Predicting Crime Incidents Systematic Approaches and Crime Analysis for Crime Reduction

Sharmeli J[1], Sethukarsi T[2]

PG Scholar, Department of Computer Science and Engineering, RMK Engineering College, Chennai, India[1]

Professor, Department of Computer Science and Engineering, RMK Engineering College, Chennai, India[2]

ABSTRACT: A crime or criminal offence is an act harmful not only to some individual but also to a community, society or the state. Such acts are forbidden and punishable by law. Crime incidents pose a serious threat to the global economy, safety and well-being of society. Data Analytics helps in discovering patterns and trends, making forecasts, finding relationships and possible explanations, mapping criminal networks and identifying possible suspects. In this we look at the use of frequent pattern analysis at the various crimes done by a criminal and predict the chance of each crime that can again be performed by that criminal. This analysis may help the law enforcement of the country to take a more accurate decision or may help in safeguarding an area. The system provides the opportunity to systematically combine various element or crime characteristics. Additionally, a comprehensive list of crime-related offences is put forward. The offences are characterized in a two-level classification system based on the occurrence of the crime criteria to assist in better classification and correlation of their respective incidents. This enables a thorough understanding of the repeating and underlying criminal activities.

KEYWORDS: Analytics, Criminal offence, mapping, prediction.

I. INTRODUCTION

Crime is a major issue which is concerned as the top priority. The data for crime often presents an interesting dilemma. While some data is kept confidential, some becomes public information. Data about the prisoners can often be viewed in the specified prison department. However, data about crimes related to narcotics or juvenile(Immature) cases is usually more restricted. Similarly, the information about the sex offenders is made public to warn others in the area in order to make the neighbor and the surroundings to be aware of the criminal, but the identity of the victim is often prevented. Thus as the analyst has to deal with all these public versus private data issues so that analysis modeling process does not infringe on these legal boundaries. Law enforcement agencies like that of police today are faced with large volume of data that must be processed and transformed into useful information. Identifying crime characteristics is the first step for developing further analysis. The knowledge that is gained from using analysis approaches is a very useful which can help and support police force[1].

According to the Federal Bureau of Investigation [2], the Internet Complaint Centre received 269 422 complaints of Internet crime in 2014, which indicates a rise of 1600% in comparison to the 16838 complaints[3]. By using this approach, the police forces can use this systematic solution to increase the level offered in the system to predict and prevent the crime related offences. Moreover, this would be useful for police resources allocation. It can help in the distribution of police at most likely crime places for any given time, to grant an efficient usage of police resources. By having all of this information available, we hope to make our society safer for the people living there and also for others who will travel there. Crime analysis is proceeding of analyzing crime. More specifically, crime analysis is the classification of dividing of acts committed in violation of laws into their parts to find out their type based on the incidents, and some data analysis. The role of the crime analysts varies from agency to agency. Statement of these findings is the objective of most crime analysis to find

meaningful information in vast amounts of data and disseminate this information to officers and investigators in the field to assist in their efforts to apprehend criminals and suppress criminal activity.

II. RELATED WORK

Using Machine Learning Algorithms to Analyze Crime Data[7] proposed by Lawrence McClendon and Natarajan Meghanathan in 2015 suggested, Data mining and machine learning have become a vital part of crime detection and prevention. In this research, we use WEKA, an open source data mining software, to conduct a comparative study between the violent crime patterns from the Communities and Crime Normalized Dataset provided by the University of California-Irvine repository and actual crime statistical data for the state of Mississippi that has been provided by neighborhoodscout.com. We implemented the Linear Regression, Additive Regression; and Decision Stump algorithms using the same finite set of features, on the Communities and Crime Dataset. Overall, the linear regression algorithm performed the best among the three selected algorithms. The scope of this project is to prove how effective and accurate the machine learning algorithms used in data mining analysis can be at predicting violent crime patterns.

Crime Prediction Based on Crime Types and Using Spatial and Temporal criminal Hotspots Tahani Almanie, Rsha Mirza and Elizabeth Lor[8] in 2015 suggested on finding spatial and temporal criminal hotspots. It analyses two different real-world crimes datasets for Denver, CO and Los Angeles, CA and provides a comparison between the two datasets through a statistical analysis supported by several graphs. Then, it clarifies how we conducted A priori algorithm to produce interesting frequent patterns for criminal hotspots. In addition, the paper shows how we used Decision Tree classifier and Naïve Bayesian classifier in order to predict potential crime types. To further analyze crimes' datasets, the paper introduces an analysis study by combining our findings of Denver crimes' dataset with its demographics information in order to capture the factors that might affect the safety of neighborhoods. The results of this solution could be used to raise people's awareness regarding the dangerous locations and to help agencies to predict future crimes in a specific location within particular time.

Data mining Techniques to Analyze and Predict Crimes proposed by S.Yamuna and N.Sudha Bhuvaneswari[4] in 2012 suggested 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. About10% of the criminals commits 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 became 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 haven't any kind of information is available before happening of such criminal acts and it result into increasing crime rate. Detecting crime from data analysis can be difficult because daily activities of criminal 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.

III. FEATURE IDENTIFICATION OF CRIME INCIDENTS

The crime issues with providing a comprehensive description are

  1 Adversity in crime descriptions and definitions that focus on different aspects.

  2 The incidents are classified as Crime demonstrate in a significant variety as features and characteristics(e.g., offender, target, and means of attack).

The prediction of future crime trends involves tracking crime rate changes from one year to the next. 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 analysis is made. 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 years records using this years data[4].

The Crime Incidents mechanism of analytics enables a better classification of a relevant incident, exact classification and monitoring of the corresponding criminal offence, and effective action in terms of counter-measures and policy generation.

A. Incident

An Incident is an instance of something happening; an event or occurrence (a violent act or assault). Incident are assigned as a specific elements from a set of already occurring incidents that would limit the capacity to describe sophisticated incidents.

B. Identified Criminal Offence

An Incident is considered as illegal when it has to be classified under an existing criminal offence. By mapping various crime incidents with the corresponding offences the authorities can devise systematic and effective ways of tacking crime. The challenge with the incidents is that they might include aspects of various known offences, they are complex in nature.

1) A thorough classification system of crime-related offences.
2) A set of specific criteria for assigning a specific criminal incident under one or more particular offences.

C. Offender

In relation to the types of offenders, forensic psychology provides descriptive analysis of criminal behavior that can potentially investigate the psychology of the crime offence. According to [5], there is a list of important theories, such as biological theories, addiction, geographical theories, arousal theories, addition, geographical theories, arousal theories and routine activity theory. However, the application of these theories to offenders might be different from face-to-face environments. There has been countless of work done related to crimes. Large datasets have been reviewed, and information such as type and the locations of crimes have been extracted to help people follow law enforcements. Existing methods have used these databases to identify crime hotspots based on locations. There are several maps applications that show the exact crime location along with the crime type for any given city. Even though crime locations have been identified, there is no information available that includes the crime occurrence date and time along with techniques that can accurately predict what crimes will occur in the future. On the other hand, the previous related work and their existing methods mainly identify crime hotspots based on the location of high crime density without considering either the crime type or the crime occurrence date and time. The crime was focusing on the existence of multi-scale complex relationships between both space and time.

IV. CLASSIFICATION SYSTEM FOR CRIME OFFENCES

Many efforts have used automated techniques to analyze different types of crimes, but without a unifying 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 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[6]. 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. SYSTEM DESIGN AND MODULES

The system Architecture of Systematic approach of Crime incident database are given by, the datasets are stored in the database, the required data are transferred in the Hadoop tool and the necessary details are retrieved by using Queries. The representation of the module Systematic process design is represented in the Fig 1. In the Systematic design the Crime database has been stored with Crime datasets and the SQOOP acts as the interface and allows the user to fetch the data between the Hadoop and Database.
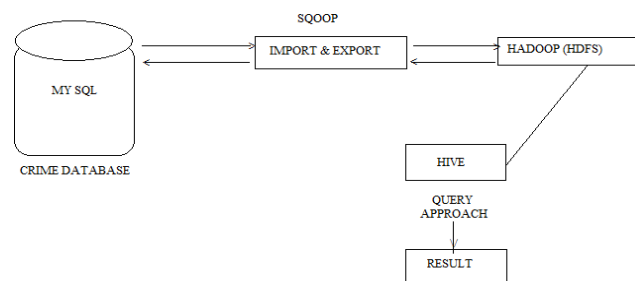

Fig 1 Systematic process design

We have four modules Data pre processing modules, Data migration module with sqoop, Data analytics module with pig, Data analytics module with hive, Data analytic module with Map Reduce. In Data preprocessing module Data set for Crime incidents are created, it contains set of information that contains details of criminal data in particular city.
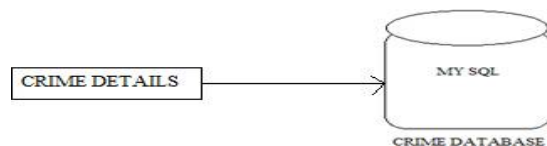

Fig 2 Analyzing Database

This data set is passed in to the database, with the help of the database we can analysis the Incidents as shown in the fig 2.

In the Data Migration module Now the dataset are transferred into Hadoop(HDFS), Sqoop is a command-line interface application for transferring data between relational databases and Hadoop. In this module we fetch the dataset into hadoop (HDFS) using sqoop Tool. Using sqoop we can perform function, like fetching the particular column detail. In the Data analytic module with Hive is a data ware house system for Hadoop. It runs SQL queries called HQL (Hive query language) which gets internally converted to map reduce jobs. Hive was developed by Facebook. Hive supports Data definition Language (DDL), Data Manipulation Language (DML) and user defined functions. In this module analysis of the dataset using HIVE tool can be done which is stored in Hadoop (HDFS). Using hive we can perform Tables creations, joins, Partition, Bucketing concept.
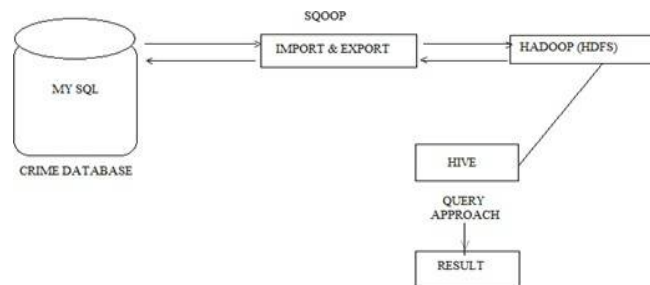
Fig 3 Hive Data analytic process

In Data analytic module with Pig we use Apache Pig is a high level data flow platform for execution, Map Reduce programs of Hadoop. The language for Pig is pig Latin. Pig handles both structure and unstructured language. It is at the top of the Map Reduce process which runs at background. Analyzing the Data set through Pig using Latin Script data flow language. In the Data analytic module with Map Reduce is a processing technique and a program model for distributed computing.
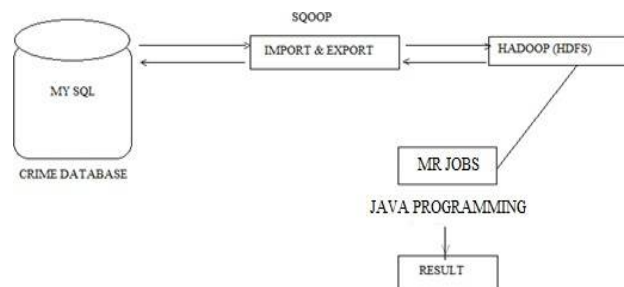

Fig 4 Data Analytic module with MapReduce Algorithm

The MapReduce algorithm contains two important tasks, Map and Reduce. The process is carried with the help of the Java programming.

VI. CONCLUSIONS AND FUTURE WORKS

We looked at the use of data analyzing techniques for identifying crime patterns using the clustering techniques. Our contribution here was to formulate crime pattern detection as machine learning task and to thereby use data analysing techniques to support police detectives in solving crimes. We identified the significant attributes; using expert based semi-supervised learning method and developed the scheme for weighting the significant attributes. Our modelling technique was able to identify the crime patterns from a large number of crimes making the job for crime detectives easier. Some of the limitations of our study includes that crime pattern analysis can only help the detective, not replace them. Also data mining is sensitive to quality of input data that may be inaccurate, have missing information, be data entry error prone etc. Also mapping real data to data analyzing attributes is not always an easy task and often requires skilled data miner and crime data analyst with good domain knowledge. They need to work closely with a detective in the initial phases. As a future extension of this study we will create models for predicting the crime hot-spots [9] that will help in the deployment of police at most likely places of crime for any given window of time, to allow most effective utilization of police resources. We also plan to look into developing social link networks to link criminals, suspects, gangs and study their interrelationships. Additionally the ability to search suspect description in regional, FBI databases [10], to traffic violation
databases from different states etc. to aid the crime pattern detection or more specifically counter terrorism measures will also add value to this crime detection paradigm.

REFERENCES

[1]   Propose Data Mining AR-GA Model to Advance Crime analysis Emad K. Jabar, Soukaena H. Hashem, Enas M. Hessian IOSR Journal of Computer Engineering (IOSR-JCE) e-ISSN: 2278-0661, p- ISSN: 2278-8727 Volume 14, Issue 5 (Sep. - Oct. 2013), PP 38-45

[2]   FBI and NW3C. (May 22, 2015). 2014 Internet Crime Report. Accessed on May 17, 2016.[Online]. Available: https://pdf.ic3.gov2014_IC3Report.pdf "IFCC 2002 Internet fraud report," FBI, Washington, DC, USA, and

[3]   NW3C, Glen Allen, VA, USA, Tech. Rep. NCJ 194344, 2003

[4]   Data mining Techniques to Analyze and Predict Crimes 1S.Yamuna, 2N.Sudha Bhuvaneswari M. Young, The Technical Writer's Handbook. Mill Valley, CA: University Science, 1989.The International Journal of Engineering and Science (IJES)

[5]   G. Kirwan and A. Power, The Psychology of Cyber Crime: Concepts and Principles. Hershey, PA, USA: IGI Glob., 2001.

[6]   Ratcliffe J.H., 2002. Aoristic signatures and spatio-temporal analysis of high volume crime patterns. J. Quantitative Criminology.

[7]   Machine Learning Algorithms to Analyze Crime Data proposed by Lawrence McClendon and Natarajan Meghanathan in 2015 Crime Prediction Based on Crime Types and Using Spatial and Temporal criminal Hotspots Tahani Almanie, Rsha Mirza and Elizabeth

[8]   C McCue, "Using Data Mining to Predict and Prevent   Violent Crimes", Whitepaper, "Oracle's Integration Hub For Justice And Public Safety",

[10]   Oracle Corp. 2004, available at: Oracle.com

[11]   Huygen, N. Helberger, J. Poort, P. Rutten, and N. Van Eijk, "UPS and downs; economic and cultural effects of file sharing on music, film and games," TNO Inf. Commun. Technol. Series, Tech. Rep. 3478 2009.

[12]   F. Oberholzer-Ge e and K. Strumpf, "The effect of file sharing on record sales: An empirical analysis," J. Polit. Econ., vol. 115, no. 1, pp. 1–42, 2007.