International Journal of Modern Engineering Research (IJMER)
www.ijmer.com Vol.2, Issue.4, July-Aug. 2012 pp-1507-1510 ISSN: 2249-6645

Witness Based Criminal Identification Using Data mining Techniques and New Gaussian Mixture Model

Uttam Mande[1], Y. Srinivas[2], J. V. R[3]. Murthy
[1](Dept of CSE, GITAM University Visakhapatnam)
[2](Dept of IT, GITAM University Visakhapatnam)
[3](Dept of CSE, J. N. T. University, Kakinada)

Abstract: *The crime history in the recent past has increased drastically, and the tendency towards the crime is also an upheaval task. Therefore to combat the crime, effective technologies are to be provided as supportive to the law enforces. Data mining concepts prove to be very useful in this direction. A major uplift in the direction of crime identification is direct evidence, which can be of crucial concern by which plausible conclusions can be drawn. Identifying the criminal by means of forensics reports is seemingly difficult. This paper introduces a novel methodology of criminal mapping based on forensics using the concepts of Data mining and New Gaussian Mixture model.*

Key words*: Data mining, New Gaussian Mixture model, Crime, criminal mapping, criminal identification, forensics.*

I. Introduction

The quest towards the luxurious life styles, environmental conditions, and uncontrollable ambitions have driven the humans towards the criminal activities[1] Many government and private agencies are rigorously marching to combat the law breakers. Today, most of the criminal records are maintained manually and on the other hand due to the lack of sufficient man power, the data entry has become a concern [2]. With this background, as the criminal activities have extended its wings at different locations, it has become an challenging task for the police to exchange the information between the police stations at different locations in solving the issues. Therefore formulation of effective mechanism to support the law enforces while investigations are very much needed.[3] This mechanism should be helpful in information sharing and thereby shaping towards a meaningful crime analysis. To support, effective methodologies that help in criminal investigations for evaluating the physical clues and determining the different strategies to be adopted for investigation are to be formulated. Therefore the physical evidences obtained from the witness and the forensics reports help towards the better criminal analysis.

Hence in this paper, a database is generated from the criminal data available from different police stations of Andhra Pradesh. The database is created by considering two basic facts viz., availability of witness at the criminal sport and the reports from the forensic labs that are collected from the clue spot[4]. The data mining concepts are exploited for mining the likelihood of the criminals based on the features/ witness available. Clustering is performing based on the type of crime. The crime activities considered in this paper are Murder, Riot, kidnap and robbery. Based on the features described about the criminal, a face is generated and the generated face is compared with that of the existing faces for finding the likelihood of the criminals[6]. These identities are further mapped with the forensic clues to formulate as unique identity. Data mining techniques help to explore the enormous data and making it possible in reaching the ultimate goal of criminal analysis by using the concepts of clustering and classification. In this paper the concept of clustering is carried out basing on the type of crime. The rest of paper is organized as follows, section -2 of the paper deals with Acquisition of features, in section- 3, details about the crime identification presented, Clustering techniques is discussed ,in section -4 the New Gaussian Mixture model is presented , experimentation is highlighted in section- 5, the section 6 of the paper focus on the conclusion

II. Acquisition of Features from the witness

Any crime investigation highlights primarily on two issues, 1) Witness available and 2) Clues available and relating these features with that of the data available regarding the criminals. The crime data base considered in this paper include are 1) robbery 2) murder 3) kidnapping 4) riots. For the identification of any crime we need to have an idea about

1)clue variables 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.

III. Crime Identification using 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.

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.

3.1 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 1categories of crimes

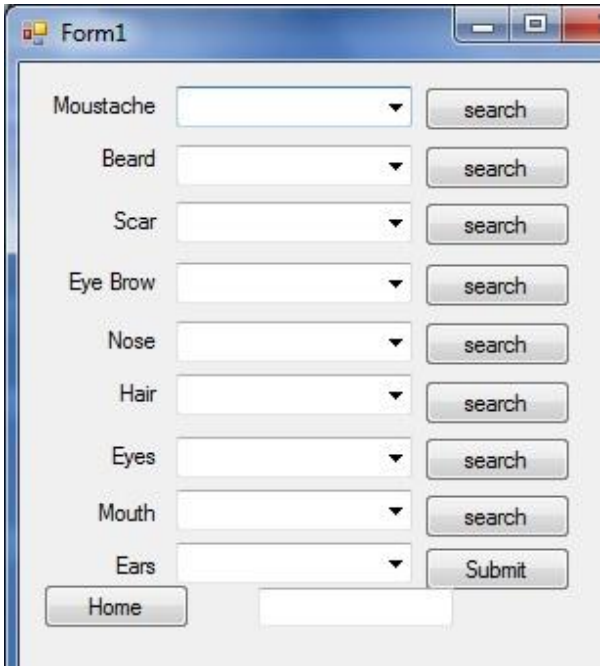
IV. New Gaussian Mixture model

New Gaussian Mixture model is utilized for the purpose of most likelihood criminal matching. This model is an extension of the existing Gaussian Mixture Model .The main advantage of using this model is that it minimizes the noise data as it includes a filter in the distribution itself. This is called a New Gaussian Mixture model since it tries to maintain the characteristics of GMM and it builds a covariance matrix for dimensionality reduction The probability density function of New Gaussian Mixture Model is given by PDF

$$f(x) = \frac{1}{\sqrt{2\pi\sigma}}\left\{\frac{\mu}{N}\frac{e^{-(x-\mu)^{\mu+1}}}{2\sigma^2} + \frac{\mu}{N}\sum_{i=1}^{n}\left(\frac{(x_i-\mu)^{\mu+1}}{2\sigma^2}\right)\right.$$

V. Experimentation

The features are taken as input from the witness available and basing on the features, the database is compared for the relevancy in crime .The features are taken as input and basing on the features, The database is compared for the relevancy in crime .The features that are considered are the standard features generally used in FIR at the police stations and is shown in the following figure-1



The various images that are retrieved using the features are displayed as follows

The outputs obtained includes cids and their filtered data is stored in the data base further investigation

filter | CAT

ID	cid	featurenam
116	1142	Feature
117	1149	Feature
118	1152	Feature
119	1155	Feature
120	1159	Feature
121	1006	Feature
27	1006	Feature
28	1006	Feature
33	1003	Feature
34	1005	Feature
35	1006	Feature
36	1009	Feature
37	1031	Feature
38	1036	Feature
39	1039	Feature
40	1042	Feature
41	1046	Feature
42	1047	Feature
43	1050	Feature
44	1053	Feature
45	1057	Feature
46	1058	Feature
47	1060	Feature

Probability density functions of the test image is mapped with the most likelihood of images present in the database and the most relevant images obtained are displayed

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 considered in this paper. The criminal is mapped by collecting the features about the crime from the witness and comparing them with that of the available from the data base. and if there is a map, the criminal can be identified. .If the data available from the witness is not sufficient, the forensics reports are also considered that are available, and correlate this report with the report of the witness to ratify criminal. Using the methodology a criminal is identified and for the uniqueness, this data is given as input to the New Gaussian mixture model to identify a unique criminal. From the above table, 101 is identified as the criminals

VI. Conclusion

This paper presents a novel methodology of identifying a criminal, in the presence 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 New Gaussian mixture model. .

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 Techniquesto 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.