Survey of Identifying Criminal Pattern using Data Mining Algorithm

Vibhuti Harishanker Jani
Dep. of Computer Engineering, SOCET

Abstract- **Data mining plays an important role in terms of prediction and analysis. Data mining is a process of extracting knowledge from huge amount of data stored in databases, data warehouses and data repositories. Crimes are a social nuisance and cost our society dearly in several ways. Crime is an interesting application where data mining plays an important role in terms of prediction and analysis. 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. This paper presents detailed study on clustering techniques and its role on crime identification. "There is not a crime, there is not a dodge, there is not a trick, there is not a swindle, there is not a vice which does not live by secrecy."**

I. INTRODUCTION

Crime is one of the dangerous factors for any country. Security of citizens is the major concern of the police. Crime analysis is the activity in which analysis is done on crime activities. Today criminals have maximum use of all modern technologies and hi-tech methods in committing crimes. It is impossible to find a country which has a crime- free society. As long as human beings have feelings they incline on attempting crimes. So the present society has also filled with various kinds of crimes. Hence, creation of data base for crimes and criminals is needed. Developing a good crime analysis tool to identify crime patterns quickly and efficiently for future crime pattern detection is challenging field.

Crime analysis is a process of gathering and analyzing recurring patterns in crime. Crime analysis aids police in intelligently making decisions and proactively planning the defense mechanism. We can analyze web information from the perspective of events and apply some research results related to the events to solve the problem of web crime mining. A web page involving a crime can be thought of as a chain of actions with series of background attributes.

II. OVERVIEW OF CRIME

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. These crimes can either be one-offs or serial. Serial crimes are 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.

The police officers or detectives use free text to record most of their observations that cannot be included in checkbox kind of pre-determined questions. While the first two categories of information are usually stored in the computer databases as numeric, character or date fields of table, the last one is often stored as free text. The challenge in data mining crime data often comes from the free text field. While free text fields can give the newspaper columnist, a great story line, converting them into data mining attributes is not always an easy job.

Crime Type	Description
Traffic Violations	Driving under the influence of alcohol, fatal/personal injury/property damage traffic accident, road rage
Sex crime	Sexual offences
Arson	Arson on buildings
Crime Type	Description
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.

Table 1: The classification of crime[7]

III. 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[5]. This information then may help with various purposes, such as the prediction of future events or behaviors.

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.

IV. CLUSTERING TECHNIQUE OF CRIME

Data clustering is a process of putting similar data into groups. A clustering algorithm partitions a data set into several groups such that the similarity within a group is larger than among groups. Clustering can also be considered the most important unsupervised learning technique; so, as every other problem of this kind, it deals with finding a structure in a collection of unlabeled data. There are so many techniques used in clustering, such as K-means algorithm, Ak-Mode Algorithm, Expectation-Maximization Algorithm [1].

There are many ways to enhance the basic K-means algorithm. But to keep things simple, we chose a very simple and efficient implementation of the K-means algorithm. For instance, we select our initial centroids by randomly choosing K documents [2]. Our reason is that we noticed that incremental updates were more effective, i.e., produced results with better overall similarity and lower entropy. For what follows we will use a bisecting K-means algorithm as our primary clustering algorithm. Note that the bisecting K-means algorithm can produce either an un-nested (flat) clustering or a hierarchical clustering. For un-nested clusters we will often "refine" the clusters using the basic K-means algorithms, but we do not refine the nested clusters. We will provide more details later. Finally, note that bisecting K-means has a time complexity which is linear in the number of documents. If the number of clusters is large and if refinement is not used, then bisecting K-means is even more efficient than the regular K-means algorithm (In this case, there is no need to compare every point to every cluster centroid since to bisect a cluster we just consider the points in the cluster and their distances to two centroids.)

V. CONCLUSION

Crime data is a sensitive domain where efficient clustering techniques play vital role for crime analysts and law-enforcers to precede the case in the investigation and help solving unsolved crimes faster. Similarity measures are an important factor which helps to find unsolved crimes in crime pattern. For identifying and predicting crime pattern we will use Genetic algorithm. This paper deals with detailed study about importance of clustering and similarity measures in crime domain identify and predict criminal patterns using Genetic algorithm. **"There are crimes of passion and crimes of logic, The boundary between them is not clearly defined."**

REFERENCES

[1] Z. Huang, "Extensions to the K-means algorithm for clustering large datasets with categorical values, "Data Mining and Knowledge Discovery, vol.2, pp.283-304, 1998.

[2] S.V. Nath, "Crime pattern detection using data mining," in Proceedings of the 2006 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology, pp. 41-44,2006.

[3] Dr. A. Bharathi, International Journal of Advance Research in Computer Science and Management Studies Volume 2, Issue 8, August 2014.

[4] Revatthy Krishnamurthy, International Journal of Data Mining Techniques and Applications, Vol 01, Issue 02, December 2012.

[5] 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.

[6] Oren Zamir, Oren Etzioni, Omid Madani, Richard M. Karp, Fast and Intuitive Clustering of Web Documents, KDD '97, Pages 287-290, 1997.

[7] Okwangale Fredrick R, Department of Networks ICT and Legal Applications, Makerere University.

[8] 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.