

AN INTELLIGENT ANALYSIS OF CRIME DATA USING DATA MINING & AUTO CORRELATION MODELS

Uttam Mande

Dept of CSE
GITAM University
Visakhapatnam

Y.Srinivas

Dept of IT
GITAM University
Visakhapatnam

J.V.R.Murthy

Dept of CSE
J.N.T.University
Kakinada

Abstract

The latest technological developments contributed significantly towards modernization, at the same time increased the concern about the security issues. These technologies have hindered the effective analysis about the criminals. Application of data mining concepts proved to yield better results in this direction. In this paper, binary clustering and classification techniques have been used to analyze the criminal data. The crime data considered in this paper is from Andhra Pradesh police department this paper aims to potentially identify a criminal based on the witness/clue at the crime spot an auto correlation model is further used to ratify the criminal.

Key words: data mining, clustering, classification, autocorrelation, crime

1. INTRODUCTION

The present day, changes in social life style and circumstances of living make the humans to come across phenomena called crime. Various agencies such as POLICE Department, CBI are working rigorously to combat the crime. But the challenges to analyze the crime and arrest the criminal activities is becoming more difficult as the crime rate is increasing [1][2], many models have been projected by the researchers for effective analysis [3][4]. the main disadvantage is that the volume of data with respect to the crime activities and criminals increased, and there is a great need for analyzing the data, hence to have a better model the knowledge about the crime & the criminal always is always advantageous. This thought has driven towards the use of data mining techniques [5] for analyzing this voluminous data .

The usage of data mining concept help to explore the enormous data and making it possible in reaching the ultimate goal of criminal analysis the usage of data mining techniques have several advantages it helps to cluster the data based on criminal /crime and thereby minimizing the search space. Based on the clusters the classification algorithm can be applied to classify the criminal in this paper we also used the auto correlation model authenticate the criminal. The rest of paper is organized as follows, section -2 of the paper deals with deep insight into fundamentals' of crime analysis, in section- 3 the concept of binary clustering is presented, the auto correlation model is discussed in section -4, experimentation is highlighted in section- 5, the section 6 of the paper focus on the conclusion

2. INSIGHT INTO FUNDAMENTALS' OF CRIME ANALYSIS

Any crime investigation highlights primarily on two issues, 1) clue/crime links 2) criminal relating/identification. Crime clues play a vital role in the proper identification of criminal. The clues help the stepping stone towards the crime analysis, and criminal relating is the mapping of the criminal based on the clues with data available in the data base, by the use of intelligent knowledge mapping.

In this paper, we have considered the crime data base of criminals involved/accused in several types of crimes the criminal activities considered in this paper are 1) robbery 2) murder 3) kidnapping 4) riots

2.1 crime links

The various crime links that were considered include

- 1) Crime location (place: restaurant, theater, road, railway station, shop/gold shop, mall, house, apartment)
- 2) Criminal attribute(hair, built, eyebrows ,nose ,teeth, beard , age group,mustache,languages known)
- 3) Criminal psychological behavior can be recognized by type of killing

We have considered the type of killing as (smooth, removal of parts, harsh) which attributes to the psychological behavior of the criminal

- 4) Modus operandi (object used for crime),
1)Pistol 2)Rope 3)Stick 4)Knife

These criminal links help to analyze the dataset there by making the crime investigators to plane for identification of the criminal.

2.2 Crime identification

In order to identify the criminals the variables/links that are identified from section 2.1 are mapped to that of the data base and previous knowledge there by solving the crime incidents,

In order to identify a criminal together with crime variables that are discussed in section 2.1 we have considered 1) witness available and 2) clues available

The various witness considered for affective identification are discussed in the above section if the evidence/witness is not available we have considered clues available from the forensics such as finger prints ,in particular cases for the identification we have considered the mapping of both witness & clues to authenticate a criminal.

In this paper we have considered binary clustering to cluster the data base based on type of crime and the classification is carried out from the feature available.

3. BINARY CLUSTERING:

In order to simplify the analysis process the huge dataset available is to be clustered. The clustering in this paper is based on the type of crime. A data set is generated from the database available from the Andhra Pradesh police department and a table is created by considering the FIR report

The various fields considered including the criminal identification numbers, criminal attributes, criminal psychological behavior, crime location, time of crime (day/night), witness /clue, the data set is generated by using the binary data of 1's & 0's, 1's indicating the presence of attribute and 0's indicating the absence of attribute then clustering of the binary data is done as proposed by Tao Hi (2005) using the binary clustering

Crimes are categorized in many ways, here we have given weights to each type of crime where weighing scheme is considered in the manner all the relative crimes will be given with near values , after applying clustering algorithm on this type of crime feature we have got four clusters of crime data they are robbery, kidnap, murder and riot

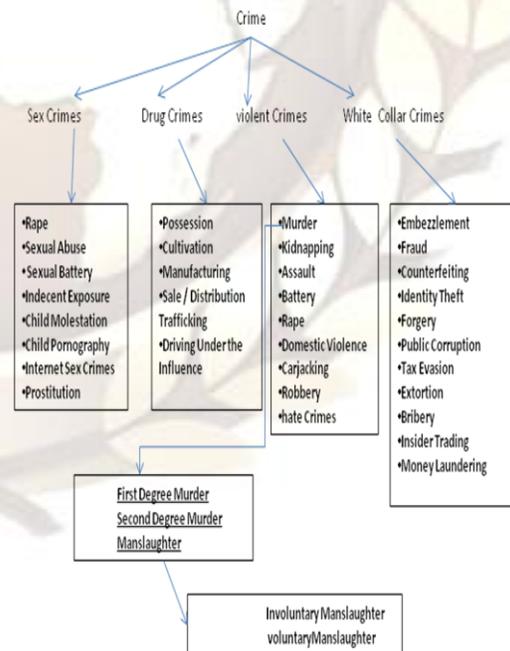


Fig 1 categories of crimes

4. AUTO REGRESSION

In the absence of witness and clues by the forensic at the criminal spot. In such situations, reverse investigation has to be done

$$C_k = \frac{1}{n} \sum_{t=k+1}^n (y_t - \bar{y})(y_{t-k} - \bar{y})$$

- Suspect are listed out by

The type of crime , Type of killing / the way in which the incident has happened.

Modus of operand – used , By the victim type ,Time of happening of the incident

$$r_k = \frac{\sum_{t=k+1}^n (y_t - \bar{y})(y_{t-k} - \bar{y})}{\sum_{t=1}^n (y_t - \bar{y})^2}$$

- If the incident is with respect to killing , then we consider the reason for killing

Dacoit and murder ,Rape and murder ,Killing took place at the time of riot

Formulae for auto correlation and regression

Crime committed due to heatedness ,No clue

- Modus of operand of crime

Knife, Pistol, Bomb,rope

- Type of killing

Harsh,Sadistic,From near / from distance

- Time of killing

Day time,Night time

- Victim type

Age, Economical status, Gender, Strong ness of victim

- Using these feature like type of kill ,associated crime mode of the crime, victim type the suspects are generated from the criminal history

Among the suspects, to identify a criminal, the correlation is to be established among the available evidences and here we use the auto correlation model to find the most likelihood criminal

The database of short listed criminals(suspects) will be given to Autocorrelation model

5. EXPERIMENTATION

The experimentation is carried out under mat lab environment .the generated database considered for experimentation is

Criminalid	Type of Crime	Mode of Crime					Place				Shop	Mal
		Knife	Pistol	Stick	Rope	Road	Rest aurant	RlyS tation	Theatre			
101	1	1	0	0	0	0	1	0	0	0	0	1
102	2	0	1	0	0	0	1	0	0	0	0	1
103	1	0	1	0	0	0	0	1	0	0	0	0
104	2	0	1	0	0	0	1	0	0	0	0	1
105	2	1	0	0	0	0	0	1	0	0	0	0
106	3	1	0	0	0	0	0	1	0	0	0	1
107	2	0	1	0	0	0	1	0	0	0	0	0
108	1	0	1	0	0	0	1	0	0	0	1	0
109	2	1	0	0	0	1	0	0	0	0	0	1
110	4	0	0	1	0	1	0	0	0	0	0	1
111	1	0	1	0	0	0	1	0	0	0	1	0
112	1	1	0	0	0	1	0	0	0	0	1	0
113	1	1	0	0	0	1	0	0	0	0	0	1
114	3	1	0	0	0	1	0	0	0	0	0	0
115	2	0	1	0	0	0	0	0	1	0	0	0
116	3	1	0	0	0	1	0	0	0	1	0	0
117	4	1	1	1	0	1	1	0	1	1	1	0
118	1	1	0	0	0	1	0	0	0	0	0	1
119	2	0	1	0	0	0	1	0	0	0	0	1
120	3	0	0	0	1	1	0	0	0	0	0	1
121	1	1	0	0	0	0	0	0	1	0	0	0
122	1	0	0	0	1	0	1	0	0	0	0	1
123	2	1	0	0	0	0	0	0	1	0	0	1

Fig2 :the snap shot of data set

If the witness is available, at the crime incident, or of the forensics reports are available, then in such cases, identification of the criminal is a different case, where the criminal is mapped with the data available at the crime spot with that of the database and if there is a map, the criminal can be identified. If the witness or forensics reports are not available, then we will take the report on the way the crime has been taken and we try to relate these features with the Autocorrelation model and try to investigate the criminal. The criminal

In the result highest positive value is considered as the most likelihood criminal

cid	correlation values
101	0.750249895876718
118	0.767249345567653
142	0.743567565585454
154	0.713427738442316

From the above table, it can be seen the Correlation value for the criminal 101 is maximum and that implies that he is having more likelihood of committing the crime. This analysis report is used by the investigating officer for further analysis

6. CONCLUSION:

This paper presents a novel methodology of identifying a criminal, in the absence of witness or any clue by the forensic experts. In these situations, in this paper we have tried to identify the criminal by mapping the criminal using the method of Auto correlation. The way in which the incident has taken place and the features of the crime are considered to ratify a criminal.

REFERENCES:

1. Carlile of Berriew Q.C “Data mining: The new weapon in the war on terrorism” retrived from the Internet on 28-02-2011
2. Cate H. Fred “Legal Standards for Data Mining” retrieved from the internet on 12-03-2011 http://www.hunton.com/files/tbl_s47Details/FileUpload265/1250/Cate_Fourth_Amendment.pdf
3. Clifton Christopher (2011). “Encyclopedia Britannica: data mining”, Retrieved from the web on 20-01-2011
4. Jeff and Harper, Jim “Effective Counterterrorism and the Limited Role of Predictive Data Mining” retrieved from the web 12-02-2011
5. U.M. Fayyad and R. Uthurusamy, “Evolving Data Mining into Solutions for Insights,” Comm. ACM, Aug. 2002, pp. 28-31.
6. W. Chang et al., “An International Perspective on Fighting Cybercrime,” Proc. 1st NSF/NIJ Symp. Intelligence and Security Informatics, LNCS 2665, Springer-Verlag, 2003, pp. 379-384.
7. H. Kargupta, K. Liu, and J. Ryan, “Privacy-Sensitive Distributed Data Mining from Multi-Party Data,” Proc. 1st NSF/NIJ Symp. Intelligence and Security Informatics, LNCS 2665, Springer-Verlag, 2003, pp. 336-342. April 2004
8. M.Chau, J.J. Xu, and H. Chen, “Extracting Meaningful Entities from Police Narrative Reports, Proc. Nat’l Conf. Digital Government Research, Digital Government Research Center, 2002, pp. 271-275.
9. A. Gray, P. Sallis, and S. MacDonell, “Software Forensics: Extending Authorship Analysis Techniqueto Computer Programs,” Proc. 3rd Biannual Conf. Int’l Assoc. Forensic Linguistics, Int’l Assoc. Forensic Linguistics, 1997, pp. 1-8.
10. R.V. Hauck et al., “Using Coplink to Analyze Criminal-Justice Data,” Computer, Mar. 2002, pp. 30-37.
11. T. Senator et al., “The FinCEN Artificial Intelligence System: Identifying Potential Money Laundering from Reports of Large Cash Transactions,” AI Magazine, vol.16, no. 4, 1995, pp. 21-39.
12. W. Lee, S.J. Stolfo, and W. Mok, “A Data Mining Framework for Building Intrusion detection

Models,” Proc. 1999 IEEE Symp. Security and Privacy, IEEE CS Press, 1999, pp. 120-132. 10. O. de Vel et al., “Mining E-Mail Content for Author Identification Forensics,” SIGMOD Record, vol. 30, no. 4, 2001, pp. 55-64.

13. G. Wang, H. Chen, and H. Atabakhsh, “Automatically Detecting Deceptive Criminal Identities,” Comm. ACM, Mar. 2004, pp. 70-76.

14. S. Wasserman and K. Faust, Social Network Analysis: Methods and Applications, Cambridge Univ. Press, 1994.

