Anamoly Detection	EMT/MET
	Recursive data mining	-
	Phishing	Invisible Witness
	Regression	-

4. Faith Ozgul, Claus Atzenbeck, AhmetCelik, Zeki Erdem, “Incorporating data Sources and Methodologies for Crime Data Mining,” IEEE proceedings, 2011.
5. H.Chen, W.ChungXu, G.Wang, Y.Qin and M.Chen, “Crime Data Mining: A General Framework and Some Examples,” computer, vol.37, 2004.
6. H.Chen, W.Chung, Y.M.Chan, J.Xu, G.ang, R.Zheng and H. Atabakch,” Crime Data Mining: An Overview and Case Studies,” in proceedings of the annual national
7. M.Chan, J.Xu,and H.Chen,”Extracting Meaningful Entities from Police Narrative Reports,” in proceedings of the National Conference on Digital Government Research,pp.271-275,2002.
8. R.Adderly and P.B. Musgrove, “Data Mining Case Study: Modeling the behavior of offenders who commit sexual assaults,” in proceedings of the 2006 IEEE/WIC/ACM Conference on Web Intelligent Agent Technology, pp.41-44, 2006.

7. Conclusion:

We conclude this paper that we analyze about the techniques implemented in data mining by using crime data. Police is the major organization to control crimes occurs in a society. Here we analyze crime data base. We give the details about where we get data resource. We give the types of crime and how it handles by data mining tool. By using this it is very useful to detect crime patterns and analyze crime analysis. It reduce further occurrence of similar incidence. It formulates strategies for crime prevention and reduction.

8. Reference:

1. 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.
2. J. Han and M. Kamber, “Data Mining: Concepts and Techniques,” Morgan Kaufmann publications, pp. 1-39, 2006.
3. J. Mena , “ Investigative Data Mining for Security and Criminal Detection”, Butterworth Heinemann Press, pp.15-16, 2003.