CSE6242 - Data and Visual Analytics
Final Report
Team 30: Crime Prediction

Kuhu Gupta[1], Preeti Maan[2], Vatsal Srivastava[3]
{kgupta307[1], pmaan6[2], vsrivastava43[3]}@gatech.edu

Introduction

Crime is everywhere and every day the law enforcement works diligently to prevent and handle criminal incidents by effectively utilizing their human, infrastructure and technical resources. Historically the police have used their experience and intuition to predict broad locations and times of crimes; now they use more sophisticated technology like mining historical crime records, criminal histories, modern equipment, but the approach is still more focused on reactive than preventive measures. In the past few years, the focus has shifted to preventive measures, mainly crime prediction by highlighting what factors lead to or facilitate a crime.

This has motivated us to analyze crime data, find correlated factors and work toward crime prediction. Much work is happening in the field of crime prediction that we are using to inform our approach.

Problem Definition

We aim to build a visual analytics tool to assist law enforcement by predicting the likelihood of different types of crime in a given geographical area in the city of Chicago.

We also want to identify a specific set of features that can be substantially responsible for increased crime instances (like a street with many abandoned houses, graffiti, etc.) so that steps can be taken to rectify them in order to reduce occurrences of crimes.

Survey

The field of crime prediction is not new but with more technological advancements, research and interest have gained momentum in this field in recent years.

Almanie, Mirza and Lor (2015) use city crime history dataset that includes location, type of crime, date, time and proposes finding temporal and spatial crime hotspots using Decision Trees and Naive Bayesian classifiers in order to predict potential crime types. While Corman and Mocan(2005) similarly proposed prediction using spatial analysis using k-nearest neighbour approach. Although it suggests overcoming limitation decision trees as models but its limited in terms of source data and is purely based on historical crime records.

Wallace and Schalliol (2015) measured the relationship between abandoned buildings and social disorder in a community. Their assumptions were based on the Broken Windows theory. While Corman and Mocan (2005) investigate the impact of economic, demographic conditions and

increased felony arrests on crime rates in New York City. It also suggests that the Broken Windows Theory has validity in the case of robbery and motor vehicle theft.

Hauk et al (2002) have used the concept space algorithm on crime data to detect abnormal activities. Once these activities are identified it may be possible to predict the next occurrence of such activity. Oatley et al (2004) have used algorithms to match and link burglary crimes together into a crime series for predicting where the next crime in that series will occur. Wang et al (2012) perform automatic semantic analysis on Twitter Posts along with dimensionality reduction via Latent Dirichlet Allocation and prediction via linear modelling for forecasting hit-and-run crimes.

Dash et al (2018) fuse the spatiotemporal information about criminal records like schools, libraries, police stations, and 311 service calls using network analytic approach to identify observations that improve the quality of prediction. They use a regression-based approach by experimenting with polynomial, auto-regressive and support vector regression methods.

According to Johnson et al (1997), deprived regions are more affected by repeat victimization and in regions with high crime event rate, and in support Johnson et al. (1997) and Townsley et al. (2000) designated as hot-spots. Several researchers like Winkel (1991); Gill and Matthews, (1994); Ericsson, (1995) Johnson and Bowers, (2004) in their research work have interviewed offenders and established that they target the same property repeatedly suggests that burglary offences cluster in space and time.

The correlation of large-scale “Point-Of-Interest” data and “taxi flow data that serve as hyperlinks” in the city of Chicago, IL is explored by Wang and Kifer (2016) to be harnessed in crime rate estimation and reduce prediction error. Wang and Kifer (2016) give us important insight to focus on seemingly trivial information to realize novel factors that might contribute to crime.

Proposed method

User Interface-

The user interface of the application encompasses a dashboard view with a calendar view date selection and digital clock view time selection. Along with it, the user could see the results for different types of crimes by selecting from a drop-down.

The major part of the dashboard is a choropleth map. A choropleth map is a thematic map in which areas are shaded or patterned in proportion to the measurement of the statistical variable being displayed on the map, such as crime probability in this scenario.

It provides an easy way to visualize how a probability varies across the wards of Chicago using colour saturation. The colour used for the choropleth are chosen sequential color schemes

derived from Cynthia A. Brewer's ColorBrewer. Since ColorBrewer publishes only discrete colour schemes, the sequential scales are interpolated using uniform B-splines.

Implementing a custom designed map by Mapbox using Leaflet.js Library, users can access the ward and street-level perspective of crime occurrences. Employing the principles of Ben Shneiderman Information Visualization mantra- "Overview first, zoom and filter, then details on demand".

Considering "details-on-demand", the user could see more information in form of a heat map about a ward by hovering the mouse on it. A heat map is a graphical representation of data where the individual values contained in a matrix are represented as colours. In this scenario, Heatmap is used to represent the correlation between different types of 311 requests and crime types in each ward.

Description

CrimeMapper is a web-based map application for predicting and exploring spatial-temporal patterns of different types of violent crimes and 311 requests in the City of Chicago. Utilizing this geo-visual analytic tool, analysts can uncover detailed spatiotemporal patterns of Chicago crimes at the ward level as well as the correlation of 311 requests, extending the value of the data served from the existing website (https://data.cityofchicago.org/).

The client-server application predicts the probability of crime incidents on a base map using Mapbox Tile Layers with Leaflet.js with supporting visualization in d3.js. For prediction the user can choose by time, (i.e., date, month and year from the calendar filter and time from the clock filter) and attribute (i.e. type of crime)

Installation and Execution-

1. In order to run the application, the system is required to have "Python 3.5 or above"
2. We need to install some package for this project. So, let's install this on the dev-env virtual environment we just created in the previous step by –

pip install -r requirements.txt

3. It's time to run the application. You just need to run the following code into your console

python app.py

4. Visit 'http://127.0.0.1:5000' in your web browser.