Crime and Criminal Activity Analysis to Forecast Crime

MD ASHRAF HUSSAIN (ROLL NO: 16)
SHOWMIK AZIZ (ROLL NO: 32)

Department Of Computer Science And Engineering

Supervisor
Md Mahmudur Rahman
Lecturar
Department Of Computer Science and Engineering
University of Dhaka

4 March 2020

Contents

Introduction

Crime is one of the most serious problems in our society. Basically crime is unpredictable and mostly unstoppable. Crime analysis is a systematic way of detecting and investigating patterns and trends in crime. Crime analysis plays a crucial role in improving public security and reducing the financial loss of crimes. The crime activities have been increased at a faster rate. Therefore, there is an urgent demand for crime prediction.

1.1 Problem Definition

Our study aims to find criminal hotspots using a set of real-world datasets of crimes. We will try to locate the most likely crime locations and their frequent occurrence time. In addition, we will predict what type of crime might occur next by a particular group in a specific location within a particular time. Finally, we intend to provide an analysis study by combining our findings of a particular crimes dataset with its demographics information.

1.2 Motivationtion

Bangladesh is one of the most populated countries with more than 160 million people. Every day, crime events are occured in so many places. Comparatively Bangladesh is relatively a poor country. It has limited resources. The number of police forces or other forces who are engaged in preventing crimes are also limited. If we can help the government or police force to utilize this limited resource in a proper way, it will be a great help to the country to reduce the criminal activities.

1.3 Objective

Nowadays crime has become a common issue in our society. Although crimes could occur everywhere, it is common that criminals work on crime opportunities they face in most familiar areas for them. Therefore, our proposed solution can potentially help people stay away from the locations at a certain time of the day along with saving lives. In addition, having this kind of knowledge would help people to improve their living place choices. On the other hand, police forces can use this solution to increase the level of crime prediction and prevention. Moreover, this would be useful for police resources allocation. It can help in the distribution of police at most likely crime places for any given time, to grant an efficient usage of police resources and reduce the financial loss of crimes. To minimize all these problems, we want to provide a data mining approach to determine the most criminal hotspots and find the type, location and time of committed crimes. We hope to raise people's awareness regarding the dangerous locations in certain time periods.

Related Works

2.1 Paper Review

Criminal activities are common around the world. Therefore, researchers have completed many works on this subject matter. In this section, we briefly introduce the work related to our study.

Almanie et al. 2015 used 3 datasets to predict the future possible crime in Denver and Los Angeles. They applied the Apriori algorithm to find frequent crime patterns in both cities. After that, they applied Decision Tree and Naive Bayesian classifiers to help predicting future crimes in a specific location within a particular time. They achieved 51% of prediction accuracy in Denver and 54% prediction accuracy in Los Angeles. They aimed to further understand their models' findings and to capture the factors that might affect the safety of neighborhoods.

Agarwal et al. 2019 worked on the features affecting the attacks and their involvement for future event prediction. Various methods for predicting the attacker groups, prediction of success and the effect due to changes in weather conditions has been proposed. The implementation of Random Forest, Logistic Regression, Naive Bayes, Decision Tree, K-Means Clustering and the dummy classifier has shown improvement.

Goel et al. 2019 analyzed the Global Terrorism Dataset and produced some beneficial and interesting results. The analysis and evaluation were done using the Orange data mining tool. Orange facilitates to analyse GTD using different data mining methods and compare their

results. The benefit of this tool is that, according to the attributes chosen, the best classifier can be found to produce more accurate results for future references and it can help to achieve global security. Being a paper of national importance and catering to global problems, this paper can be further used on a larger scale to look into the future of terrorism. The data can be drilled down to look specifically for India particularly, or could be rolled up and viewed on a global level as well. The data set can be diced or sliced too, based on what kind of information needs to be extracted.

Sivaranjani et al. 2016 presented the method to predict and forecast crimes in six cities of Tamilnadu. Clustering techniques were used for crime detection and classification techniques were used for crime prediction. The K-Means clustering, Agglomerative hierarchical clustering and DBSCAN clustering were implemented and their performance was evaluated based on accuracy. On comparing their performance the DBSCAN clustering gives result with high accuracy and effectively forms clusters than other two algorithms. The KNN classification is used for predicting crimes based on similarity search. Thus this system assists law enforcement agencies for an improved and accurate crime analysis. this work can be extended to have improved classification algorithms to enhance privacy and other security measures to protect the crime data.

Zhao and Tang 2017 propose a novel framework TCP, which captures temporal-spatial correlations including intra-region temporal corre- lation and the inter-region spatial correlation for crime prediction. TCP utilizes heterogeneous urban sources, e.g., public security data, meteorological data, point of interests (POIs), human mobility data and 311 complaint data. They evaluate their approach with extensive experiments based on real-world urban data about New York City. The results show that their framework can accurately predict crime numbers in the future, temporal-spatial correlations can help crime prediction and more temporal-spatial patterns could be used to advance crime prediction.

In Yamini 2019, a multiple clustering approach is proposed based on fuzzy clustering theory. The FCM algorithm works how an individual data point has been grouped in the multiple clusters. The final results are used to analyze the crime prone states in US so that it can be stopped by enhancing the security level in those regions. The results are only helpful

for crime analysis but there is a requisite of analyzing the crime patterns that can occur in future. The prediction of crimes is impossible, but it can be prevented if the time in which the crime is going to happen is known. In future, the pattern analysis of imminent crime can be performed using association rule mining along with proposed system. Moreover, the work can be extended to predict the time in which crime may happen.

Jia 2019 used some machine learning and data mining techniques for crime prediction and analysis. Classification analysis algorithms uses attributes in the data set to predict the value of one or more variables taking discrete values.Regression analysis algorithms use the properties of the data set to predict the value (e.g. profit and loss) of one or more variables taking continuous values. Predictive variable importance of crime types for Random Trees, Bayesian Network and Neural Networks. Model accuracy for Random Trees is much higher than Bayesian Networks. The performance of the Bayesian network algorithm is relatively poor, probably due to the existence of certain random factors in various crimes and related features.

CHAPTER 3

Summary

One of the paper shows that temporal-spatial correlations can help crime prediction.If we use more temporal-spatial patterns could be used to advance crime predictions.

One of the paper uses only urban data but it is possible to use other sources such as criminal networks, social networks and urban environments.

One of the paper can be used to analyze the crime prone states in the US. But there is a requisite of analyzing the crime patterns that can occur in future. So prediction algorithms can also be used with this proposed system.

Some paper uses clustering approaches for detecting and predicting crimes. This can be improved by using improved clustering approaches which will increase the performance highly. Data preprocessing can also improve performance.

In another work, authors tried to predict crimes based on crime types using spatial and temporal criminal hotspots. But their accuracy was not quite good. Accuracy might be improved using some other classification methods.

Another work made a comparison of machine learning approaches in the prediction of crimes. They only used supervised methods to implement some features. But more features could be introduced by using some unsupervised methods.

Bibliography

Agarwal, Palak, Mahak Sharma and Satish Chandra (2019). 'Comparison of Machine Learning Approaches in the Prediction of Terrorist Attacks'. In: *2019 Twelfth International Conference on Contemporary Computing (IC3)*. IEEE, pp. 1–7.

Almanie, Tahani, Rsha Mirza and Elizabeth Lor (2015). 'Crime prediction based on crime types and using spatial and temporal criminal hotspots'. In: *arXiv preprint arXiv:1508.02050*.

Goel, Mansi, Nikitasha Sharma and Mahendra Kumar Gurve (2019). 'Analysis of Global Terrorism Dataset using Open Source Data Mining Tools'. In: *2019 International Conference on Computing, Power and Communication Technologies (GUCON)*. IEEE, pp. 165–170.

Jia, Xueming (2019). 'Crime Prediction Using Data Mining and Machine Learning'. In: *The 8th International Conference on Computer Engineering and Networks (CENet2018)*. Vol. 905. Springer, p. 360.

Sivaranjani, S, S Sivakumari and M Aasha (2016). 'Crime prediction and forecasting in Tamilnadu using clustering approaches'. In: *2016 International Conference on Emerging Technological Trends (ICETT)*. IEEE, pp. 1–6.

Yamini, M Premasundariand C (2019). 'A violent crime analysis using fuzzy c-means clustering approach'. In: *ICTACT Journal on Soft Computing* 9.3, pp. 1939–1944.

Zhao, Xiangyu and Jiliang Tang (2017). 'Modeling temporal-spatial correlations for crime prediction'. In: *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*, pp. 497–506.