Implementation of Data Mining for Effective Crime Investigation: A Case study of Maharashtra Police

Prof. Harinath N. Ramesh - Prof. Ashish S. Desai
Prof. Prasanna R. Rasal - Prof. Rajendra D. Kumbhar

ABSTRACT:

The research study highlights the present status of crime investigation and need of data mining implementation for crime investigation support to the police force. Data Mining plays very important role in analysis of large amount of data to find various hidden patterns. Crime investigation has very significant role in functioning of police system in any country. Nowadays, due to the extensive technological advancements, the people with criminal intentions can easily commit criminal acts. It has become a major challenge for police system to detect and prevent such crimes, as there is no any kind of information is available before happening of such criminal acts. The present paper highlights the conceptual framework of various data mining tools which can be used for crime investigation.

Keywords:
Data Mining, Crime, NCRB, Investigation, CCIS, CIA, technological advancements, Classification, clustering, association rule mining, Data Warehouse.

I. INTRODUCTION:

Police plays an important role in civil administration in India. The Constitution of India assigns responsibilities for the maintenance of law and order to the country. Police force is in insufficient ratio as compare to population. It leads to increasing graph of crime in nation, another reason is lack of use of information technology in investigation and in prevention. Criminal use very sophisticated technology and finds new way to commit crime. Keeping these things in mind govt. of India take a step towards and designed CCIS model. In 1986 Govt. of India introduced National Crime Record Bureau (NCRB) Under SCRR. The Crime Record Bureau (CRB) for state and District crime record bureau (DCRB) for districts has been created in order to make use of information technology. The Government of India designed Crime Criminal Information System (CCIS) to store and retrieve crime and criminal records. To provide the input to CCIS, the crime integrated police application systems is designed. In desire of technological innovations in database technologies such as data warehousing and data mining, the present CCIS is only limited to the informative purposes. The large database of CCIS can be used for analysis purpose which can help to our police system to uncover various hidden patterns of criminal acts which helps to detect and prevent such crimes.

II. MAHARASHTRA POLICE: AN OVERVIEW.

To understand current scenario of crime detection, we need to know structure and hierarchy, technological usage of police force.

The Head of state police is Director General of Police (DGP). The state is divided into administrative units called as Districts. A group of districts called as region and Head for each region is Deputy Inspector General of Police (DIGP). Superintendent of Police (SP) is Head for district and is assisted by Additional Superintendent
of Police (Addl. SP) and Deputy Superintendent of Police (DySP) in each district.

Maharashtra, the third largest State in Republic of India, has one of the largest Police departments in the country. Besides 233 Indian Police Service officers holding the State rank, it consists of 171 Superintendents of Police, 290 Deputy Superintendents of Police, 3136 Inspectors, 2641 Assistant Police Inspectors, 3001 Sub Inspectors and 1,48,912 more members of a mandatory Bilingual Maharashtra, a highly industrialised State with dense urban conglomerates, has adopted Commission rates system for policing its large cities. The State has 10 Commissionerates and 35 district police units. Details about these units as well as special units of Maharashtra Police Department are available under sub-head “Districts & Commissionrates and “Special Units of MPD” on the menu bar of the home page [13].

III. DATA WAREHOUSING AND DATA MINING TECHNIQUES:

A data warehouse is a subject-oriented, integrated, time-variant and non-volatile collection of data in support of management’s decision making process. (W.H. Inmon-1990)

So, data warehouse can be said to be a semantically consistent data store that serves as a physical implementation of a decision support data model and stores the information on which an enterprise needs to make strategic decisions.

Need of data warehousing:

Data warehousing developed, despite the presence of operational databases due to following reasons:

1) An operational database is designed and tuned from known tasks and workloads, such as indexing using primary keys, searching for particular records and optimizing Processing Of AP queries in operational databases would substantially degrade the performance of operational tasks.

2) An operational database supports the concurrent processing of multiple transactions. Concurrency control and recovery mechanisms and OLAP query often need read-only access of data records for summarization and aggregation. Concurrency control and recovery mechanisms, if applied for such OLAP operations, may jeopardize the execution of concurrent transactions.

3) Decision support systems require data that are not typically stored in operational databases. These data are typically obtained from other systems using Extract, Transform and Load (ETL) procedures.

4) Decision support needs coordination of data from heterogeneous sources, and requirements, which are continuously updated too.

Nowadays, the implementation of data warehousing is required, which consists of following activities:

- Define the architecture, do capacity planning and select the storage servers, database and OLAP server and tools
- Integrate the servers, storage and client tools
- Design the warehouse schema and views
- Define the physical warehouse - conversion, data placement and partitioning and access methods
- Connect the sources using gateways, ODBC drivers or other wrappers
- Design and implement scripts for data extraction, cleaning, transformation, load and refresh
- Populate the repository with the schema and view definitions, scripts, data footprint, etc.
- Design and implement end-user applications
- Roll out the warehouse and applications

DATA MINING:

Data mining is basically used to find out unknown patterns from a large amount of data. There are several tools available for this purpose, including Microsoft Excel. There are two approaches to the implementation of data mining. One is to copy data from one database into another and mine it. Other approach is to mine the data within a data warehouse. There are various data mining techniques [12] available, as follows:

A. Classification:

Classification is used to classify datasets into a certain number of predefined classes or certain criteria. The data with similar common properties are classified into predefined classes.

B. Clustering and segmentation:

This technique is used to segment a database into subsets, or clusters, based on set of attributes. It is a method to group data into classes with similar characteristics in which the similarity of attributes...
IV. DATA MINING PROCESS IN CRIME INVESTIGATION

The SCRB, Maharashtra collects crime data from various CRBs using CCIS. The data stored in CCIS at SCRB, Maharashtra is used as a Data Warehouse of Criminal data in Maharashtra. The process of collecting crime data from various CRBs is normally done in CCIS along with its various details such as Crime Register Date, Type and Section, Complainant Name, Major Head, District, Police Station, Brief Facts, Accused Name, Offence Place, Crime Register Status, etc.

A data preprocessing is a process of data gathering, which is collected from CCIS. Preprocessing removes noise, inconsistencies and specification of boundaries. Text mining is more important technique used in this process. Text Mining is the discovery by computer of new, previously unknown information, by automatically extracting information from different written resources. A key element is the linking together of the extracted information together to form new facts or new hypotheses to be explored further by more conventional means of experimentation. A related application is automatic detection of fraud, such as in credit card usage. Analysis look across huge numbers of crime case details to find uncommon fraud patterns. (1) Due to recent advancements in technology, there are multiple different kinds of fraud such as identity theft, electronic fraud, and more.

Data mining is a part of the overall procedure in crime data analysis and detection. After preprocessing of crime data, the next step is knowledge discovery.

After preprocessing, the data mining process would be started. The data mining processes include classification, clustering, association, data mining and pattern analysis. These are used for discovering crime data patterns. For a data mining process, one should have a predefined task set for making a model. For example, the criminal case under investigation is interested in identifying criminal, crime-related information, that can be used to understand crime patterns.

The figure 1 shows the crime Data Mining process to knowledge discovery.

Figure 1: Crime Data Mining Process
attributes such as age, type of weapon they had used, location, time etc. that could be beneficial to trap suspect and accordingly investigation can be done. These patterns can be displayed in the form of charts or charts.

The data mining is therefore a must choice that can be used to support a Criminal Detection and investigation. It reduces data analysis time consuming and enhances a frontier of crime analysis. The knowledge from crime analysis can be available by which the data mining is used as a core process in data analysis.

V. CONCLUSIONS:


Figure 2. Number of cases for investigation and Investigation Completion proportion

Incidence of cognizable crimes (IPC) under different crime heads during 2000-2007 (Table 1.1)

There is huge gap between number cases on board and completion of investigation due to many reasons which are stated below.

- Integrated Mechanism for investigation. The Common Integrated Crime Data Mining Cell (CICDMC) must be formed to help the investigation officer on requirement.

- Technology Usage: Police must use the intelligence technology for investigation. As on today they are mostly investigate with traditional way, on contrary criminals are using very sophisticated technology and often finds the loop hole, due to which, there is tremendous increase in crime ratio.

- It is crucial to change the procedure of investigation and make it more efficient.

- It is necessary to put some importance on the training of the police officers.

- It is necessary to provide the investigators necessary training.

- The police force should be well equipped with proper infrastructure and proper working material and equipment to be increased.

- It is necessary to maintain inter-sectional department.

- Most of the officers, Constables, PSI, work 16 to 18 hours, and it affects on their behaviour, working strategy, attitude etc. ultimately it affects on investigation, so to reduce the working hour it is necessary to recruit additional force.

ACKNOWLEDGMENT:

The researchers are grateful to the authors, writers and editors of the books and articles, which have been referred for preparing the presented research paper.

It is the duty of the researchers to remember their parents whose blessings are always with them.

REFERENCES:


http://herbamic.org/index.html 24 December 2009 2.20 pm

http://indianpolice.gov.in/article/24 December 2009 4.50 pm

http://mahapolice.gov.in disputetree village 24 December 2009 12.30 pm

Alex Berson, Stephan J. Smith (2004) Data Warehousing, Data Mining, & OLAP, FIA McGRAW HILL Publications New Delhi

Mark Hadley (2003).

http://people.ischool.berkeley.edu/~herrf/Ktextmining.html 26 December 2009 5.10 pm

### Table 1:2

Incidence of reportable crimes (IPC) under different crime heads during 2016-2017

<table>
<thead>
<tr>
<th>Sr. No.</th>
<th>Year</th>
<th>Crime Group</th>
<th>Crime Category</th>
<th>Area</th>
<th>Rural</th>
<th>Urban</th>
<th>Rural Deaths</th>
<th>Urban Deaths</th>
<th>Rural Incidents</th>
<th>Urban Incidents</th>
<th>Rural Clearance</th>
<th>Urban Clearance</th>
<th>Total Clearance</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2016</td>
<td></td>
<td></td>
<td>1</td>
<td>1639</td>
<td>1639</td>
<td>2424</td>
<td>4428</td>
<td>1639</td>
<td>1639</td>
<td>1639</td>
<td>1639</td>
<td>1639</td>
</tr>
<tr>
<td>2</td>
<td>2017</td>
<td></td>
<td></td>
<td>2</td>
<td>2022</td>
<td>2022</td>
<td>2802</td>
<td>5462</td>
<td>2022</td>
<td>2022</td>
<td>2022</td>
<td>2022</td>
<td>2022</td>
</tr>
<tr>
<td>3</td>
<td>2018</td>
<td></td>
<td></td>
<td>3</td>
<td>2213</td>
<td>2213</td>
<td>2982</td>
<td>5679</td>
<td>2213</td>
<td>2213</td>
<td>2213</td>
<td>2213</td>
<td>2213</td>
</tr>
<tr>
<td>4</td>
<td>2019</td>
<td></td>
<td></td>
<td>4</td>
<td>2153</td>
<td>2153</td>
<td>2982</td>
<td>5679</td>
<td>2153</td>
<td>2153</td>
<td>2153</td>
<td>2153</td>
<td>2153</td>
</tr>
<tr>
<td>5</td>
<td>2020</td>
<td></td>
<td></td>
<td>5</td>
<td>2074</td>
<td>2074</td>
<td>2982</td>
<td>5679</td>
<td>2074</td>
<td>2074</td>
<td>2074</td>
<td>2074</td>
<td>2074</td>
</tr>
<tr>
<td>6</td>
<td>2021</td>
<td></td>
<td></td>
<td>6</td>
<td>2067</td>
<td>2067</td>
<td>2982</td>
<td>5679</td>
<td>2067</td>
<td>2067</td>
<td>2067</td>
<td>2067</td>
<td>2067</td>
</tr>
<tr>
<td>7</td>
<td>2022</td>
<td></td>
<td></td>
<td>7</td>
<td>2074</td>
<td>2074</td>
<td>2982</td>
<td>5679</td>
<td>2074</td>
<td>2074</td>
<td>2074</td>
<td>2074</td>
<td>2074</td>
</tr>
</tbody>
</table>