AN APPLICATION TO PREDICT THE CRIME

[1]K.SUBRAHMANYA KOUSIK, [2]Dr. M.SAMPATH KUMAR

[1]M.TECH SCHOLAR, [2]PROFESSOR

Department of Computer Science & Systems Engineering,
Andhra University College of Engineering (A), Visakhapatnam, India

Abstract: Crime detection is the one of the big challenge to police and crime detectives. So many cases are in pending state. In this project we suspect the criminal and predict the crime using existing data. That means we have some existing data. When the case was filed into the police station, we will give input to the application. Then that application compares the input with the existing database and gives some results which are matching to the database with similarity percentage. Then we have some idea that which records have similarity percentage >60, we can investigate those criminal and suspect the criminal and predict the crime.

I. INTRODUCTION

Historically crimes are solved manually only if they have evidence and proper clue. Due to that manual so many cases are in pending and so many criminals have a chance to escape. If police have clues, it takes a lot of time to trace the case and suspect the criminal and predict the crime. Another issue of historical system is there in no proper format of data. If we want to access previous criminal data we need to go through the paper work. There is no proper data storage. So i would like to design an application and software to able trace and suspect the criminals. Due to this application we can trace the cases as able as possible and we can trace the cases up to 70%. Criminals are creating social nuisance and disturbance to the public. Any criminal often repeat the crime if crime done without any failures and if he didn't found by police or any public, he will repeat the pattern of the crime. So based on that technique we can suspect the criminal and predict.

II. CRIME ANALYSIS PROCEDURE

This crime analysis is the tedious process to the police or crime detectives. We have an existing database which contains all criminal data. With the help of the data we can analysis the crime and suspect the criminal. I design an application which connects to that database. The police officer have to login to go through the application. After he logged in successfully, he can give an input to the application. That application checks with that existing database and give some result which the input similar to the data in the existing database. Similarly that application gives a similarity percentage between the input value and similarity values of the existing database. Based on the similarity percentage we can suspect the criminal. If the similarity percentage is greater than 60%, the police team investigates them and there is a chance to trace the criminals. Due to this procedure we can trace the cases easily and it takes less time.
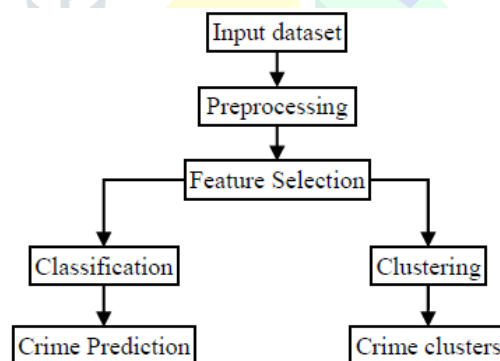

Fig.1. crime prediction and crime clustering

III. ALGORITHM USED

In this section, the proposed Naive Bayesian network model is formulized for the criminal prediction problem. Naïve Bayesian classifier is a simple probabilistic classifier which works by applying the Bayes' theorem along with Naïve assumptions about feature independence. Despite its independence assumption, Naïve Bayesian classifier is proved to be quite useful in modelling the conditions of complex real-world problem. Throughout this section, the incident variables are indicated in bold and lowercased fonts, whereas their values are in italic fonts with uppercased capital. The proposed model, shown in below figure is constructed to express the probability of each criminal when the incident date and location, crime and criminal names, and criminals' acquaintances are given as clues. When used as a learner, the best model is the one that maximizes the learning performance, rather than best expressing the true relations among variables. In this context, crime, location and date variables are considered to act on the criminal's variable. The network formed by the criminal's acquaintances has also influences on criminals. But here, those except the acquaintances and acquaintances of the acquaintances (secondary acquaintances) are not considered due to its negligible effect on the result and reducing the computational complexity.
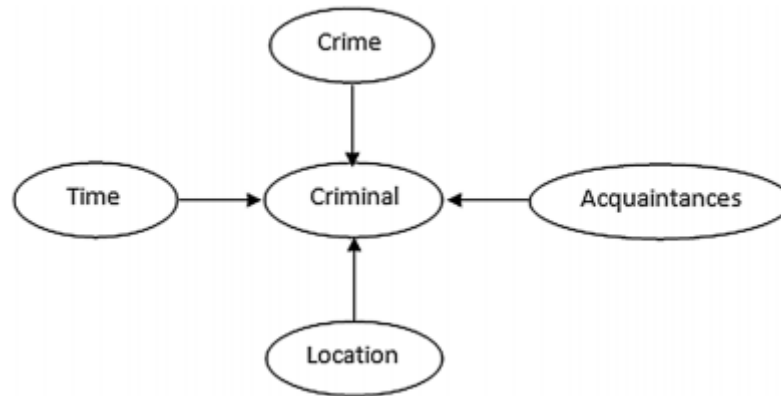
Fig.2. system model for suspect criminal

With the proposed Naive Bayesian network, the criminal probability is:

P(criminal|location,time,crime,acquaintances) = P(criminal,location,time,acquaintances)/P(location,time,crime,acquaintances)   Eq.1

P(criminal|location,time,crime,acquaintances)= P(location) * P(Criminal | Location) * P(time) * P(Criminal | Time)

* P(crime) * P(Criminal | Crime) * P(acquaintances) * P(Criminal | Acquaintances)   Eq.2

IV. CRIMINAL ANALYSIS METHODS

I.   **By Evidence:** Generally, if any incident happened there is some clue is there. We take that clue as evidence and go through the existing database and compare with that database. Then we have some results will display. For all records we have a similarity percentage is there. So which records have similarity percentage is higher than 60, we have slide that records and investigation will happen about that records.

II.   **By Repeated Sequence**: Generally, the criminal follow the same steps which he followed previously to do the crime. If the crime which done previously is success he will repeat the step which he done previously. So we have that data also with the help of that data we can suspect the criminal by comparing the previous steps he follow. Then we can predict the crime how the crime happened.

III.   **By Location:** The criminals have some favoured locations. That means if the crime done that particular location, there is maximum possibilities of particular criminals that location tagged to some criminals. We can predict the crime using the location based also. How we can predict the crime using location means, while giving input to application it compares to the existing data and give some results that whether crime exists in that location previously.

IV.   **By Time:** Time is the major factor to trace the case. Because the criminal have some favour time to involve the crime. So we can trace the case by time also. While comparing with the existing database with the current values with respect to time, it shows some records with similarity percentage. We can sort out those record based on the similarity percentage. Which similarity percentage is greater than 60%. So we can investigate those criminal to trace the case.

V.   CONTROL FLOW

Login: Initial we need to login to enter into the home page. We need to give user name and password. If the values are authenticated then we enter into the home page. If the values are not authenticated the application will give an error message. In this page there are another two buttons are there. In this field we have reset and exit functions also. We can reset the values of user name and password at the same time we can exit from the page.

Home: The second step in this application is home page. In this page we have three modules are there. The first module performs store the data into the respective databases. That means all the data can be stored in their respective positions. The second module is the classification part. In this part we can classify the data into two sub modules. The sub module is pending cases and another sub module is traced cases. The records which have criminal status is "No", those records are transferred into pending cases. The criminal status is "Yes" those records are transferred into Traced cases.

Trace the case: In this part, we give an id of the case to the application. It checks with the database and give some similarity record with similarity percentage. Based on the similarity percentage we can investigate the criminals and find the criminals and trace the case and close the case. After tracing the case the status of the case is changed in the database at the same time the data is transferred into the traced cases. After Successful tracing the criminal, the record will update. That means the status of the crime and name of the criminal will be updated.

VI. RESULTS AND DISCUSSION

Table 1. quality analysis

Method	Input	Action Performed	Outcome
Login	Username, Password	User Authentication	If both username and password both are correct, switch to home page, else give an error message
New Case	Values of Sno,casetype,victim,location,time,date,criminalfound,criminal,clues	Load the data into the database.	Store the data which given by user have to load into database successfully. If any record which associated with same victim name it should inform to the user.
Pending Cases	--	Classification	Retrieve all the data which criminalfound='No'
Solved Cases	--	Classification	Retrieve all data from database which criminalfound='Yes'
Trace The Case (Get Details)	S.no of the case which is in pending state	Retrieving Data	Retrieve all information which belongs to the given serial no
Trace the case (Trace)	S.no of the case which is in pending state	Prediction	Retrieve all information which similar to the given serial number.
Trace the case (transfer to traced cases)	S.no of the case which is in pending state	Update	Update crime status and name of the criminal.

For performing the quantitative analysis of the methods taken, the performance metric value needed to be computed and they are to be compared with the other. Hence, for performing the calculations of the performance metric there are a few formulas which can be utilized for achieving the performance value from the dataset. The formulae for the calculation of the performance metrics are given below in table.

Table.2. quantitative analysis

METRIC	FORMULA
Accuracy	(TP+TN) / (TP+TN+FP=FN)
Error Rate	100 - Accuracy
Precision	TP/(TP+FP)
Recall	TP/(TP+FN)
F- Value	2/(1/P+1/R)

Table.3 Performance Analysis

S.NO	NAME OF THE METHOD/ SYSTEM	PERFORMANCE METRIC	PERFORMANCE VALUE	
1	Navai baise classification with GA for optimizing the parameters	Accuracy of Prediction	Optimized parameter	93.64%
			Non- Optimized parameter	87.74%
		Classification Error	Optimized parameter	7.36%
			Non- Optimized parameter	11.26%
		Fitness Function	Optimized parameter	72.28%
			Non- Optimized parameter	72.48%
2	SIIMCO	Recall	0.62	
		Precision	0.56	
		F-Value	0.59	
	Crime Net Explorer	Recall	0.36	
		Precision	0.41	
		F-value	0.38	
	Log Analysis	Recall	0.53	
		Precision	0.51	
		F-value	0.52	

VII. CONCLUSION

In this paper, we have studied some known approaches for crime prediction and suspect criminal concerned with data mining. Although many papers have been studied, only those papers with background in the crime prediction and criminal identification papers are compared with a theoretical study. Each paper has their own advantages and disadvantages. Each paper has its own individual approach for solving the crimes and criminal prediction. This is a practical study for several methods in identification of crime and criminals which include Text based methods, crime patterns and crime evidence based methods and finally Prisoner based methods. The data mining techniques studied from this application can be applied for identifying the criminals in the society and also for providing a better future to live in.

VIII. REFERENCES

[1] Nazlena Mohamad Ali1, Masnizah Mohd2, Hyowon Lee3, Alan F. Smeaton3, Fabio Crestani4 and Shahrul Azman Mohd Noah2 ,2010 Visual Interactive Malaysia Crime News Retrieval System

[2] M. H. Dunham, Data Mining: Introductory and Advanced Topics, Prentice Hall , 2002.

[3] https://www.edureka.co/blog/implementing-kmeans-clustering-on-the-crime-dataset/

[4] Malathi , A; Santhosh Baboo , S, 2011 An Enhanced Algorithm to Predict a Future Crime using Data Mining

[5] Aravindan Mahendiran, Michael Shuffett, Sathappan Muthiah, Rimy Malla, Gaoqiang Zhang,2011 Forecasting Crime Incidents using Cluster Analysis and Bayesian Belief Networks

[6] K. Zakir Hussain, M. Durairaj and G. Rabia Jahani Farzana ,2012 Application of Data Mining Techniques for Analyzing Violent Criminal Behavior by Simulation Model

[7] Rasoul Kiani, Siamak Mahdavi and Amin Keshavarzi, "Analysis and Prediction of Crimes by Clustering and Classification", International Journal of Advanced Research in Artificial Intelligence, Vol. 4, No. 8, pp. 11-17, 2015.

[8] Jeroen S. De Bruin, Tim K. Cocx, Walter A. Kosters, Jeroen F. J. Laros and Joost N. Kok, "Data Mining Approaches to Criminal Career Analysis", Proceedings of 6th IEEE International Conference on Data Mining, pp. 1-7, 2006.

[9] Kamal Taha and Paul D. Yoo, "SIIMCO: A Forensic Investigation Tool for Identifying the Influential Members of a Criminal Organization", IEEE Transactions on Information Forensics and Security, Vol. 11, No. 4, pp. 811-822, 2016.

[10] Tong Wang et al., "Learning to Detect Patterns of Crime", Joint European Conference on Machine Learning and Knowledge Discovery in Databases, pp. 515-530, 2013.