IARJSET

ISSN (Online) 2393-8021
ISSN (Print) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology
Vol. 3, Issue 2, February 2016

Crime: Classification and Pattern Prediction

Akshay Kumar Singh[1], Neha Prasad[2], Nohil Narkhede[3], Siddharth Mehta[4]

Department of Computer Engineering, AISSMS COE, Pune, India[1, 2, 3, 4]

Abstract: In the present era of information age, access to a variety of information is easy which has resulted in a rampant increase in crime rate. Criminals are using technology for improving the traditional crime modus operandi. Though law enforcers have knowledge of the misuse of such technology but they are lagging behind in adapting to technology to curb crime. Hence, there is a huge difference in adapting to technology between the two, criminals and law enforcement agencies. In order to reduce the crime rate, there is a need to analyse crime and prevent those crimes. In this paper, we describe a system to analyse crime and how we are going to increase the accuracy of crime prediction to prevent crime. We consider various crime factors to accurately predict crime if a similar crime pattern is observed.

Keywords: Naive Bayes classifier, Apriori Algorithm, Decision Tree, Mongo DB, Web Crawler, Crime Class, Crime Factors.

I. INTRODUCTION

Crime is a behaviour reflected or an act by an individual or a group of individuals that causes a disturbance in social peace. Various crimes are found in our society like a crime against a person, property and authority or an organization. With the advent of the information age, crime has increased rampantly because of easy accessibility of information to criminals. Criminals use the information to commit a crime with more ease. The law enforcers must also include technology to fight crime and reduce the crime rate. In spite of digitization in almost every field, law enforcement agencies still have not incorporated technology to its potential to solve crimes. Nowadays, we have the capacity to collect data and analyse them to find out patterns in order to predict what data we will get in future and at the same time develop methods to use this data in advance for the betterment of the society. In many countries across the globe, law enforcement agencies have started to analyse the crime data collected from past criminal activities and predict future crime hotspots or patterns of crimes to deploy preventive measures to prevent crime. These agencies have various crime data analysts and software's to analyse past and present crime and predict expected crime in near future. In our country India, we still have not deployed the usage of data analysis techniques to analyse crime data on a large scale, although we have collected crime data over the years. The main cause for this delay is we still do not have accurate systems to analyse the collected data and predict the crime pattern or crime hotspot effectively. Though many initiatives have been started over the years in various research centres or high-level crime investigation agencies but not for the local law enforcement agency. So, in order to have crime analysis techniques for local agencies, we have proposed a system with a goal to achieve more accurate prediction of crime pattern and crime category**.**

II. PROPOSED WORK

Our system is an improvisation over another system. We have proposed to analyse crime pattern of a place and predict what category of crime may occur, in near future, in other areas showing similar signs of crime pattern. We are utilizing various crime factors for this purpose and maximize the accuracy of our prediction by incorporating as many crime factors as possible. The various crime factors we are considering are given in Table I-.(Crime factors).

TABLE I CRIME FACTORS

Location	**Time and Date**	**Nature**	**Age**
Education	Object	Category	Agency
Group	Cause	Effect	Type of loss

Our system comprises of four modules to achieve final prediction of crime which are as follows:

A. Collection of data.
B. Classification of data into Crime classes.
C. Pattern Identification.
D. Prediction of crime category.

III. COLLECTION OF DATA

In order to predict crime class by analysis of the crime pattern in a region, we need to collect a variety of crime data. This data is collected by a web crawler from various news feeds, blogs and articles over the Internet. The web crawler will crawl through various articles over the internet and store relevant content after pre-processing as an unstructured data record in MongoDB. Pre-processing is basically done to remove ambiguity, functional words and words whose frequency is not more as compared to other words. These words collected from a document are unstructured data categorized under the various crime factors we are considering. Also, we are going to have a test database to map similarity between words in the document while pre-processing. Data collected in this stage will be used for classification of records into various crime classes.

Data shall be collected and pre-processed and stored in the format as shown in Table II. - (Collected data format expected). The attributes are arranged in a specific order to

IARJSET
ISSN (Online) 2393-8021
ISSN (Print) 2394-1588
International Advanced Research Journal in Science, Engineering and Technology
Vol. 3, Issue 2, February 2016

classify these documents under various crime classes effectively. The table shows an example of how the data will be stored in MongoDB.

TABLE II COLLECTED DATA FORMAT EXPECTED

Doc ID	Attributes From Article After Pre-processing In a Record
1	Pune, Monday morning, Burglary, kids, lock pick, local police, stolen jewellery, cash.
2	Pune, Night New Year's Eve, Chain Snatching, motorcycle, two bikers, multiple victims.
3	Pune, Weekdays, Rush hour, Pick pocketing, trains, male offender, wallets cash.

IV. CLASSIFICATION OF DATA INTO CRIME CLASSES

This is the second major stage after pre-processing the data preceded by data collection stage. We are using a classifier know as Naive Bayes classifier in this stage. It is a supervised learning model which is known for its good performance and accuracy when compared to other classification techniques particularly for document classification as reported by S.L.Ting [4].

It uses a statistical and a probabilistic approach which when given as an input provides a probability distribution for all classes having their posterior calculated probabilities. This classifier classifies the collected documents into a type by assigning each of them a class. The output we get is the P (c|d), the probability of a provided document (d) belonging to a class (p).

The document under the process contains some keywords whose occurrences are calculated within that particular document for finding the posterior probability. So a test set is developed in which there are sets of words having their prior calculated probabilities and occurrences with reference to the document, which helps in classifying the document into the right class(for a new document). So, we are training the model on a test dataset so that it can further classify the upcoming unknown documents according to the highest posterior probability.

Also, we have fixed the zero-frequency problem. While evaluating the overall probability for a document, probability for each keyword belonging to respective classes is calculated. So, if any one of the probability comes out to be zero, the whole probability comes out to zero as it is a multiplication of all the probabilities. In order to soothe this calculation, we replace the zero by one (+1) thus giving accurate classification. For eg: P (B) * P (C/D) * P (E/D) * P (F/D) and if P (E/D) is zero then we replace its value by one (+1) to achieve uniform distribution.

Since we are gathering data from news feeds we have access to local crimes only and not high priority national crimes so once the data is gathered we are going to classify that data records in the database into various Crime Classes. As shown in Table III.-(Crime classes) the various crime classes considered by our system.

Table IV- (Crime classification expected) shows how the examples in Table II.-(Collected data format expected) are classified into what crime classes.

TABLE III CRIME CLASSES

Pick pocketing	Burglary	Chain Snatching	Fraud
Break-Ins	Property damage	Environment damage	Assault

TABLE IV CRIME CLASSIFICATION EXPECTED

Doc ID	Crime Category Class
1	Burglary
2	Chain Snatching
3	Pick pocketing

V. PATTERN IDENTIFICATION

The Apriori property is based on the following, if an itemset I does not satisfy the minimum support threshold, min sup, that is, P(I) < min sup then the itemset I is not frequent. If an item A is added to the itemset I, then the resulting itemset (i.e., I [A]) cannot occur more frequently than I. Therefore, I [A] is not frequent either, that is, P(I [A]) < min sup.

Apriori property states that every nonempty subset of a frequent itemset must also be frequent. Apriori algorithm is used to calculate itemsets that are most frequent. These generated frequent item sets will be used to identify crime hotspots and gain insight into crime pattern.

Apriori algorithm uses Association Rules for Crime Analysis. We have to identify trends and patterns in crime. For finding crime pattern that occurs frequently, we are using Apriori algorithm. Apriori algorithm is used to determine association rules which generate trends in the database. Apriori algorithm helps in getting better insights in the database.

Various attributes which can be considered while generating frequent itemsets are location-related attributes like the place, time, day, type of crime committed (like burglary, murder etc.), the age of victim and offender, the presence of VIP. These attributes can be matched with current VIP events to predict crime.

Now once we have our crime class for the data collected we are going to find a general pattern for each crime class based on the classified records. This pattern will help us to predict future crime if similar pattern detected. The pattern generally represents similar attributes or crime factors.

For example, the pattern for the three crimes as described above give in Table V-. (Expected pattern for crime).

TABLE V EXPECTED PATTERN FOR CRIME

Burglary	House alarm, lock pick, cash, jewellery, break in, empty houses
Chain Snatching	Gold, women, roads, malls, bikers, teenagers, crowded locality
Pick-pocketing	Wallets, cash, personal belonging, Trains, Buses, Crowded, Male

DOI 10.17148/IARJSET.2016.3210

VI. PREDICTION OF CRIME CATEGORY

We are using decision trees for prediction. Decision tree has internal nodes which signify the presence of an attribute and following the path the node takes represents the next sequence of attribute occurrence. The decision tree has an edge over neural networks in the sense that it can be easily understood and interpreted. Unlike decision trees, neural networks have somewhat limited utility in many of the necessary public safety functions because they are relatively opaque. It is not possible to just look at a neural net and understand the nature of the associations, which significantly limits their applicability in certain tasks. Therefore, in many situations, it is important to compromise somewhat on accuracy in an effort to identify an actionable model that can be used in the operational setting [2]. The other advantages include its robust nature and also it works well with large data sets. This feature facilitates the algorithm to make better decisions regarding variables [3].

Decision trees capitalize on the fact that criminal behaviour is relatively foreseeable or homogeneous, particularly regarding common or successful MO features [2]. Decision tree concept is being used to characterize and model known patterns of behaviour. It is applied to new data in an effort to quickly identify previously observed, known patterns and categorize unknown behaviour [2].
The goal is to develop a group of decision rules that can be used to determine a known outcome. [2] Decision trees work by repeatedly subdividing the data into groups based on identified predictor variables, which are related to the selected group membership. In other words, this technique creates a sequence of decision rules that is used to separate data into specific, predetermined groups. [2]

Prediction of crime class possible if similar pattern detected as determined by the Pattern Identification Module i.e. similar attributes found for an area where crime occurrence highly possible.

For example, if in an area we witness crime factors like house alarm quite often needed to repair, people living in are mostly out on vacations, not a good neighbourhood, rich people living and distance from police station quite large. Analysing these attributes the system shall determine that Burglary is most likely to occur and if a burglary has occurred in past we can determine a pattern and time frame, where and when the burglary can occur.

VII. EXPECTED SYSTEM RESULT

Our system gives the result in the form of crime class or category of crime by which effective measure to avoid crime can be deployed by the local law enforcement agencies to secure the neighborhood. Also, we can predict the time frame for the crimes if the collected data is quite rich.

As proposed, we are also initially creating a test database for simplifying the classification and prediction. This database will consider various details regarding previous crimes of all the crime classes under consideration. Data shall include the details about the crime with respect to the various crime factors. Using this test database, we are going to train our system so that in future it can accurately determine the category of crime and also successfully generate the most likely crime pattern for a particular area.

Since we are collecting data for our system from the news feeds and newspaper articles over the internet, the severity of crime is low but mostly such crimes occur frequently and by preventing or reducing such crime rate we can have a crime reduced society and the local law enforcement agencies can effectively deploy their resource to protect the society from such crimes and enforce preventive measures in various areas.

Once our system proves to be successful for local law enforcement agencies in curbing the crime rate we can deploy it on a large scale for various types of crimes and not limited to the crime category we have selected.

REFERENCES

[1] Isuru Jayaweera, Chamath Sajeewa, Sampath Liyanage, Tharindu Wijewardane, Indika Perera and Adeesha Wijayashri, “Crime Analytics: Analysis of Crimes through Newspaper Articles” IEEE ISBN 9781479917402, 2015
[2] Colleen McCue, “Data Mining and Predictive Analysis Intelligence Gathering and Crime Analysis” ISBN 9780080464626.
[3] Lior Rokach and Oded Maimon. Decision trees. In Oded Mairnon and Lior Rokach, editors, the “Data Mining and Knowledge Discovery Handbook”, pages 165-192. Springer, 2005.
[4] S.L. Ting, W.H. Ip, Albert H.C. Tsang, “Is Naive Bayes a Good Classifier for Document Classification?” International Journal of Software Engineering and Its Applications, Vol. 5, No. 3, July 2011.
[5] Malathi. A and Dr. S. Santhosh Baboo. Article: an enhanced algorithm to predict a future crime using data mining. International Journal of Computer Applications, 21(1):1-6, May 2011. Published by Foundation of Computer Science.
[6] Hsinchun Chen, Wingyan Chung, Jennifer Jie Xu, Gang Wang Yi Qin and Michael Chau ,”Crime Data Mining: A General Framework and Some Examples” IEEE Computer Society April 2004 ISBN 0018916204.
[7] GU Yunhua, SHEN Shu, ZHENG Guansheng,” International Journal of Digital Content Technology and its Applications.” Volume 5, Number 6, June 2011.
[8] Bruno Laporais Pereira and Wladmir Cardoso Brandão, “ARCA: MINING CRIME PATTERNS USING ASSOCIATION RULES”, 11th International Conference Applied Computing 2014.
[9] P. Chamikara, D. Yapa, R. Kodituwakku and J. Gunathilake, “SL Secure Net : intelligent policing using data mining techniques,” International Journal of Soft Computing and Engineering, vol. 2, no. 1, pp. 175-180, 2012.
[10] R. Krishnamurthy and S. Kumar, “Survey of data mining techniques on crime data analysis,” International Journal of Data Mining Techniques and Applications, vol. 1, no. 2, pp. 117-120, December 2012.
[11] V. Nath, “Crime pattern detection using data mining,” in Web Intelligence and Intelligent Agent Technology Workshops, Hong Kong, 2006, pp. 41-44.