FEATURE SPECIFIC CRIMINAL MAPPING USING DATA MINING TECHNIQUES AND GENERALIZED GAUSSIEN MIXTURE MODEL

Uttam Mande
Dept of CSE
GITAM University
Visakhapatnam
mineuttam@gmail.com

Y. Srinivas
Dept of IT
GITAM University
Visakhapatnam
sriteja.y@gmail.com

J.V.R. Murthy
Dept of CSE
J.N.T. University
Kakinada
mjonnalagedda@gmail.com

Abstract:
Lot of research is projected to map the criminal with that of crime and it is observed that there is still a huge increase in the crime rate due to the gap between the optimal usage of technologies and investigation. This has given scope for the development of new methodologies in the area of crime investigation using the techniques based on data mining, image processing, forensic, and social mining. In this paper, presents a model using new methodology for mapping the criminal with the crime. This model clusters the criminal data basing on the type crime. When a crime occurs, based on the eye witness specified features, the criminal is mapped. Here we propose a novel methodology that uses Generalized Gaussian Mixture Model to map the features specified by the eyewitness with that of the features of the criminal who have committed the same type of the crime, if the criminal is not mapped, the suspect table is checked and the reports are generated.

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

1. Introduction
With the quest towards luxurious life, people to fulfill their ambitions are getting themselves indulged towards the crime and criminal activities. The national crime record bureau states that, there are nearly 230000 crimes which occurred upto 2010. The Governments is working rigorously towards the steps to be taken to decrease the crime rate, using
1) Prevention of the crime occurrence
2) Finding the accused and punishing

1.1 Prevention of the crime occurrence(before crime occurrence)
It always better to prevent the occurrence crime instead of taking the measures after occurrence of the crime. For this, the analysis of crime data plays a vital role in prediction of the occurrence of the future crime. Much work has been reported in this areas and it is mostly confined in predicting the crime occurrence based on the location, which helps for to analyze the locations chosen by the criminals. Analyzing the data of crime, finding the frequent locations they are using the for particular type of crime, time of crime, and social status of the people, crime links etc. will be of more use. This helps the law keepers to predict the crime and take the nessacary action.

1.2 Finding accused and punishing(after the occurrence of the crime)
Punishment creates a fear among the criminals and thereby acts as a preventive measure to stop the others criminals to get attracted towards the criminal activities; this directly implies that if a criminal is identified and punished rigorously for his criminal activities, indirectly leads towards the crime prevention. As per the latest statistics, it can be observed that more than 70% per cases are not solved and pending which implies increase in the rate of the crime, and as per the survey reports, the crime activities are mostly repeated by the old criminals who have committed similar crimes earlier, so it proves that if the old crimes are solved, indirectly helps to prevent the new crimes. With this gap in mind, in this paper, we have developed an model which helps the police department and other investigation agencies to find the criminal at the earliest. This model is confined to situations where the eye witness is there at the place of crime. for our model we have collected a data which is generated from the criminal data available from different police stations. 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[uttam et al. 2012]. The data mining concepts are exploited for mining the likelihood of the criminals based on the features/ witness available. Based on the type of the crime the clustering is performed. Here the weight are assigned to the each crimes and the they are clustered using the k-means clustering. We have got mainly four clusters. 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 [Uttam et al. 2012]. 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 earlier work The rest of paper is organized as follows, section 2 of the paper describes about the feature extraction, section 3 will give the details about the crime identification presented, Clustering techniques is discussed, in section 4 the generalized Gaussian Mixture model is presented, experimentation is highlighted in section 5, the section 6 of the paper focus on the conclusion.

2. Features Extraction:
The way of doing crime investigation is depended on different factor like, 1) type of crime. 2) Witness available and 3) Clues available etc.. Using these factors the crimes can be mapped with criminal who are having the same pattern. In the crime database we have to consider the features which are available at the crime spot. For the identification of any crime we must depend on the different crime variables like 1) Clue variables 2) Criminal relating/identification variables. 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.

3. Crime variables for criminal mapping

The important crime variables used for crime mapping include

1) Type of crime
   a. (murder, rape, riot, kidnap)
2) Way of doing the crime killing
   a. Criminal psychological behavior can be recognized here
      like for murder (harsh, smooth, removing parts)
3) Time of crime
   a. (time, day time, night)
4) Modus operandi
   a. (object used for crime), 1) Pistol 2) Rope 3) Stick 4) Knife
5) Crime location
   a. (place: restaurant, theater, road, railway station, shop/gold shop, mall, house, apartment)
6) Physical attributes of criminal
   a. (hair, built, eyebrows, nose, teeth, beard, age group, mustache, languages known)

These criminal variables help to analyze the dataset there by making the crime investigators to plane for identification of the criminal. Chose of the variables is depended on the particular scenario of the crime occurred. In this paper we have considered the k- means clustering to cluster the data base based on type of crime and the classification is carried out from the feature available.

3. k-means 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 indexing the each feature with index value as shown in figure 3.1

![Fig 3.1](image-url)
Depending on the different offences and the victims the crime are categorized as shown in the Fig 3.2. For the different crimes different sections are given .hence when ever crime occurs until the investigation completes, law enforces can not define the crime to exact category ,this implies that the at early stages the crime is defined in generalized way

![Fig 3.2 categories of crimes](image)

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 .as shown in figure 3.3

When a crime occurs the data to be analyzed is taken from any one dataset of the cluster in which this crime comes. Here the features specified by the eye witness are considered and they are given to mapping model which calculates the PDF’s of features using Generalized Gaussian mixture model and compare with PDF’S of the criminals features .

4.1 Generalized Gaussian Mixture Model (GGMM)

Generalized Gaussian Distribution

In this section we briefly discuss the probability distribution and its properties used in the crime identification algorithm. Let the crime data values intensities in the entire crime data obtained by crime data values intensities is a Random Variable and follow a Finite Generalized Gaussian Mixture Distribution. It is also assumed that the entire crime data is a collection of ‘K’ crime data regions, then the crime data values intensities in each crime data region follows a Generalized Gaussian Distribution. The probability density function is

\[
f(z|\mu, \sigma, P) = \frac{1}{2\Gamma(1+\frac{1}{P})A(P,\sigma)} e^{-\left(1-\frac{1}{P}\right)\left(\frac{z-\mu}{\sigma}\right)^{p\sigma}}
\]

\[
\text{Where} \quad \sigma > 0, A(P,\sigma) = \left[\frac{\sigma^2}{\Gamma(1+\frac{1}{P})}\right]^{\frac{1}{P}}.
\]

The parameter \(\mu\) is the mean, the function \(A(P,\sigma)\) is an scaling factor which allows that the \(\text{Var}(Z) = \sigma^2\), and ‘P’ is the shape parameter. When \(P=1\), the corresponding Generalized Gaussian corresponds to a Laplacian or Exponential Distribution. When \(P=2\), the corresponding Generalized Gaussian corresponds to a Gaussian distribution. In limiting cases \(P \to +\infty\), equation (4.1) converges to a uniform distribution in \((\mu - \sqrt{3}\sigma, \mu + \sqrt{3}\sigma)\) and when \(P \to 0+\), the distribution becomes a degenerate one in \(Z = \mu\).

5. 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 that are considered are the standard features generally used in FIR at the police stations and is shown in the following figure 5.1
The features which are given as input are given for the calculation of PDF using the Generalized Gaussian Mixture model, these PDF’s are compared with PDF’s of the criminals features in the data base and the criminals with nearer values are displayed as shown in the fig 5.2. The outputs obtained includes cids and their filtered data is stored in the data base further investigation here the metrics like the MSE and PSNR are used for the evaluation of the outputs as shown in the fig 5.3.

Here the as shown in the table the values of MSE and PSNR are analyzed and the criminal with highest PSNR is considered to be most likelihood criminal as shoen in the fig 5.5 the criminal with cid 1013 is considered as most likelihood criminal with highest PSNR.

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 Generalized Gaussian mixture model to identify a unique criminal. From the above table, 1013 is identified as the criminals

6. 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 Generalized Gaussian mixture model.

References:

1. Carlile of Berriew Q.C “Data mining: The new weapon in the war on terrorism” retrieved 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


4. Jeff and Harper, Jim “Effective Counterterrorism and the Limited Role of Predictive Data Mining” retrieved from the web 12-02-2011


14. S. Wasserman and K. Faust, Social Network Analysis: Methods and Applications, Cambridge Univ. Pr is available as a Word file, <copyright.doc>.