Bayesian Prediction In Police-Crime Investigation Management System

J.Sethuraman [#1], K.R.Sekar [*2], N.Senthil Selvan [#3]

[#] Computer Science Department, Sastra University
Thanjavur, India
[1] sethuraman@cse.sastra.edu
[3] senthilselvan@cse.sastra.edu
[*] Sastra University
Thanjavur, India
[2] sekar_kr@cse.sastra.edu

***Abstract*— The paper titled as "Bayesian prediction in Police-crime investigation management system" is a web based project that provides an online portal for ordinary citizens to file complaints on crimes, view their status, and report missing persons without having to visit an actual police station. Similarly the intelligence department can show most-wanted person through online and tracing the area of the accused involved in the crime by applying bayes algorithm. The users have to create a login for accessing the data to file a complaint and they can authorize them by using "Voter ID" or "Aadhar card" documents issued by Government. Admin has a login for accessing the data through the portal and give response to the user's issues. The Investigator can investigate the details and predicts the area of the accused who involved in the crime.**

Keyword-Bayesian Prediction, Crime Management System

I. INTRODUCTION

The traditional way of visiting a police station to file a complaint is time consuming, and provides minimum transparency. In the existing system only we can see the details of particular information about the police stations in our state, the existing system has more workload for the authorized person, but in the case of the proposed system, the user can register in the application and send the crime report and complaint about a particular social crime or a person.

The proposed system allows a user to login using his unique credentials – voter ID, Aadhar Card the identification documents issued by the government. In addition to filing complaints, a visitor can view updates that are issued by the administrator, and also provide feedback to higher officials in the absence of adequate action. The administrator can view the complaints and view user's feedback and update the criminal history sheet in the database. The proposed system fast tracks the existing processes by a huge margin, while also providing security, data consistency, better service via a user-friendly medium. After getting with registered complaints, the investigation process will takes place.

At the end of investigation, the previously processed government criminal records will be taken as training sets and witness record will be taken as a pattern. BAYES Theorem will be applied to trace the area of the accused involved in the crime/suspect. When implemented at scale, the system could fast-track the judicial process and make way for lesser crimes.

The operations carried are creating a user interface whereby storage and retrieval process will be carried out, through which crime reports will be registered. Pattern will be created through another interface for applying BAYES algorithm to trace the target attribute. The main module is the Investigator.

A. The Investigator

1)View Complaints: The police department officials can view the complaints that are forwarded to them by the admin.

2)Investigation Pattern: This helps the police to investigate the witness about the crime.

3)Predict the Area of Criminal: With the help of the witness, pattern are entered into database and Bayesian theorem is used to predict the area of the criminal with the training set that is available in the database.

4)Close complaint: Once the investigation is over, the accused is found out and the criminal history record is being updated into the database. These closed complaints will help as the training set for the future complaints

II. LITERATURE SURVEY

In vehicular homicide investigations, it is hard to find the culprit. The Bayesian evaluation of forensic evidence is proposed to overcome this problem. In the proposed method, the Bayes' Theorem is applied to the facts in a vehicular homicide investigation. An initial analysis of the crash dynamics in comparison with the

injury pattern and ejection status of the surviving occupant versus that of the decedent suggested that the survivor was the driver. The analysis results with estimated true and false positive rates will form the basis for a Bayesian calculation of the posterior probability of the survivor's guilt given the evidence. [1]

Bayesian rough set model (BRSM) is the hybrid development between rough set theory and prediction by bayes law, will deal the practical problems which could not be effectively handled by original rough set model. In this paper, the equivalence between two kinds of current attribute reduction models in BRSM for binary decision problems is proved. Furthermore, binary decision problems are extended to multi-decision problems in BRSM.Based on them, the approaches to knowledge reduction in BRSM can be obtained which corresponds well to the original rough set methodology. [2]

The naive Bayes classifier is proposed for the machine learning task of classification, which has been found to give good performance. Naive Bayes and other standard algorithms are applied to the database application, in which the variables are highly non-normal and found that the algorithm performed well while predicting a class that had been derived from the same data. The proposed algorithm outperformed naive Bayes in all four cases and outperformed multinomial logistic regression (MLR) in two cases where non-normal distributions are observed. [3]

Detailed and clear information is required for performing digital forensic analyses. The proposed method describes a comparative performance analysis of Bayesian and neural network techniques to classify the state of file system activities in terms of execution of applications based on the pattern of specific file manipulation. In particular, it tells the construction of a Bayesian networks and neural networks from the predetermined knowledge of the file system manipulation artifacts and its corresponding metadata information by a set of software applications. [4]

Most of the irrelevant and redundant features will delay and degrades the performance of the detection process. The proposed study is to identify important input features in building intrusion detection system (IDS) which is more efficient. Further analysis of the performance of three standard feature selection methods using Correlation-based Feature Selection, Information Gain and Gain Ratio. In the proposed method, Feature Vitality Based Reduction Method is used to identify important reduced input features. Naive bayes classifier is applied on reduced datasets for intrusion detection. [5]

A new idea has been implemented using the naïve Bayes inference to filter all the Spam and junk mails from e-mail. The effects of many parameters such as corpus size and feature extraction methods are found. Then the results of this approach are compared with many other published statistical spam filtering approaches. [6]

During the predictions using Bayesian inference, the loss occurrence will be high. Using the Bayesian predictions under Kullback – Leibler Divergence with the minimal conditions we can derive the finite bounds for the loss incurred. The variety of settings is available for the Bayesian predictions. The predictions come under model averaging, under model misspecification, arbitrary loss functions and non-stationary environment predictions. [7]

III. PROPOSED METHODOLOGY

A. Working Mechanisms

P (A/B) = [P (B/A) * P (A)] ÷ P (B)

P (A) → Probability of the class which is going to be predicted.

P (B) → Probability of the given pattern which predicts the class.

For Bayes implementation, we need a Training set and Pattern. In the training set, we have to find the overall probability of the class that we are going to predict. After finding the probability of the class we need to find the probability for all the given attributes in the training set.

For example:

Finding the Area's probability (the class which is going to predict):

Sample space = 50

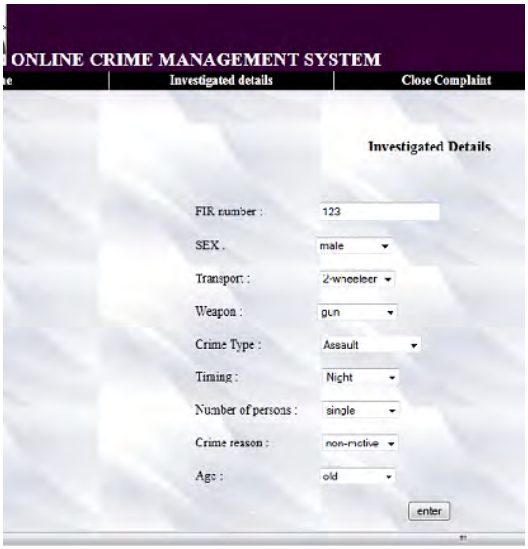
1. KK nagar = P(c1) → 12 /50 = 0.24
2. Tiruverumbur = P(c2) → 21 /50 = 0.42
3. Kattur = P(c3) → 07 /50 = 0.14
4. Vengur = P(c4) → 10 /50 = 0.20

S.no	Sex	Transport	Weapon	Crime Type	Timing	No. of. Persons	Motive	Age	Area
1.	Male	Walk	Knife	Chain	Evening	Single	No	Middle	Kattur
2.	Male	2-Wheeler	Wood	Assault	Morning	Gang	Yes	Young	Thiruverumbur
3.	Male	4-Wheeler	Rod	Murder	Night	Single	Yes	Old	Vengur
4.	Male	4-Wheeler	Gun	Robbery	Night	Gang	No	Young	Thiruverumbur
5.	Female	2-Wheeler	Rod	Chain	Night	Single	No	Middle	Thiruverumbur
6.	Male	4-Wheeler	Gun	Murder	Morning	Single	Yes	Old	Kk nagar
7.	Male	Walk	Wood	Chain	Evening	Single	No	Young	Kk nagar
8.	Female	4-Wheeler	Knife	Murder	Morning	Gang	Yes	Middle	Vengur
9.	Male	2-Wheeler	Gun	Robbery	Evening	Single	Yes	Old	Thiruverumbur
10.	Male	2-Wheeler	Knife	Assault	Morning	Gang	No	Young	Thiruverumbur
11.	Male	Walk	Rod	Chain	Evening	Single	No	Middle	Kattur
12.	Female	4-Wheeler	Gun	Murder	Night	Gang	Yes	Young	Thiruverumbur
13.	Male	2-Wheeler	Knife	Robbery	Night	Single	No	Old	Vengur
14.	Male	2-Wheeler	Knife	Murder	Morning	Single	Yes	Old	Thiruverumbur
15.	Female	2-Wheeler	Gun	Robbery	Afternoon	Gang	No	Young	Kk nagar
16.	Male	4-Wheeler	Rod	Murder	Night	Single	Yes	Old	Thiruverumbur
17.	Male	4-Wheeler	Acid	Rape	Afternoon	Single	No	Middle	Kk nagar
18.	Male	2-Wheeler	Rod	Assault	Afternoon	Gang	Yes	Young	Vengur
19.	Female	2-Wheeler	Knife	Chain	Evening	Single	No	Old	Thiruverumbur
20.	Male	4-Wheeler	Gun	Murder	Night	Single	Yes	Middle	Kattur

Fig 1. Training Set for Bayes Theorem(Simulated Government Records)

Attributes	Sub - Attributes	P(C1)	P(C2)	P(C3)	P(C4)
1. Sex	Male	0.22	0.45	0.17	0.14
	Female	0.26	0.33	0.06	0.33
2. Vehicle	2wheeler	0.21	0.47	0.05	0.26
	4wheeler	0.25	0.50	0.15	0.10
	Nil	0.27	0.18	0.27	0.27
3. Weapon	Knife	0.14	0.27	0.14	0.14
	Wood	0.28	0.42	0.28	0
	Rod	0.16	0.16	0.08	0.58
	Gun	0.31	0.50	0.12	0.06
	Acid	1.00	0	0	0
4. Crime type	Assault	0.14	0.57	0.	0.28
	Murder	0.18	0.62	0.06	0.12
	Rape	1.00	0	0	0
	Chain	0.25	0.33	0.16	0.25
	Robbery	0.30	0.30	0.20	0.20
	Cheat	0.25	0	0.50	0.25
5. Timing	Morning	0.25	0.33	0.08	0.33
	Afternoon	0.42	0.42	0	0.14
	Evening	0.23	0.53	0.23	0
	Night	0.16	0.38	0.16	0.27
6. No. of criminals	Single	0.29	0.41	0.16	0.12
	Gang	0.15	0.42	0.10	0.31
7. Motive	Yes	0.17	0.60	0.04	0.17
	No	0.29	0.25	0.22	0.22
8. Age	Young	0.25	0.40	0.10	0.25
	Middle	0.25	0.31	0.25	0.18
	Old	0.21	0.57	0.07	0.14

Fig 2. Values of all Attributes



1.fig:attribute 8sInvestigation pattern with :

After entering the above investigated details, the pattern will be created which then be evaluated with the training set (simulated government records) and starts its **"prediction"** i.e. the location of the accused involved in the crime will be traced through the created pattern. Here the major eight attributes that influences the crime,

- Sex
- Transport
- Weapon
- Crime type
- Timing
- Number of persons
- Crime reason
- Age

B. Problem Definition

Let us take the random pattern that is the investigated details about the accused involved in crime.

Pattern set= {Male, 2Wheeler, Gun, Assault, Night, Single, Non-motive, Old}

C1 = [(0.0010109) * (0.00000014475)] / (0.998875789)

= 1.66074703 x 10^-10.

Similarly,

C2 = 1.044784825 x 10^-7.

C3 = 1.290893913 x 10^-9.

C4 = 9.107771385 x 10^-9.

Since **C2 > C1, C3, C4** the predicted area of the accused is in Tiruverumbur.

C. Purity Level of Attribute Using Gini Index

By Applying Gini index we can evaluate the purity of the attribute. For Gini index application we have to find the main entropy value followed by the overall gain of the attributes.

Area: Entropy = [1- (12/50) ^2 – (21/50) ^2 – (7/50) ^2 = (10/50) ^2] = 0.7064.

Fig 2. Results

D. Gini Index Table Summarization

D – Main entropy value, R – Overall value of the attribute, I – Overall gain of the attribute.

Attributes	Sub - Attributes	R	Overall (R)	I
1. Sex	Male	0.782	1.692	-0.986
	Female	0.910		
2. Vehicle	2wheeler	0.870	2.676	-1.969
	4wheeler	0.862		
	Nil	0.943		
3. Weapon	Knife	0.891	4.608	-3.902
	Wood	0.951		
	Rod	0.903		
	Gun	0.882		
	Acid	0.980		
4. Crime type	Assault	0.860	5.509	-4.803
	Murder	0.867		
	Rape	0.980		
	Chain	0.936		
	Robbery	0.948		
	Cheat	0.920		
5. Timing	Morning	0.930	3.447	-2.741
	Afternoon	0.860		
	Evening	0.740		

	Night	0.917		
6. No. of criminals	Single	0.812	1.693	-0.986
	Gang	0.881		
7. Motive	Yes	0.802	1.608	-0.902
	No	0.805		
8. Age	Young	0.882	2.688	-1.981
	Middle	0.917		
	Old	0.888		

IV. CONCLUSION

The **Bayesian Prediction in police-crime investigation management system** is a web-based application for primarily providing training to the intelligence officers whoever tends to predicts the area of the accused.

This software application has been computed successfully and was also tested successfully by taking "test cases". It is user friendly, and has required options, which can be utilized by the investigator to perform the desired operations.

The goals that are achieved by the software are:

- Instant access.
- Improved productivity.
- Optimum utilization of resources.
- Efficient management of records.
- Simplification of the operations.
- Less processing time and getting required information.
- User friendly.

REFERENCES

[1] Michael D. Freeman, Annette M. Rossignol, Michael L. Hand - *"Applied forensic epidemiology: The Bayesian evaluation of forensic evidence in vehicular homicide investigation"*, Journal of Forensic and Legal Medicine, Volume 16, Issue 2, February 2009, Pages 83-92.

[2] Hongyun Zhang, Jie Zhou, Duoqian Miao, Can Gao - *"Bayesian rough set model: A further investigation"*, International Journal of Approximate Reasoning, Volume 53, Issue 4, June 2012, Pages 541-557.

[3] Daniele Soria, Jonathan M. Garibaldi, Federico Ambrogi, Elia M. Biganzoli, Ian O Ellis - *"A non-parametric version of the naive Bayes classifier"* , Knowledge-Based Systems, Volume 24, Issue 6, August 2011, Pages 775-784.

[4] Muhammad Naeem Ahmed Khan - *"Performance analysis of Bayesian networks and neural networks in classification of file system activities"*, Computers & Security, Volume 31, Issue 4, June 2012, Pages 391-401.

[5] Saurabh Mukherjee, Neelam Sharma - *"Intrusion Detection using Naive Bayes Classifier with Feature Reduction"*, Procedia Technology, Volume 4, 2012, Pages 119-128.

[6] Lawrence M.Rudener & Tahung Liang - *"Automated essay scoring using bayes theorem"* JTLA - The Journal of Technology, Volume 1, Number 2 , June 2002, Pages 1966-1994, University of Maryland.

[7] Alessio Sancetta – *"Universality of Bayesian predictions"*, International Society for Bayesian Analysis, Volume 7, Issue 1, October 2012, Pages 1-36.

[8] Emanuele Olivetti, Sriharsha Veeramachaneni, Ewa Nowakowska - *"Bayesian hypothesis testing for pattern discrimination in brain decoding"*, Pattern Recognition, Volume 45, Issue 6, June 2012, Pages 2075-2084.

[9] Shikui Tu, Lei Xu - *"A theoretical investigation of several model selection criteria for dimensionality reduction"*, Pattern Recognition Letters, Volume 33, Issue 9, 1 July 2012, Pages 1117-1126.

[10] Rui Yang, Sigurdur Olafsson - *"Classification for predicting offender affiliation with murder victims*" , Expert Systems with Applications, Volume 38, Issue 11, October 2011, Pages 13518-13526.

[11] Janick V. Frasch, Aleksander Lodwich, Faisal Shafait, Thomas M. Breuel - *"A Bayes-true data generator for evaluation of supervised and unsupervised learning methods"* , Pattern Recognition Letters, Volume 32, Issue 11, 1 August 2011, Pages 1523-1531.

[12] G. Alan Wang, Homa Atabakhsh, Hsinchun Chen - *"A hierarchical Naïve Bayes model for approximate identity matching"*, Decision Support Systems, Volume 51, Issue 3, June 2011, Pages 413-423.

[13] A. Thomas, B. John Oommen - *"The fundamental theory of optimal "Anti-Bayesian" parametric pattern classification using order statistics criteria"*, Pattern Recognition, Volume 46, Issue 1, January 2013, Pages 376-388.

[14] Sean L. Humpherys, Kevin C. Moffitt, Mary B. Burns, Judee K. Burgoon, William F. Felix - *"Identification of fraudulent financial statements using linguistic credibility analysis"* , Decision Support Systems, Volume 50, Issue 3, February 2011, Pages 585-594.

[15] Mohamed M. Mostafa - *"A neuro-computational intelligence analysis of the global consumer software piracy rates"* , Expert Systems with Applications, Volume 38, Issue 7, July 2011, Pages 8782-8803.
[16] Chien-Lung Chan, Hsien-Wei Ting - *"Constructing a novel mortality prediction model with Bayes theorem and genetic algorithm"* , Expert Systems with Applications, Volume 38, Issue 7, July 2011, Pages 7924-7928.
[17] Pierrick Bruneau, Marc Gelgon, Fabien Picarougne - *"A low-cost variational-Bayes technique for merging mixtures of probabilistic principal component analyzers"*, Information Fusion, Volume 14, Issue 3, July 2013, Pages 268-280.
[18] Sheng-Tun Li, Shu-Ching Kuo, Fu-Ching Tsai - *"An intelligent decision-support model using FSOM and rule extraction for crime prevention"* , Expert Systems with Applications, Volume 37, Issue 10, October 2010, Pages 7108-7119.
[19] E.W.T. Ngai, Yong Hu, Y.H. Wong, Yijun Chen, Xin Sun - *"The application of data mining techniques in financial fraud detection: A classification framework and an academic review of literature"* , Decision Support Systems, Volume 50, Issue 3, February 2011, Pages 559-569.
[20] Wen-Hsi Chang, Jau-Shien Chang - *"An effective early fraud detection method for online auctions*", Electronic Commerce Research and Applications, Volume 11, Issue 4, July–August 2012, Pages 346-360.
[21] Xiaomo Jiang, Sankaran Mahadevan, Angel Urbina - *"Bayesian nonlinear structural equation modeling for hierarchical validation of dynamical systems"* , Mechanical Systems and Signal Processing, Volume 24, Issue 4, May 2010, Pages 957-975.
[22] E. Fersini, E. Messina, F. Archetti - *"Emotional states in judicial courtrooms: An experimental investigation"*, Speech Communication, Volume 54, Issue 1, January 2012, Pages 11-22.