Criminal behavior analysis and segmentation

Problem statement

Crime analysis is a systematic way of detecting and investigating patterns and trends in crime. Manual analysis of all crime cases is a cumbersome, time-consuming process prone to errors and biases. Crime prediction and criminal identification are the major problems to the police department as there are voluminous data of crime exist. The potential crime analysis can give us the meaningful information about the pattern of the crime and can help in locating or segmenting the criminals. One can group the criminals based on the pattern followed by them during crime. In this research, a novel Machine Learning based crime analysis method will be exploited for finding out the pattern of the crimes and criminals in the Indian context. Nowadays, Machine Learning based solutions are becoming popular in solving and giving much accurate results. One of the challenges in applying Machine Learning based crime and criminal pattern analysis is, applying such models for the Indian context where there is a large geographical data. Another challenge to the crime pattern analysis are the large size of crime datasets, and a potentially large collection of interesting crime patterns. New Machine Learning based analysis could help the agency quickly and consistently discover important patterns in crime occurrences and empower police officers and analysts to enhance crime resolution rate and can increase operational efficiency. On successful completion of the project it will deliver the functionality that will enable to analyze the crime dataset and produce meaningful insights.

Background

Various studies have explored patterns of crime occurrence. These patterns can be divided into two categories: spatial and temporal patterns. Despite the challenges, crime pattern analysis remains an active research area in computer vision and Machine Learning in recent years. Numerous approaches have been proposed over these years. Earlier methods of crime pattern and criminal analysis were based on data mining based algorithms to help detect the crimes patterns and speed up the process of solving crime. The drawbacks of the data mining based approaches are, they do not give better results in terms of detecting or predicting the crime and criminals. Fuzzy association rule mining has proven useful for crime analysis, and has utility for other crime-related data sets. Earlier computer systems were rule based but with the introduction of machine learning, now machine can learn from the data and can act accordingly. Several Crime prediction methods have utilized a variety of machine learning techniques, such as regression analysis, kernel density estimation (KDE) and support vector machine (SVM). It has an obvious extension to applications due to the potential for improving and stopping crime cases and detecting the criminals.

Methodology

Architecture of the crime pattern analysis is shown in fig 1.

Step 1: Data collection and dataset preparation
This will involve collection of crime dataset from available sources.

Step 2: Developing a Machine Learning based analysis and prediction model(K-Means)
The input to the model will be the crime dataset. The dataset may need to be pre-processed to suit the requirement of analysis work. In pre-processing part some non-relevant fields can be discarded to reduce the dimension of the data and structure of the data can be

improved. Finally, pre-processed data will be divided into training and testing part, training data will be used to train the machine learning model and testing data will be used to validate the model. Based on the new crime data input potential criminals based on the crime pattern can be predicted. The clusters of the criminals based on their geographical area can be created. Different cities in which crime is higher then other part of country can be visualized. For prediction Logistic Regression can be performed. It is suggested that more powerful Machine or Deep learning model should be explored and experimented to produce more accurate results.

Step 3: training and experimentation on datasets

The crime analysis model will be experimented on the large-scale datasets of Indian crime cases and dataset that will be populated based on local context as part of this project. The dataset may require preprocessing of text data to be processed.
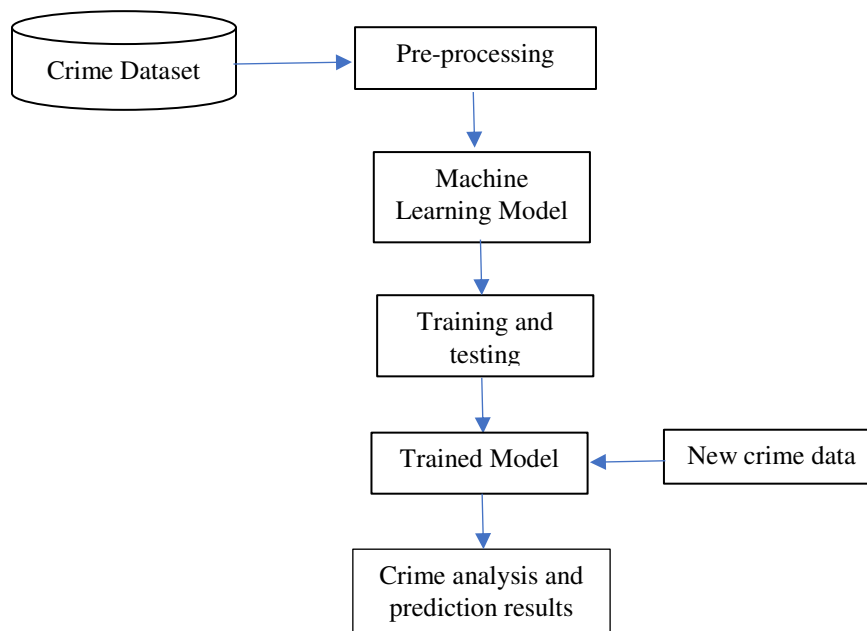

Fig 1. Architecture of K-Means clustering based Crime behavior analysis model

Experimental Design

Dataset

Indian Crime dataset are one of the mostly used dataset for the analysis of crime and criminals. It contains complete information about various aspects of crimes that happened in India since 2001. The dataset contains the State-wise data and is classified according to 40+factors. More information about the dataset is available at [4]. Another dataset [5] is based on the crime that occurred in the City of Chicago from 2001 to present.

Evaluation Measures

Measures such as accuracy, Mean Average Precision (mAP), Mean Squared Error(MSE) and Mean Absolute Error(MAE) will be calculated and compared with existing methods.

Software and Hardware Requirements

For the development and experimentation of the project, Python based Computer Vision and Machine Learning libraries will be exploited. Specifically, libraries such as Keras, TensorFlow, and other text based libraries will be used. Training will be conducted on NVIDIA GPUs for training the end-to-end version of machine learning model.

References:

1. Sivaranjani S, Sivakumari S, Aasha M. "*Crime prediction and forecasting in Tamilnadu using clustering approaches*". In Emerging Technological Trends (ICETT), International Conference on 2016 Oct 21 (pp. 1-6). IEEE.
2. Buczak AL, Gifford CM. "Fuzzy association rule mining for community crime pattern discovery". In ACM SIGKDD Workshop on Intelligence and Security Informatics 2010 Jul 25 (p. 2). ACM.
3. Nath SV. "*Crime pattern detection using data mining*". In Web intelligence and intelligent agent technology workshops, 2006. wi-iat 2006 workshops. 2006 ieee/wic/acm international conference on 2006 Dec (pp. 41-44). IEEE.
4. https://www.kaggle.com/rajanand/crime-in-india
5. https://catalog.data.gov/dataset/crimes-2001-to-present-398a4