

Datamining Techniques to Analyze and Predict Crimes

¹S.Yamuna, ²N.Sudha Bhuvaneshwari

¹M.Phil (CS) Research Scholar

School of IT Science,

Dr.G.R.Damodaran College of science Coimbatore

²Associate Professor MCA, Mphil (cs), (PhD)

School of IT Science

Dr.G.R.Damodaran College of science Coimbatore

Abstract

Data mining can be used to model crime detection problems. Crimes are a social nuisance and cost our society dearly in several ways. Any research that can help in solving crimes faster will pay for itself. About 10% of the criminals commit about 50% of the crimes. Data mining technology to design proactive services to reduce crime incidences in the police stations jurisdiction. Crime investigation has very significant role of police system in any country. Almost all police stations use the system to store and retrieve the crimes and criminal data and subsequent reporting. It becomes useful for getting the criminal information but it does not help for the purpose of designing an action to prevent the crime. It has become a major challenge for police system to detect and prevent crimes and criminals. There hasn't any kind of information is available before happening of such criminal acts and it results into increasing crime rate. Detecting crime from data analysis can be difficult because daily activities of criminals generate large amounts of data and stem from various formats. In addition, the quality of data analysis depends greatly on background knowledge of analyst, this paper proposes a guideline to overcome the problem.

Keywords – Analyze, crime, decision, investigation, link, report,

Date of Submission: 11, December, 2012



Date of Publication: 25, December 2012

I INTRODUCTION

Criminology is an area that focuses the scientific study of crime and criminal behavior and law enforcement and is a process that aims to identify crime characteristics. It is one of the most important fields where the application of data mining techniques can produce important results. Crime analysis, a part of criminology, is a task that includes exploring and detecting crimes and their relationships with criminals [1].

The high volume of crime datasets and also the complexity of relationships between these kinds of data have made criminology an appropriate field for applying data mining techniques. Identifying crime characteristics is the first step for developing further analysis. The knowledge that is gained from data mining approaches is a very useful tool which can help and support police forces. According to Nath (2007), solving crimes is a complex task that requires human intelligence and experience and data mining is a technique that can assist them with crime detection problems. The idea here is to try to capture years of human experience into computer models via data mining. The criminals are becoming technologically sophisticated in committing crimes. Therefore, police needs such a crime analysis tool to catch criminals and to remain ahead in the eternal race between the criminals and the law enforcement [1]. The police should use the current technologies to give themselves the much-needed edge. Availability of relevant and timely information is of utmost necessity in conducting

of daily business and activities by the police, particularly in crime investigation and detection of criminals. Police organizations everywhere have been handling a large amount of such information and huge volume of records.

An crime analysis should be able to identify crime patterns quickly and in an efficient manner for future crime pattern detection and action. crime information that has to be stored and analyzed, Problem of identifying techniques that can accurately analyze this growing volumes of crime data, Different structures used for recording crime data.

II CRIME PREDICTION

The prediction of future crime trends involves tracking crime rate changes from one year to the next and used data mining to project those changes into the future. The basic method involves cluster the states having the same crime trend and then using "next year" cluster information to classify records. This is combined with the state poverty data to create a classifier that will predict future crime trends. To the clustered results, a classification algorithm was applied to predict the future crime pattern. The classification was performed to find in which category a cluster would be in the next year. This allows us to build a predictive model on predicting next year's records using this year's data. The decision tree algorithm was used for this purpose [2]. The generalized tree was used to predict the unknown crime trend for the next year. Experimental results proved that the technique used for prediction is accurate and fast.

III MATCHING CRIMES

The ability to link or match crimes is important to the Police in order to identify potential suspects, asserts that 'location is almost never a sufficient basis for, and seldom a necessary element in, prevention or detection', and that non-spatial variables can, and should be, used to generate patterns of concentration. To date, little has been achieved in the ability of 'soft' forensic evidence to provide the basis of crime linking and matching based upon the combinations of locational data with behavioural data.

IV CRIME FRAMEWORK

Many efforts have used automated techniques to analyze different types of crimes, but without a uni-fying framework describing how to apply them. In particular, understanding the relationship between analysis capability and crime type characteristics can help investigators more effectively use those techniques to identify trends and patterns, address problem areas, and even predict crimes.

The framework shows relationships between datamining techniques applied in criminal and intelligence analysis and the crime types, there were four major categories of crime data mining techniques: entity extraction, association, prediction, and pattern visualization. Each category represents a set of techniques for use in certain types of crime analysis. For example, investigators can use neural network techniques in crime entity extraction and prediction[3]. Clustering techniques are effective in crime association and prediction. Social network analysis can facilitate crime association and pattern visualization. Investigators can apply various techniques independently or jointly to tackle particular crime analysis problems.

V. Clustering Techniques

Given a set of objects, clustering is the process of class discovery, where the objects are grouped into clusters and the classes are unknown beforehand. Two clustering techniques, K-means and DBScan algorithm are considered for this purpose.

The algorithm clusters the data into m groups where m is predefined.

Input – Crime type, Number of Clusters, Number of Iteration

Initial seeds might produce an important role in the final result.

Step 1: Randomly Choose cluster centers

Step 2: Assign instances to clusters based on their distance to the cluster centers.

Step 3: centers of clusters are adjusted.

Step 4: go to Step 1 until convergence

Step 5: Output X0, X1, X2, X3

Output:

X0: Crime is steady or dropping. The Sexual Harassment rate is the primary crime in flux. There are lower incidences of: Murder for gain, Dacoity, rape.

X1: Crime is rising or in flux. Riots, cheating, Counterfeit. There are lower incidences of: murder and kidnapping and abduction of others[4].

X2: Crime is generally increasing. Thefts are the primary crime on the rise with some increase in arson. There are lower incidences of the property crimes: burglary and theft.

X3: Few crimes are in flux. Murder, rape, and arson are in flux. There is less change in the property crimes: burglary, and theft. To demonstrate at least some characteristics of the clusters.

From the clustering result, the city crime trend for each type of crime was identified for each year. Further, by slightly modifying the clustering seed, the various states were grouped as high crime zone, medium crime zone and low crime zone.

Output Function of Crime Rate = $1/\text{Crime Rate}$

Crime rate is obtained by dividing total crime density of the state with total population of that state since the police of a state are called efficient if its crime rate is low i.e. the output function of crime rate is high.

Clustering techniques were analyzed in their efficiency in forming accurate clusters, speed of creating clusters, efficiency in identifying crime trend, identifying crime zones, crime density of a state and efficiency of a state in controlling crime rate.

VI. VALIDATION

Validation is a key component to providing feasible confidence that any new method is effective at reaching a viable solution, in this case a viable solution to the malicious detection problem. Validation is not only comparing the results to what the expected result should be, but it is also comparing the results to other published methods.

The values which include true positive rate (TPR), false positive rate (FPR), accuracy and precision. TPR, also known as recall, "is the proportion of relevant data retrieved, measured by the ratio of the number of relevant retrieved data to the total number of relevant data in the data set." In other words TPR is the ratio of actual positive instances that were correctly identified. FPR is the ratio of negative instances that were incorrectly identified[5]. Accuracy is the ratio of the number of positive instances, either true positive or false positive, that were correct. "Precision is the proportion of retrieved data that are relevant, measured by the ratio of the number of relevant retrieved applications to the total number of retrieved applications," or the ratio of predicted true positive instances that were identified correctly.

		Actual	
		Positive	Negative
Predicted	Positive	a	B
	Negative	c	D

All of these values are derived from information provided from the truth table. A truth table, also known as a confusion matrix, provides the actual and predicted classifications from the predictor.

$$TPR = \frac{a}{a+b}$$

$$FPR = \frac{b}{b+d}$$

$$Accuracy = \frac{a+d}{a+b+c+d}$$

$$Precision = \frac{a}{a+b}$$

where, a(true positive) is the number of malicious applications in the data set that were classified as malicious applications, b(false positive) is the number of benign applications in the data set that were classified as malicious applications, c(false negative) is the number of malicious applications in the data set that were classified as benign applications, and d(true negative) is the number of benign applications in the data set that were classified performance calculations that were used in effort for validation of the predicted results.

VII. DATA MINING TECHNIQUES

Entity extraction has been used to automatically identify person, address, vehicle, narcotic drug, and personal properties from police narrative reports. Clustering techniques have been used to automatically associate different objects (such as persons, organizations, vehicles) in crime records. Deviation detection has been applied in fraud detection, network intrusion detection, and other crime analyses that involve tracing abnormal activities. Classification has been used to detect email spamming and find authors who send out unsolicited emails[6]. String comparator has been used to detect deceptive information in criminal records. Social network analysis has been used to analyze criminals' roles and associations among entities in a criminal network.

VIII. CRIME DETECTION

Intelligence agencies are actively collecting and analyzing information to investigate terrorists activities. Local law enforcement agencies have also become more alert to criminal activities in their own jurisdictions. When the local criminals are identified properly and restricted from their crimes, then it is possible to considerably reduce the crime rate.

Criminals often develop networks in which they form groups or teams to carry out various illegal activities. Data mining task consisted of identifying subgroups and key members in such networks and then studying interaction patterns to develop effective strategies for disrupting the networks. Data is used with a concept to extract criminal relations from the incident summaries and create a likely network of suspects[6]. Co-occurrence weight measured the relational strength between two criminals by computing how frequently they were identified in the same incident.

IX. COMBATING CRIMES

Using data mining, various techniques and algorithms are available to analyze and scrutinize data. However, depending on the situation, the technique to be used solely depends upon the circumstance. Also one or more data mining techniques could be used if one is inadequate. Data mining applications also uses a variety of parameters to examine the data start investigation as to the likely causes of the attack and the individuals who might have responsible attack. We have stated that crime investigation remains the prerogative of the law enforcement agencies concern, but computer and computer analysis can be useful in solving detecting.

X. CLASSIFICATION OF DATA

Classification in a broad sense is a data mining technique that produces the characteristics to which a population is divided based on the characteristic. The idea is to define the criteria use for the segmentation of a population, once this is done, individuals and events can then fall into one or more groups naturally. classification divides the population (dataset) based on some predefined condition.

When classification is used, existing dataset can easily be understood and it will in no doubt help to predict how new individual or events will behave based on the classification criteria. Data mining creates classification models by examining already classified data (cases) and inductively finding a predictive pattern. These existing cases may come from an historical database, such as people who have already undergone a particular medical treatment or moved to a new long-distance service[7]. They may come from an experiment in which a sample of the entire database is tested in the real world and the results used to create a classifier.

XI. DISCOVERING DISCRIMINATION

Discrimination discovery is about finding out discriminatory decisions hidden in a dataset of historical decision records. The basic problem in the analysis of discrimination, given a dataset of historical decision records, is to quantify the degree of discrimination suffered by a given group (e.g. an ethnic group) in a given context with respect to the classification decision (e.g.intruder yes or no).

Discrimination techniques should be used in training data are biased towards a certain group of users (e.g.young people), the learned model will show discriminatory behavior towards that group and most requests from young people will be incorrectly classified as intruders.

Additionally, anti-discrimination techniques could also be useful in the context of data sharing between IDS (intrusion detection systems)[8].Assume that various IDS share their IDS reports that contain intruder information in order to improve their respective intruder detection models.

XII. LINK ANALYSIS

Link Analysis (LA) is another data mining technique that is useful in detecting valid and useful patterns. The theoretical framework of Link Analysis (LA) is based on the fact that events are linked to one another and are hence mutually exclusive.Link Analysis framework is that if A is link to B and B in linked to C and C to D, then A could be linked to D. link analysis can be employed by enforcement investigators and intelligence Analysts connect networks of relationships and contacts hidden in the data. link analysis one needs to reduce the graphs so that the analysis is manageable and not combinatorial explosive.

Link analysis can then be used to analyze the activities of individuals by forming a link of their activities. These links might be in form of telephone conversation, places visited, bank transactions etc.

XIII. FINANCIAL CRIME DETECTION

Financial crime here refers to money laundering, violative trading, and insider trading. The Financial Crimes operates with an expert system with Bayesian inference engine to output suspicion scores and with link analysis to visually examine selected subjects or accounts. Supervised techniques such as case-based reasoning, nearest neighbour retrieval, and decision trees were seldom used due to propositional approaches, lack of clearly labelled positive examples, and scalability issues. Unsupervised techniques were avoided due to difficulties in deriving appropriate attributes[9]. It has enabled effectiveness in manual investigations and gained insights in policy decisions for money laundering.

Use of large amounts of unstructured text and web data such as free-text documents, web pages, emails, and SMS messages, is common in adversarial domains but still unexplored in fraud detection literature. Link Discovery on Correlation Analysis which uses a correlation measure with fuzzy logic to determine similarity of patterns between thousands of paired textual items which have no explicit links. Itcomprises of link hypothesis, link generation, and link identification based on financial transaction timeline analysis to generate community models for the prosecution of money laundering criminals[9]. Relevant sources of data which can decrease detection time, expert systems and clustering it is for finding and predict early symptoms of insider trading in option markets before any news release.

XIV. CRIME REPORTING SYSTEMS

The data for crime often presents an interesting dilemma. While some data is kept confidential, somebecomes public information. Data about the prisoners can often be viewed in the county or sheriff's sites. However, data about crimes related to narcotics or juvenile cases is usually more restricted. Similarly, the informationabout the sex offenders is made public to warn others in the area, but the identity of the victim is often prevented. Thus as a data miner, the analyst has to deal with all these public versus private data issues so that data mining modeling process does not infringe on these legal boundaries[10]. 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, convertingthem into data mining attributes is not always an easy job. We will look at how to arrive at the significant attributes for the data mining models.

XV. CONCLUSION

Data mining applied in the context of law enforcement and intelligence analysis holds the promise of alleviating crime related problem. Using a wide range of techniques it is possible to discover useful information to assist in crime matching,not only of single crimes, but also of series of crimes.In this paper we use a clustering/classify based model to anticipate crime trends. The data mining techniques are used to analyze the crime data from database. The results of this data mining could potentially be used to lessen and even prevent crime for the forth coming years. we believe that crime data mining has a promising future for increasing the effectiveness and efficiency of criminal and intelligence analysis.

REFERENCES

- [1] Levine, N., 2002. CrimeStat: A Spatial Statistic Program for the Analysis of Crime Incident Locations (v 2.0). Ned Levine & Associates, Houston, TX, and the National Institute of Justice, Washington, DC. May 2002.
- [2] Ewart, B.W., Oatley, G.C., & Burn K., 2004. Matching Crimes Using Burglars' Modus Operandi: A Test of Three Models. Forthcoming.
- [3] Ratcliffe J.H., 2002. Aoristic signatures and spatio-temporal analysis of high volume crime patterns. *J. Quantitative Criminology*.
- [4] Chen, H., Chung, W., Xu, J.J., Wang,G., Qin, Y., & Chau, M., 2004. Crime data mining: a general framework and some examples. *IEEE Computer* April 2004, Vol 37.
- [5] Polvi, N., Looman, T., Humphries, C., & Pease, K., 1991. The Time Course of Repeat Burglary Victimization. *British Journal of Criminology*
- [6] Brown, D.E. The regional crime analysis program : A frame work for mining data to catch criminals," in *Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics*.
- [7] Abraham, T. and de Vel, O. Investigative profiling with computer forensic log data and association rules, *Proceedings of the IEEE International Conference on Data Mining*
- [8] Townsley, M. (2006) Spatial and temporal patterns of burglary: Hot spots and repeat victimization in an Australian police divison. Ph.D. thesis, Griffith University, Brisbane, Australia
- [9] Mannheim, H. *Comparative Criminology (Volume 1)*. London: Routledge and Kegan Paul.
- [10] De Bruin, J.S., Cocx, T.K., Kusters, W.A., Laros, J. and Kok, J.N. Data mining approaches to criminal career analysis," in *Proceedings of the Sixth International Conference on Data Mining*.

Biographies and Photographs

I am S.Yamuna, Pursuing Mphil(CS) in GRD college Of Science As a full time Scholar And I did my Post Graduate in Msc(CT)in Coimbatore Institute Of Technology, Coimbatore and did my Under graduate in BSC(CS) in GRD college of Science..And My Research Guide is N.Sudha Bhuvaneswari working as an Associate Professor MCA, Mphil (cs), (PhD) School Of IT Science
Dr.G.R.Damodaran College of science Coimbatore.