2011 International Conference on Information and Electronics Engineering
IPCSIT vol.6 (2011) © (2011) IACSIT Press, Singapore

An intelligent Analysis of a City Crime Data Using Data Mining

Malathi. A [1], Dr. S. Santhosh Baboo [2] and Anbarasi. A [3]

[1] Assistant professor, Department of Comuter Science, Govt Arts College, Coimbatore, India.

[2] Readers, Department of Computer Science, D. G. Vaishnav College, Chennai, India.

[3] Associative Professor, Department of Civil Engineering, Kumaraguru college of Technology, Coimbatore, India.

Abstract. There had been an enormous increase in the crime in the recent past. The concern about national security has increased significantly since the 26/11 attacks at Mumbai, India. However, information and technology overload hinders the effective analysis of criminal and terrorist activities. Crime deterrence has become an upheaval task. The cops in their role to catch criminals are required to remain convincingly ahead in the eternal race between law breakers and law enforcers. Data mining applied in the context of law enforcement and intelligence analysis holds the promise of alleviating such problem. In this paper we use a clustering/classify based model to anticipate crime trends. The data mining techniques are used to analyze the city crime data from Police Department. The results of this data mining could potentially be used to lessen and even prevent crime for the forth coming years.

Keywords: Cluster, Classify, Crime Analysis and weka

1. Introduction

The concern about national security has increased significantly since the terrorist attacks on November 26, 2008 at Mumbai. Intelligence agencies such as the CBI and NCRB (National Crime Record Bureau) are actively collecting and analyzing information to investigate terrorists' activities [12]. Local law enforcement agencies like SCRB(State Crime Record Bureau) and DCRB(District Crime Record Bureau)/CCRB (City Crime Record Bureau) have also become more alert to criminal activities in their own jurisdictions. One challenge to law enforcement and intelligence agencies is the difficulty of analyzing large volumes of data involved in criminal and terrorist activities. Data mining holds the promise of making it easy, convenient, and practical to explore very large databases for organizations and users. In this paper, we review data mining techniques applied in the context of law enforcement and intelligence analysis.

2. Data Mining Overview

In this paper we review the Crime Data Mining in two directions

1. Crime Types and security concerns
2. Crime Data Mining Approaches and technique

Crime types and security concerns

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). 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 following are the different types of crimes

- Property crime
- Violent Crime
- Crime against Women and Child
- Traffic Violations

+ Corresponding author. Tel.: + (919942526000).
E-mail address: (malathi.arunachalam@yahoo.com).

- Cyber Crime and
- Others

Crime data mining approaches and techniques

Data mining is defined as the discovery of interesting structure in data, where structure designates patterns, statistical or predictive models of the data, and relationships among parts of the data [1]. Data mining in the framework of crime and intelligence analysis for national security is still a young field.

The following describes our applications of different techniques in crime data mining. *Preprocessing* has been used to keep the data set ready for the process. *Entity extraction* has been used to automatically identify person, address, vehicle, and personal properties from police narrative reports [2]. *Clustering* techniques has been used to cluster the city crime data mining depends on the crimes. *Classification* has been used to detect criminal data from the city crime data base. *Social network analysis* has been used to analyze criminals' roles and associations among entities in a criminal network [9].

3. Data Mining Techniques and concepts

Pre-processing

The data set was made available by the department of Police. The range of years available and utilized was between 2000 and 2009.

Data Attributes

The following yearly attributes were presented and used in the data set for the city crime statistics [7][8]

- Property
  - Murder for gain
  - Dacoity
  - Prep.& Assembly For Dacoity- explanation
  - Robbery
  - Burglary
  - Theft
- Violent
  - Murder
  - Attempt to commit murder
  - C.H.Not Amounting to murder
  - Hurt/Grievous Hurt
  - Riots
- Crime against Women and Child
  - Rape
  - Dowry Death
  - Molestation
  - Sexual Harassment
  - Cruelty by Husband and her relatives.
- Others
  - Kidnapping & Abduction of others
  - Criminal Breach of Trust
  - Arson
  - Cheating
  - Counterfeiting
  - Others IPC crimes

4. Implementations In WEKA Tool

In this paper there are various features that are implemented to support various tasks that can be performed on a given dataset, they are:

Filters: Filters as the word suggest filters out unwanted
data from the dataset. There are two different types of filters that are implemented in this paper, they are:
o **Supervised:** Interface for filters that make use of a class attribute. Ex: AddClasification, ClassOrder, Discretize, Nominal-To-Binary, Resample, SpreadSubSample,

o **Unsupervised:** Interface for filters that do not need a class attribute. Ex: Standardize, String-To-Nominal, String-To-Word-Vector, Replace, ReplaceMissingValues.
Classifiers: Classification techniques are used to predict nominal and numeric attributes. Some of the algorithms that are implemented are: NaiveBayes, Randomizable-Classifier, etc.
Association: This is an important step in crime data mining; this process helps find association between different activities such as crime type, population, year of crime and etc.,. Ex: Apriori, Associator.
Clustering: Cluster models are used to find groups in the dataset with similar instances. Ex: Simple K-means, DBScan, EM.
Data Visualization: This is one of the important parts of data mining, where it helps determine the difficulty of the learning problems
Data fetching is an important step in crime data mining. Data is fetched from different sources, such as a database from CSV formatted files where they are converted to the ARFF file format. A new instance is created in WEKA to store the imported dataset. An instance is an object that stores all the attributes and data objects of the dataset. Functions such as filter, clustering, association, etc are performed on such instances. This can be achieved by using SQL statements to fetch the data based on the dataset attribute. After the datasets are fetched into the application different data mining tasks are performed on them, these methods are explained below.
Filtering
Filtering is an initial step that is performed on the datasets inorder to eliminate attributes that are unnecessary for the data
mining process. This is an important task as some data mining
tasks such as clustering (DBScan, EM) and Association (Apriori) involving large amounts of data takes up large amount of time and memory. Not performing this step will slow down the application, to counter this problem the initial dataset instances which are populated by all the database attributes are filtered down to the most useful attributes depending on the user's interest and requirements.

5. Clustering in Weka

The first task is the prediction of the size of the population of a city [6]. The calculation of per capita crime statistics helps to put crime statistics into proportion. However, some of the records were missing one or more values. Worse yet, half the time, the missing value was the "city population size", which means there was no per capita statistics for the entire record. Over some of the cities did not report any population data for any of their records. To improve the calculation of "yearly average per capita crime rates", and to ensure the detection of all "per capita outliers", it was necessary to fill in the missing values. The basic approach to do this was to cluster population sizes, create classes from the clusters, and then classify records with unknown population sizes [3]. Why use clustering to create classes? Classes from clusters are more likely to represent the actual population size of the cities. The only value needed to cluster population sizes was the population size of each record. These values were clustered using

"weka.clusterers.EM -I 100 -N 10 -M 1.0E-6 -S 100"

6. Prediction of Crime Trends

The next task is the prediction of future crime trends. This meant we tracked crime rate changes from one year to the next and used data mining to project those changes into the future. The basic method here is to cluster the cities having the same crime trend, and then using "next year" cluster information to classify records [11]. This is combined with the state poverty data to create a classifier that will predict future crime trends. Few "delta" attributes were applied to city crime clustering: Murder for gain, Dacoity, Prep.&Assembly For Dacoity, Robbery, Burglary, Theft, Murder, Attempt to commit murder, C.H.Not Amounting to murder, Hurt/Grievous Hurt, Riots, Rape, Dowry Death, Molestation, Sexual Harassment, Kidnapping & Abduction of others, Criminal Breach of Trust, Arson, Cheating, Counterfeiting, and Others IPC crimes. These attributes were clustered using

'Weka 3.5.8's, Simple EM (expectation maximization)'

with parameters of "EM -I 100 -N 4 -M 1.oE-6 -S 100" [4]. EM is a deviation of K-Means clustering. Four clusters were chosen because it produced a good distribution with a relatively easy to interpret set of clusters [5]. Usually, the high level interpretation of clusters from an unsupervised algorithm is not easily defined. However, in this case, the four clusters produced had the following attributes: Note: The clusters are ordered from best to worst.

1) C0: Crime is steady or dropping. The Sexual Harassment rate is the primary crime in flux. There are lower incidences of: Murder for gain, Dacoity, Preparation for Dacoity, rape, Dowry Death and Culpable Homicide.
2) C1: Crime is rising or in flux. Riots, cheating, Counterfeit, and Cruelty by husband and relatives are the primary crime rates changing. There are lower incidences of: murder and kidnapping and abduction of others.
3) C2: 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.
4) C3: 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,

7. City Crime Analysis

Looking at the number of property crimes, it looks like crime at the city has been going down since 2004 except in 2009. But the number of violent crimes has been in flux.

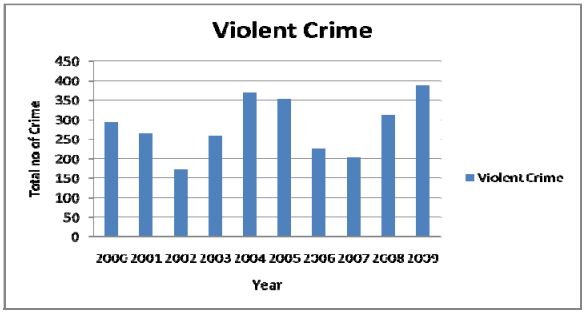

FIGURE 1: Analysis of Violent Crime

Crime against women includes Rape, Dowry Death, Molestation, Sexual Harassment, Cruelty by husband or relatives, Kidnapping & Abduction of women & Girls. Crimes against the women have been going down since 2000, but once again the same thing has been increased since 2004 till 2008, finally it started coming down.

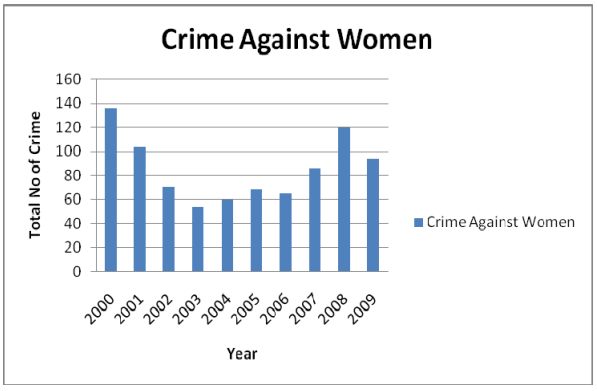

FIGURE 2: Crime against Women – Analysis.

8. Conclusion & Future Work

Data mining applied in the context of law enforcement and intelligence analysis holds the promise of alleviating crime related problem. In this paper we use a clustering/classify based model to anticipate crime trends. The data mining techniques are used to analyze the city crime data from Tamil Nadu Police Department. The results of this data mining could potentially be used to lessen and even prevent crime for the forth coming years. From the encouraging results, we believe that crime data mining has a promising future for increasing the effectiveness and efficiency of criminal and intelligence analysis. Many future directions can be explored in this still young field. Visual and intuitive criminal and intelligence investigation techniques can be developed for crime pattern [10].

9. Acknowledgements

We are highly indebted to Dr. Sylendra Baboo IPS, Commissioner of Police, Coimbatore. We are also thankful to him and his team members for sharing their valuable knowledge in this field

10.REFERENCES

[1] Fayyad, U.M., & Uthurusamy, R. (2002). Evolving data mining into solutions for insights. Communications of the ACM, 45(8), 28-31.

[2] Chau, M., Xu, J., & Chen, H. (2002). Extracting meaningful entities from police narrative reports. In: Proceedings of the National Conference for Digital Government Research (dg.o 2002), Los Angeles, California, USA.

[3] Classification via Decision Trees in weak, Depaul University, http://maya.cs.depaul.edu/~classes/ect584/WEKA/classify.html

[4] Frank Dellaert，The expectation Maximization Algorithm, Technical Report, Georgia Institute of Technology, 2002

[5] Weka 3 - Data Mining with Open Source Machine Learning Software in Java. 7 Dec. 2008 http://www.cs.waikato.ac.nz/ml/weka/

[6] Stephen Schneider, Predicting crime: a review of the research,Department of Justice Canada, 1-2, 2002

[7] Major Crime Trends – Tamil Nadu http://www.tnpolice.gov.in/crimeprofile.html

[8] Major Crime Trends – Tamil Nadu http://www.tnpolice.gov.in/CAWChart.html

[9] David S. Coppock, Why Lift? Data Modeling and Mining, DM review online, 2002

[10] Shyam Varan Nath, Crime Pattern Detection Using Data Mining, Proceedings of the 2006 IEEE/WIC/ACM international conference on Web Intelligence and Intelligent Agent Technology , 41-44, 2006

[11] University Crime Data Mining http://www.cse.msu.edu/~kingstua/Team3/

[12] Manish Gupta et al./ Crime Data Mining for Indian Police Information System, Proceeding of the 2008 Computer Society of India.