Chicago City Crime Data Analysis (Data Mining)

Jeevith Hebbaka Mallikarjunaiah (A20382020)

Phanindra Chandraprakash (A20376416)

Information Technology Management Department

Illinois Institute of Technology, Chicago, Illinois

ITMD 525-01 (Spring 2017)

Abstract

Crimes are anti-social activities which causes society profoundly in innumerable ways. There had been humongous expansion in the crime in the recent past. It has become the most critical concern about the national security, especially after the 9/11 attacks at World Trade Center at New York city. However, the stupendous development in information and technology overwhelmingly bottlenecks the sufficient analysis of criminal and terrorist activities. Obstructing or controlling criminal activities has become a turbulence task. In the ever-ending race between law breakers and law enforcers, the job of the police officers or law enforcers in their roles to grab the criminals, these people must have to be ahead in the race. Data mining applied in this perspective gives fruitful results to assuage criminal activities. In this paper, we visualize the crime patterns and trends by using classification based models. Results from this analysis can be used to improve safety measures for the forth coming years.

Keywords: *crime patterns, classification, data mining, law enforcers*

Introduction

Determining solution for crimes has been the prerequisite of the law enforcement specialists. With the advent of computerized systems, it has become more convenient to track the data of criminals, which will speed up the process of solving crimes. The main objective of this project is to perform data mining operation on Chicago City Crime Data to detect crime pattern and suggest relevant safety measures depending upon the trend and pattern of the crime which has occurred over a period. More precisely, we will use clustering based models to help in identification of crime pattern. We will be predicting future crimes based on the training data set.

Data Sets

The data sets are taken from "City of Chicago – Data Portal". This data set reflects the criminal activities reported in the city of Chicago from 2001 to present date. This work will be on Crime Analysis domain and this data set has around 62 lakhs of records. This data set has 22 features and from that we will be choosing only 3 features to perform data mining to predict the crime rates or criminal activities for next five years.

Research Problems

The research problem is to identify patterns of change in crime and location through a fixed timeline to be able to predict crime rates and locations in the future. For this, batches of years have to be used as training sets to note after batches of how many years the patterns change. The challenge is determining the batch size and heterogeneity in the change across different time periods. The aim is identifying patterns and changes in them during the period 2001-present and predict crimes and locations of crime.

- Date
- Primary Type
- Location Description

Potential Solutions

For time being we are considering KNN and Naïve Bayes classifiers and thinking of using batches of 5 years as interval, other solutions can be varying the interval and trying other classification algorithms. Plan to perform experiments include varying the parameters of the model like what 'K' value to use and checking if one single interval of change is consistently observed or where there are significant periods of crime before activity and successive periods of activity in the same period and a way to deal with this heterogeneity.

Evaluation

For each algorithm use K Fold cross validation and use metrics such as accuracy, precision recall and Area under the ROC curve to check the performance of the model. Other evaluation considerations include time and space efficiency in terms of memory used and potential for a distributed setup.

Expected Outcomes

The expected outcome will be the solution to prevent or control crime based on the prediction made through KNN and Naïve Bayes classifiers. More details will be provided on this while preparing final project report.