Crime Pattern Detection Using Historical Data

Imran Ahmed, Neha Jaison, Shahabas Ahammed, Shireen Shajahan

[1]*Student, Dept. of Computer Science and Engineering, IES College of Engineering, Kerala, India*
[2]*Student, Dept. of Computer Science and Engineering, IES College of Engineering, Kerala, India*
[3]*Student, Dept. of Computer Science and Engineering, IES College of Engineering, Kerala, India*
[4]*Student, Dept. of Computer Science and Engineering, IES College of Engineering, Kerala, India*

ABSTRACT

The objective of the project is to create a platform where the police or responsible authorities can predict the location and observe patterns of crime. With help of crime data modeling the crime pattern can be detected or worked upon. Crime data analysis can even speed up the resolution to the court cases with ease and timely. Such types of crime prediction approach work on patterns of crimes in terms of similarity of attack or crime such as time of incidence, the methodology of crime is same etc. Collect the data: In this phase the data from various sources is collected from various government sources, social media platforms about the incident like Facebook and blogs. It may expect that the data received may be in unstructured form with different types and sizes. Classification; The standard algorithm useful for classification which may be applied is Naïve bayes Classifier. This classifier will determine the probability of falling into different classes of crimes and the crime predicted to belong to a specific class the probability is highest. Identify pattern: Next phase in the methodology is to find the sequence of crimes which are similar in nature and belong to the same class. Such a pattern may be identified as suggested in literature through the apriori algorithm. For the prediction outcome the decision trees are used. Here in the decision tree each internal node has a test for occurrence on an incidence and outcome are normally yes or no and based on the outcome again the next level of internal node has another question test and so on to reach a final decision for prediction. This type of analysis on historical data can also help in determining the criminal profile based on the characteristic behavior inferred from the data. It can help the investigator to accurately predict the profile of unknown criminals.

Keyword: *Crime data analysis, Machine Learning. Data Analytics, Classification*

1.   INTRODUCTION

The crime incidents area unit continued to grow in rate and complexity. Crime cannot be foreseen since it's neither systematic nor random. In many criminal investigations, crime analysis includes processes to find the criminal of incidents. The discovery of possible suspects in advance is an important requirement for security forces to optimize the usage of human and technical resources and carry out important functions in the prediction of crime. Crime Records Bureau crimes like burglary, arson, etc. have decreased while crimes like murder, sex abuse, gang rape, etc. have increased. Even though we cannot predict the result with 100% accuracy, the application gives the result more chances of attempting it. Crimes nowadays are increasing day by day and with different levels of intensity and versatility. The result is a great loss to society in terms of monetary loss and social loss and further, it enhances the level of threat against the smooth livelihood in the society. To overcome this problem, the computing era can help to reduce crime or even may help predict crime so that sufficient measures can be taken to minimize the loss of property and life. The crime rate prediction strategies can be applied to historical data available in the police records

by examining the data from various angles including the reason for crime, and frequency of similar kinds of crimes at specific locations with other parameters to prepare a model for the crime prediction. It is a major challenge to understand the versatile data available to us and then model it to predict future incidence with acceptable accuracy and further reduce the crime rate.

2. LITERATURE SURVEY

Crimes are a social irritation and cost our society deeply in several ways. Any research that can help in solving crimes quickly will pay for itself. About 10% of the criminals commit about 50% of the crimes [9]. The system is trained by feeding previous years record of crimes taken from legitimate online portal of India listing various crimes such as murder, kidnapping and abduction, dacoits, robbery, burglary, rape and other such crimes.

As per data of Indian statistics, which gives data of various crime of past 14 years (2001-2014) Sunil Yadav , Meet Timbadia , Ajit Yadav , Rohit Vishwakarma , Nikhilesh Yadav have created a regression model and the crime rate for the following years in various states can be predicted [8]. They have used supervised, semi-supervised and unsupervised learning techniques [4] on the crime records for knowledge discovery and to help in increasing the predictive accuracy of the crime. This work will be helpful to the local police stations in crime suppression.

According to a paper by Veyus Gunes, an independent researcher in Le Vans, France proposed a general framework as a further step in designing intelligent machines.The work is not solely focused on recent problems but the ones that have been relevant for the past 60 years. It is of great importance that machines have realistic models and classification paradigms(ie. techniques and algorithms) for the targeted area of research. They have mentioned standards, prototypes and models as having the same meaning, and features, coordinates, parameters, variables, criterions and attributes as having the same meaning, for a unified approach to AI. A brief taxonomy of existing intelligent machines and the involved research and application fields will be introduced. Then the Turing test will be discussed. After that, a definition of spaces and a framework for intelligent machines will be introduced. Finally, the last section includes a conclusion and some prospects.

Shyam Varan Nath used data mining to model crime detection problems. Crimes are a social nuisance and cost our society dearly in several ways. Any research that can help in solving crimes faster will pay for itself. Here he looked at the use of clustering algorithms for a data mining approach to help detect the crime patterns and speed up the process of solving crime. He even looked at k-means clustering with some enhancements to aid in the process of identification of crime patterns. He applied these techniques to real crime data from a sheriff's office and validated his results. He also used a semi-supervised learning technique here for knowledge discovery from the crime records and to help increase the predictive accuracy. He also developed a weighting scheme for attributes to deal with limitations of various out of the box clustering tools and techniques. He states that this easy to implement data mining framework works with the geospatial plot of crime and helps to improve the productivity of the detectives and other law enforcement officers. It can also be applied for counter terrorism for homeland security

Zengli WangEarly research found that crimes are not randomly distributed but tend to follow patterns. For example, researchers observed that disproportionate crimes occur shortly after the previous crimes and call this phenomenon repeat victimization [4,5,6,7]. In addition, many places exhibit similar crime distribution patterns due to their similar geographic features [8,9]. According to displacement or diffusion of benefits, nearby locations may experience repeated high crime risk due to the dynamic nature of criminal opportunities [10,11,12,13,14,15,16,17]. Many studies have been conducted to detect and interpret these phenomena [4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22]. However, they are typically designed to investigate one particular pattern type and often overlook the complexity of crimes [17]. Crime changes across space, time, and culture [23,24]. Moreover, they are not generalized enough to detect, display, and interpret combinations of those patterns simultaneously. Considering the complexity of crimes, he studied a comprehensive model. Utilizing graph representation, he efficiently and effectively discovered crime patterns, including repeat and near-repeat (RNR) co-occurrence, and geographic shift. Besides, he introduced and studied a combination of these patterns. Compared to previous studies, he aimed to detect, demonstrate, and analyze crime relations that have not been studied yet. The research results are expected to improve crime research and analysis by providing multiple crime patterns. This method can be extended to other research areas that use point events as research objects.

Lalitha Saroja Thota, Mohrah Alalyan, Al-Otaibi Awatif Khalid, Fabiha Fathima made cluster analysis by using k-means cluster algorithm on the criminal dataset of India. The CrimeInfo report of National Crime Records Bureau (NCRB), India collects, analyzes and publishes the crime data. The crime profiling and zoning can be modeled with utilization of data mining. The cluster input is used to create a custom India map with the cluster zones of states. The custom map displays an overall crime profile of states which helps police and law enforcement departments to take additional preventive measures to combat against the crime and plan advanced investigation strategies. The crime

trend and zoning knowledge can also be helpful in cautioning police to increments and reductions in levels of actions.

Ravi Kumar & Bharti Nagpal had introduced a paper with a solution to the crime prediction problem using Naive Bayes classifier, which includes finding the most likely criminal of a particular crime incident when the history of similar crime incidents has been provided with the incident-level crime data. The incident-level crime data is provided as a crime dataset which includes incident date and location, crime type, criminal ID and the acquaintances are the attributes or crime parameters. The acquaintances are the suspects whose names are either directly involved in the incident or indirectly the acquaintances of the criminal. Acquiring a real-time crime dataset is a difficult process in practice due to confidentiality principles. So, crime datasets are used for the inputs using the state of the art methods. The proposed system is tested for the crime prediction problem using data learning, and the experimental results show that the proposed system provides better results and finding of the potential solutions and crime patterns.

Data mining is a powerful tool which enables investigators to explore large criminal and crime databases quickly and efficiently. D.E. Brown (1998) [3] constructed a software framework called ReCAP (Regional Crime Analysis Program) for mining data in order to catch professional criminals using data mining and data fusion techniques.In 2009, Li Ding et al. [4] propose an integrated system called PerpSearch that takes a given description of a crime,including its location, type, and the physical description of suspects (personal characteristics or vehicles) as input. To detect suspects, the system will process these inputs through four integrated components: geographic profiling, social network analysis, crime profiles, and physical matching. Essentially, geographic profiling determines "where" the suspects are, while other components determine "who" the suspects are.
In 2012, K. Zakir Hussain, et al. [5] presented Criminal behavior analysis by using data mining techniques. He elaborated the criminal behavior analysis of the offenders by using data mining techniques.
In 2013,Kaumalee Bogahawatte & Shalinda Adikari [6] developed a system named Intelligent Criminal Identification System (ICIS) that can potentially identify a criminal based on the evidence collected from the crime spot. The solution has provided for three crime categories namely robbery, burglary and theft out of 2 1 categories of grave crimes.

The theory is formed around the assumption that human beings are born neutral (i.e. neither with conformity nor unruly). Humans learn through imitation. People learn from the social surroundings and the behavior of others around them. In the first part of this paper the author talks about all the strains that are responsible for developing negative characteristics amongst individuals. One such character which is a commonality in criminals is 'apathy' (i.e. lack of empathy or desire for the welfare of others). [2] The process of determining such personality traits is called criminal profiling. [3] The process helps to determine motive by categorizing the individual's behavioral tendencies, demographic variables and geographic location. It forms a database of an individual's life instances and their psychological consequences over his psyche which further determines whether the individual has a tendency to become a repetitive offender or not. This paper discusses how a criminal mind is formed and how different psychological, social and biological factors contribute to this in different stages of their lives. Guilt and free will generate complex emotions in an individual's mind. Crimes can be committed by people who are completely aware of what they are doing which is possibly being committed due to a strained social circumstance. It can also be committed by people who are going through deep psychological issues who commit them for pleasure and exercise a certain form of sadism. All these are results of human behaviour and it follows a certain pattern. As mentioned in the beginning of this paper it is said that an individual's actions are more likely to be an imitation or a response to the surroundings they were constantly subjected to. These are key details in determining not only the pattern of crime but also the root causes of a criminal personality development.

3. CONCLUSIONS

We looked at the use of data mining for identifying crime patterns using the clustering techniques. Our contribution here was to formulate crime pattern detection as a machine learning task and to thereby use data mining to support police detectives in solving crimes. We identified the significant attributes; using expert based semi-supervised

learning methods and developed the scheme for weighting the significant attributes. Our modeling technique was able to identify the crime patterns from a large number of crimes making the job for crime detectives easier. Some of the limitations of our study include that crime pattern analysis can only help the detective, not replace them. Also, data mining is sensitive to quality of input data that may be inaccurate, have missing information, be data entry error prone etc. Also mapping real data to data mining attributes is not always an easy task and often requires skilled data miner and crime data analyst with good domain knowledge. They need to work closely with a detective in the initial phases. As a future extension of this study, we will create models for predicting the crime hot-spots that will help in the deployment of police at most likely places of crime for any given window of time, to allow most effective utilization of police resources. Key features of this project are displaying hospital success rates and creating an access to hospital records to start an investigation, providing a database to update new cases with the aim of digitizing FIRs. The old FIRs can also be viewed and the soft copies of these documents can be accessed with ease.We also plan to look into developing social link networks to link criminals, suspects, gangs and study their interrelationships. Additionally, the ability to search suspect descriptions in regional FBI databases, to traffic violation databases from different states etc. to aid the crime pattern detection will also add value to this crime detection paradigm. We have only used open source softwares for this project and there is no hardware implementation making this project highly financially feasible. Our project has incorporated some necessary features in the crime detection part and it is extremely user friendly.

4. REFERENCES

[1]Gera P, Vohra R (2014) City crime profiling using cluster analysis. Int J Comput Sci Inf Technol 5(4):5145–5148

[2]S. Dzeroski, "Towards a General Framework for Data Mining", Proc. 5th International conf. on Knowledge discovery in inductive databases. Springer, Berlin, (2007), p. 259.

[3] D.E. Brown, The regional crime analysis program (RECAP): A framework for mining data to catch criminals, in Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics, Vol. 3, pp. 2848-2853, 1998.

[4] K. Zakhir Hussain, M. Durairaj and G. Rabia lahani Farzana ,Criminal behavior analysis by using data mining techniques, IEEEInternational Conference On Advances In Engineering, Science And Management (ICAESM-2012) March 30, 31, 2012, pp.656-658,2002.

[5]C. H. Chen and P. G. P. Ho, "Statistical pattern recognition in remote sensing", Pattern Recognition, 41 (9) (2008), p. 2731

[6]Greenwald, H., & Roche, P. Q. (1959). The Criminal Mind: A Study of Communication Between Criminal Law and Psychiatry. American Sociological Review, 24(5), 741. doi: 10.2307/2096192

[7]Cowan, P. A., & Cowan, C. P. (n.d.). From family relationships to peer rejection to antisocial behavior in middle childhood. Children's Peer Relations: From Development to Intervention., 159–177. doi: 10.1037/10653-009.