

Prediction of Crime Pattern and Suspects Using Datamining Techniques

¹Aishwarya DS, Madhumalathi S, Manisha UA, Sushma K M, Ravikumar V G*

Department of CSE, GSSSIETW, Mysuru, Karnataka, India

DOI: <https://doi.org/10.21467/proceedings.1.28>

* Corresponding author email: ravikumarvg@gsss.edu.in

Abstract

Crime prevention and detection has become important in today's society. However, the crime data previously stored from various sources have a tendency to increase steadily. As a consequence, the management and analysis with huge data are very difficult and complex. To solve this problem, data mining techniques employ many learning algorithms to extract hidden knowledge from huge volume of data. This project involves in the prediction of crime patterns and suspects based on the previously collected data thus helping the law enforcers to prevent and reduce the rate of crimes and also help in solving crimes quickly.

Index Terms- crime pattern prediction, suspect prediction, data mining, association mining, Apriori algorithm, classification rule C4.5.

1 INTRODUCTION

Data mining is the phenomenon of finding patterns from large data sets. It is an essential process where intelligent methods are applied to extract data patterns. Data mining techniques can be used to analyze the historical data to find trends and patterns related to crime. To extract the hidden knowledge, these are the initial important factors:

- 1) The data used for analysis requires accuracy and efficiency.
- 2) Knowledge and experiences of specialists.
- 3) The knowledge results obtained from data mining.

This project makes use of previously stored data set and with the help of data mining techniques it helps to find patterns and trends in crimes. It can help solve the crimes with more speed and also help to predict the possible suspects involved in a particular crime. This system is a web enabled application It is an automation for complaints registration, crime pattern prediction and suspect prediction based on the previous crime details This system makes use of data mining technique called as "Apriori algorithm" for crime patten predictionand "classifier C4.5" to predict suspects. The Input given to the system is previous crime activities, date and location. The expected output is crime patterns and list of suspects.



2 LITERATURE SURVEY

“Crime Prediction and Forecasting in Tamilnadu using Clustering Approaches Article Structure”-By S Sivaranjini and Dr. S Sivakumari. This paper gives a systematic way of detecting and investigating patterns and trends in crime. This work uses a various clustering approaches of data mining to analyze the crime data of Tamilnadu. The crime data is extracted from National Crime Records Bureau (NCRB) of India. It consists of crime information about six cities namely Chennai, Coimbatore, Salem, Madurai, Thirunelveli and Thiruchirapalli from the year 2000-2014 with 1760 instances and 9 attributes to represent the instances. K-Means clustering, Agglomerative clustering and Density Based Spatial Clustering with Noise (DBSCAN) algorithms are used to cluster crime activities based on some predefined cases and the results of these clustering are compared to find the best suitable clustering algorithm for crime detection. The result of K-Means clustering algorithm is visualized using Google Map for interactive and easy understanding. This work helps the law enforcement agencies to predict and detect crimes in Tamilnadu with improved accuracy and thus reduces the crime rate. “Mining crime data by using new similarity measure”-By Guangzhu Yu, Shihuang Shao, Bing Luo. This paper introduces the Clustering technique to the field of crime data analysis for finding crime suspects but compared with the high requirements for public security work, traditional clustering methods used in existing application systems provide a relatively low accuracy. To solve the problem, paper proposes a new similarity measure method, i.e., Segmented Multiple Metric Similarity Measure (SMMSM), to improve the accuracy of similarity measure. In this method, attributes are divided into different groups according to their importance to similarity, compensation relationships do not exist among attributes in different groups. The new measure is scalable with dimensionality of data and is both suitable for numeric data and categorical data. Experiments on real data show that our method has higher accuracy than other measures. “Crime Analysis using k-means Clustering”-By Anant Joshi, A. Sai Sabitha, Tanupriya Choudhury. This paper focuses on the Analysis of crime for providing safety and security to the civilian population. Using data mining, we can discover critical information which can help local authorities detect crime and areas of importance. The main purpose of this paper is to analyze the crime which entails theft, homicide and various drug offences which also include suspicious activities, noise complaints and burglar alarm by using qualitative and quantitative approach. Using K-means clustering data mining approach on a crime dataset from New South Wales region of Australia, crime rates of each type of crimes and cities with high crime rates have been found. There are many social, temporal, spatial and demographic factors which assist the police in the evaluation of crime. A lot of data is collected when a crime investigation is finished and this data can be used to find patterns. It can allow us to find geographical locations where a particular type of crime happens frequently and allow police agencies to respond better.

3 METHODOLOGY

3.1 The Algorithm used to predict crime patterns is Association rule- Apriori algorithm

The steps are as follows:

STEP 1: Scan the Crime data set and determine the support(s) of each item.

STEP 2: Generate L1 (Frequent one item set).

STEP 3: Use Lk-1, join Lk-1 to generate the set of candidate k - item set.

STEP 4: Scan the candidate k item set and generate the support of each candidate k – item set.

STEP 5: Add to frequent item set, until C=Null Set.

STEP 6: For each item in the frequent item set (L) generate all non-empty subsets.

STEP 7: For each non empty subset determine the confidence. If confidence is greater than or equal to this specified confidence. Then add to Strong Association Rule.

Sample Example

City Name - Mysuru

The crime types considered are theft, snatching, murder, dacoit and burglary.

Table 1: Crime Type.

TYPE ID	CRIME TYPES
1	Theft, Murder, Dacoit.
2	Theft, Murder, Burglary.
3	Theft, Snatching, Murder, Burglary.
4	Snatching, Burglary.

The minimum support and the minimum confidence are defined as follows:

Minimum Support = 50%

Minimum Confidence = 80%

Item set consists of Theft, Snatching, Murder, dacoity and Burglary as crime types. The frequent item set obtained after calculating the support is given by the below table.

Table 2: Frequent Item Set (L)

ITEMS	SUPPORT
Theft	75%
Snatching	50%
Murder	75%
Dacoit	75%
Theft, Murder	75%
Theft, Burglary	50%
Snatching, Burglary	50%
Murder, Burglary	50%
Theft,Murder,Burglary	50%

Based on the frequent item set the crime types which have the minimum support are considered for generating the confidence table.

Table 3: Generate Confidence

RULE X	RULE Y	CONFIDENCE
{theft}	-> {murder}	100%
{murder}	-> {theft}	100%
{theft}	-> {burglary}	66%
{burglary}	-> {theft}	66%
{snatching}	-> {burglary}	100%
{burglary}	-> {snatching}	66%
{burglary}	-> {murder}	66%
{theft}	-> {murder, burglary}	66%
{murder}	-> {theft, burglary}	66%
{burglary}	-> {theft, murder}	66%
{murder, burglary}	-> {theft}	100%
{theft, burglary}	-> {murder}	100%
{theft, murder}	-> {burglary}	66%

After applying the strong association rule the final pattern of crime can be generated which is represented below.

Crime pattern

1. {Snatching} -> {burglary}
2. {Murder, burglary} -> {theft}
3. Theft, burglary} -> {murder}
4. {Theft} -> {murder}
5. {Murder} -> {theft}

This is the final output of one module- i.e.: pattern prediction module.

3.2 The algorithm used to predict suspects is Classification rule C4.5

The steps are as follows:

Step 1: Scan the dataset (storage servers)

Step 2: for each attribute a, calculate the gain [number of occurrences]

Step 3: Let a_best be the attribute of highest gain [highest count]

Step 4: Create a decision node based on a_best – retrieval of nodes [patient] where the attribute values matches with a_best.

Step 5: recur on the sub-lists [list of patient] and calculate the count of outcomes [Stages] – termed as sub nodes. Based on the highest count we classify the new node.

Sample Example-Crime Type - Robbery

Attributes (Clues) – vehicle, vehiclecolor, height [m=3]

Subject (Suspects) – Anoop, Sujay [p=1/2=0.5]

Table 4: Training Dataset

Name	Vehicle (X,Y,Z)	Vehicle color (A,B,C)	Height (P,Q,R)	Culprit
Anil	X	A	P	Anoop
Kumar	X	B	Q	Anoop
Ajay	Y	B	P	Sujay
Naveen	Z	A	R	Anoop
Akash	Z	A	Q	Sujay

New Complaint Features – Akul vehicle -X, vehicle color-A, height-R. Anoop/sujay?

Feature Count (X) in the dataset = 2

Feature Count (A) in the dataset = 3

Feature Count (R) in the dataset = 1

Table 5. Descending order of count.

FEATURE	COUNT
A	3
X	2
R	1

A – Anoop (2) & Sujay (1)

Table 6. Output- The list of probable suspects.

SUSPECTS	PRIORITY
Anoop	2
Sujay	1

4 CONCLUSIONS

Crime related data is sensitive and has a large domain and therefore there is a need of some efficient clustering techniques and algorithms which will help the crime analysts and law enforcers retrieve the data and information and draw patterns and conclude to a result which will help in the investigation process. This project makes use of previously collected crime data and in conjunction with a priori and c4.5 algorithms it predicts the crime patterns and suspects respectively. The algorithms can be developed in such a way that solves the unsolved crimes faster and also predicts the patterns in a more precise and accurate way.

5 REFERENCES

- [1] Sunil Yadav, Rohith Vishwakarma, “Crime Pattern Detection, Analysis And Prediction”, 2017
- [2] NELSON Bilorian, MARIO Ferrandez, “Crime Prediction Using Pattern And Context”, 2017
- [3] Maria Jeseca C. Baulo, CHARLI S. Mazan, “Geospatial-Temporal Analysis And Classification Of Criminal Data In Manila”, 2017
- [4] Mohammad Al Boni And Matthew S. Gerber, “Predicting Crime With Routine Activity Patterns Inferred From Social Media”, 2016

-
- [5] Nurul Hazwani Mohd Shamusuddin, NOR Azizah Ali, "An Overview On Crime Prediction Methods", 2017
 - [6] S.SIVARANJANI, DR.SIVAKUMARI, "Crime Prediction And Forecasting In Tamilnadu Using Clustering Approaches", 2016
 - [7] GUANGZHU Yu, SHIHUANG Sho And Bing Lua, "Mining Crime Data By Using New Similarity Measure", 2008
 - [8] Chhaya Chachan And Smritisehgal, "A Review: CRIME Analysis Using Data Mining Techniques And Algorithms", 2017
 - [9] HSINCHUN Chen And Wingyan Chung, "Crime Data Mining: A General Framework And Some Examples", 2004
 - [10] SANDHYA Harikumar And Divya Usha, "Apriori Algorithm For Association Rule Mining High Dimensional Data", 2016
 - [11] SHASHI Kanth Kumar And Amrithpal Kaur, "Constraint Data Mining Using Apriori Algorithm With AND Operation", 2016
 - [12] WANG Renli, DAI Yusming Dong Liming, "The Application Of Apriori-BSD Algorithms In Medical Records Data Mining", 2016
 - [13] P. Thongtae And S.Srisuk, "An Analysis Of Data Mining Application In Crime Domain", 2008
 - [14] SHIJU Sathya Devan And Devan M.S, "Crime Analysis And Prediction Using Data Mining", 2014
 - [15] PRIYANKA Dhaka And Rahul Johari, "CRIB: Cybercrime Investigation, Data Archival And Analysis Using Bih Data Tool", 2016
 - [16] ZHANG Ying, "Analysis Of Crime Factors Correction Based On Data Mining Technology", 2016
 - [17] LOUISE F.G Underson, "Using Data Mining And Judgment Analysis To Construct A Predictive Model Of Crime", 2002
 - [18] ANANT Joshi And A.Sai Sabitha, "Crime Analysis Using K-Means Clustering", 2017
 - [19] Gaurav Govind Kumar, Santhoshkumar P, "A Survey On Crime Data Analysis Using Data Mining Techniques", 2017
 - [20] Shyam Varan Nath, "Crime Pattern Detection Using Data Mining", 2006