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

Prof. Hanmant N. Renushe[1], Prof. Abhijit S Desai[2],
Prof. Prasanna R. Rasal[3], Prof. Rajendra D. Kumbhar[4]

[1][2][3][4] Bharati Vidyapeeth Deemed University Institute of Management and Entrepreneurship Development, Sangli (Maharashtra)
Karmaveer Bhaurao Patil Institute of Management Studies & Research, Satara (M.S.), India

[1]hanmant.renushe@bharatividyapeeth.edu, [2]abhijit.desai@bharatividyapeeth.edu
prasanti.rs@gmail.com, [4]rdkumbhar@gmail.com

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 presented paper highlights the conceptual framework of various data mining tools which can be used for crime investigation.

Keywords:

Data Mining, Crime, NCRB, Investigation, CCIS, CIFA, 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 responsibility law and order in the country. Police force is in insufficient ratio as compare to population[1], it leads to increasing graph of crime in nation, another reason is lack of use of information technology in investigation and in prevention. Criminal uses very sophisticated technology and finds new way to commit crime. Keeping these things in mind govt. of India take a step forward and designed CCTNS model. In 1986 Govt. of India created National Crime Record Bureau(NCRB)[2] Under NCRB the state crime record bureau [SCRB] for state and District crime Record Bureau [DCRB] for districts has been created In order to making 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 common integrated police application was also designed. In despite 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 be helpful to our police system to uncover various hidden patterns in 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 a 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 forces in the country. Besides 236 Indian Police Service officers borne on the State cadre, it consists of 121 Superintendents of Police, 230 Deputy Superintendents of Police, 3136 Inspectors, 2641 Assistant Police Inspectors, 8001 Sub Inspectors and 1,48,912 men members of constabulary. Maharashtra, a highly industrialized State with five urban conglomerates, has adopted Commissionerates 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 & Commissionerates and "Special Units of MPD" on the menu bar of the home page.[3]

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 OLAP 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 needs 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) Operational databases maintain only current data whereas decision support requires historical data which operational databases do not typically maintain.

4) Decision support needs consolidation of data from heterogeneous sources, and its quality often requires comparison of old and new data.

Steps involved in designing of a data warehouse

Designing a data warehouse is a complex process which consists of following activities:

- Define the architecture, do capacity planning and select the storage servers, database and OLAP servers and tools.
- Integrate the servers, storage and client tools.
- Design the warehouse schema and views.
- Define the physical warehouse organization, 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 schemas and view definitions, scripts and metadata.
- 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 popular tools of data mining to rule discovery, classification. There are two approaches to the implementation of data mining. First is to copy data from data warehouse and mine it. Other approach is to mine the data within a data warehouse. There are various data mining techniques [5] available as follows:

A. Classification :

Classification is used to classify database records on number of predefined classes on certain criteria. The data with sharing common properties are specified 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 identical characteristics in which the similarity of intra-class is

C. Association:

D. Decision Tree :

A decision tree can predict the customer can be viewed as tree, each branch is a classification question and leaves of the tree are partitions of data set with their classification. It divides data on each branch point without losing any of the data. The number of churners and nonchurners is conserved as we move up or down the tree. ID 3, C4.5, CART and CHAID are some algorithms used in this technique.

E. Neural Networks :

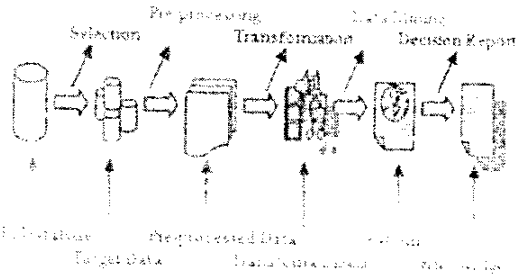
True neural networks are biological systems that detect patterns, make predictions and learn. The artificial neural networks are computer programs implementing sophisticated pattern detection and machine learning algorithms on a computer to build predictive models for historical databases.

IV. DATA MINING PROCESS IN CRIME INVESTIGATION

The SCRB, Maharashtra collects crime data from various CRB's using CCIS. The data stored in CCIS at SCRB, Maharashtra is works as a Data Warehouse of Criminal data in Maharashtra. The process of collecting crime data from each DCRB is normally uses CR no. along with its various details such as Crime Register Date, Acts 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. Analysts look across huge numbers of

Figure 1: Crime Data Mining Procedure



Crime & riminal Information System stores number of crime under different headings such as Murder, Attempt to Commit Murder C.H. Not Amounting Murder, Rape Kidnapping & Abduction , Dacoity , Preparation & Assemble For Dacoity , Robbery, Riots Arson, Dowry Deaths, Kidnapping & Abduction Of Women & Girls, Molestation, Sexual Harassment, Cruelty By Husband And Relatives, Importation Of Girls, Criminal Breach Of Trust, Cheating, Counterfeiting, Burglary, Theft etc. so related data should be taken for pre-processing for example the crimes related to Murder or Robbery then the noisy data or inconsistent data must removed, after removing inconsistent data the data is mined using data mining tool and patterns are generated, these patterns may shows inter-relation between crimes using year

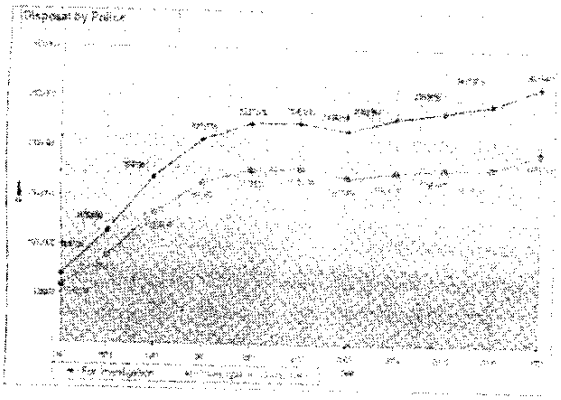
attributes such as age type of weapon they had used, location , time etc. that could be beneficial to trap suspect and accordingly investigation can be done. These patterns can be displayed in the form of charts or chart.

The data mining is therefore a most choice that can be used to support a Criminal Detection and investigation. It reduces data analysis time consuming and enlarges 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 :

The Current situation of IPC Crime for investigation and crime investigation completed [7] represented in figure 2.

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 necessary to change the procedure of and appointment of the IPS officers.
- Training to police is not adequate [7] & [8] Specific training to police is necessary.
- Training must be provided to the investigating personnel on regular basis.
- The police force should be well equipped with powerful weapons and proper shielding material and strength has to be increased.
- It is necessary to encourage Physical sport and games.
- Most of the offices, 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 forces.

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://ritawriter.wordpress.com/2007/08/20/police-people-to-police-ratio/, 23 December 2009 4:30 pm.

http://ncrb.nic.in/origin.html , 24 December 2009 2:20 pm

http://statistics.gov.in/ abouttis 23 December 2009 4:50 pm

http://mahapolice.gov.in/ disputefreevillage 24 December 2009 12:30 pm

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

Marti Hearst(2003), http://people.ischool.berkeley.edu/~hearst/text-mining.html 26 December 2009 3:10 pm

National Crime Record Buerau , CCISMLe http://ncrb.nic.in/cii2007/cii-2007/CII_CHARTS07.pdf

Table 1.1

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

Sl. No.	Year	Murder	Attempt to Commit Murder	Culpable Homicide not amounting to Murder	Rape	Kidnapping & Abduction	Dacoity	Preparation and assembly for Dacoity	Robbery	Burglary (House Breaking)	Theft	Riots
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
1	2000	37949	36743	6073	16496	22891	5525	1226	25406	165191	336668	93464
2	2001	36222	35123	5967	16075	23046	5604	1618	12913	163138	329662	78326
3	2002	35230	34390	5524	16373	24330	6105	1251	11754	158671	341542	67590
4	2003	32716	35342	5029	15847	19992	5203	1393	11512	153724	329307	57515
5	2004	33658	37860	5470	18233	21757	5111	1340	11602	155197	329537	57154
6	2005	32715	36181	5618	18359	19705	5161	1054	11003	151166	321611	56245
7	2006	32467	37406	5535	19348	21991	4761	1021	10496	151266	319644	56441
8	2007	32318	37851	5621	20737	24551	4574	1021	10136	151268	327042	58615

Sl. No.	Year	Criminal Breach of Trust	Cheating	Counterfeiting	Arson	Hurt	Dowry Deaths	Molestation	Sexual Harassment	Cruelty by Husband or Relatives	Importation Of Girls	Causing Death by negligence	Other IPC Crimes	Total Cognizable crimes under IPC
(1)	(2)	(14)	(15)	(16)	(17)	(18)	(19)	(20)	(21)	(22)	(23)	(24)	(25)	(26)
1	2000	14581	41701	2291	16382	241550	6995	9746	11046	45778	64	17050	727067	1771084
2	2001	14798	44727	1985	16534	271487	6851	34124	9781	48170	114	17152	701362	1769408
3	2002	14527	46271	1502	17620	269025	6822	33943	10455	45237	76	16434	709747	1768179
4	2003	13432	47478	2055	18365	261444	6208	32839	10326	50703	46	16572	750412	1716329
5	2004	14178	51959	1529	19637	276848	7026	34567	10201	52121	59	16423	741033	1832010
6	2005	13572	53605	2363	18451	270861	6787	34178	9961	58319	149	21698	741577	1822602
7	2006	13635	58076	2189	18480	264748	7618	36617	9665	63128	57	78513	775697	1878293
8	2007	15531	65395	2204	19024	273067	8093	38724	10490	75810	61	66278	825036	1989673