Big Data Domain Research

In Law Enforcement

Abstract:

Big data is helping law enforcement in a big way. Law enforcement uses big data for recognizing both the criminal offense and outlaw. In the 19th Century, Italian physician Cesare Lombroso won fame for claiming to be able to identify criminals by their faces. If only it were that easy to spot the bad guys. Crime has patterns just like everything else humans do when we're seen as a large enough group. Individual behavior is hard to predict, but determining the average behavior of a group and then matching individuals to that template to determine “fit” can be accurate. We can achieve this using big data to determine patterns of crime and criminal. In this paper I propose how big data is used in the Law Enforcement Domain to determine a pattern and how data is being collected, analyzed and challenges faced while processing data in this Domain.

Introduction:

Police authorities are using Big Data to find patterns and build connections so they can prevent and solve crime. Crime analysis can help police leaders better understand underlying economic and demographic factors for a given area that might help explain why trends occur and appropriate responses can then be crafted. Big data can find anomalous patterns in behavior and solve financial fraud. Durham police department has shut down a “cash for crash” scam which involves an organized crime group defrauding insurance companies by calming several times for same accident. Durham (N.C.) police department’s analytical services group used IBM analytics technology and reduced violent crime by more than 50 percent during a four-year span in a two-square-mile region. Now let’s take a case study and discuss about the data sources, data collections, preparation of the data and initial analysis.

Case Study:

Problem:

Police officers sometimes feel like they’re playing Whack-A-Mole, they work to decrease illegal activity in a high-crime area, only to have it pop up somewhere else or come back the moment they leave. We need to figure out a way that produces long-term positive outcomes.

Solution:

Joel Caplan, an assistant professor in Rutgers University’s School of Criminal Justice partnered with two Rutgers colleagues, professors Eric Piza and Leslie Kennedy worked on this problem and developed a risk terrain modeling. To solve this problem they followed an approach that takes an area and blends its history of crime with data on local behavioral, physical and demographic characteristics to create a map of locations with the greatest crime risk. Risk terrain modeling considers and examines the local behavioral, physical and demographic characteristics that contribute to crime and helps police officers predict where new hot spots could arise. It shows on map the underlying features of the environment that are attractive for certain types of illegal behavior, and in doing so, we’re able to assign probabilities of crime occurring. For example, while doing this with the risk terrain map for shootings in Irvington, they could quickly noticed that the environmental context of lower-risk places often included churches or had street intersections with traffic lights nearby. Traffic lights might represent places with higher-than-normal volumes of people/traffic. These mitigating features of the environment help police to control crime in that region and gives a long term positive outcome.

Data sources:

The major sources of crime data in the United States are the Uniform Crime Reports (UCR), the National Incident-Based Reporting System (NIBRS) and the national survey of crime victimization is another way in which criminal activity is measured, but it focuses on the experiences of victims.

Data collections:

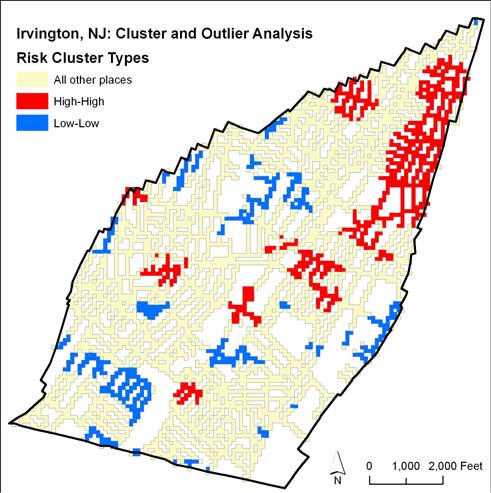
Big data flow in from gunshot sensors, surveillance video, social media, Images, blood spatters and the like. The use of mobile phone data and demographic data can also be used to predict crime geographically

Problem Statement:

I have a dataset that reflects reported incidents of crime that occurred in the City of Chicago from 2001 to present. The Data is extracted from the Chicago Police Department. I will analyze and spot the hot spot regions of crime on the map. Addition to spotting hot spots I will analyze what demographic factors acted as catalyst for crime rate at that particular region. As of now I am still in search for demographical data of Chicago city.

Preparation of the data and initial analysis:

After performing analysis on crime and demographic data, Caplan and his team has come up with a map showing crime hot spots on risk terrine mode of Irvington city. These maps help police to control crime in that region and gives a long term positive outcome.



Software and Algorithms used for analysis:

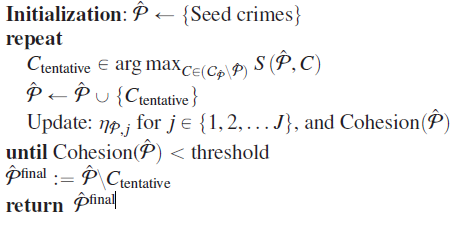
Risk Terrain Modeling:

Risk terrain modeling is used for crime analysis and other spatial risk analysis (SRA) that utilizes a demographic information system to attribute qualities of the real world to places on a digitized map. RTMDx (Risk Terrain Modeling Diagnostics) is a software that helps to identify the spatial attractors and generators of crime incidents. RTMDx automates the steps of RTM by Selecting an outcome event, Choosing a study area and a time period, Obtaining base maps, Identifying all possible risk factors. RTMDx performs conjunctive analysis which involves combination of factors appear in a local study area. This analysis can then be used to assign a description to an area that can be compared to similar areas in a location. For example, setting "bars close to schools", shows a high rates of crime. Now we identify the locations of these combinations and determine how their particular characters are conducive to crime.

Series Finder Algorithm:

Series Finder is a pattern detection algorithm that grows a pattern of discovered crimes from within a database, starting from a seed of a few crimes. a series or pattern crime are committed by the same individual or group. The algorithm it searches through the database looking for similarities between crimes in a growing pattern and in the rest of the database, and tries to identify the modus operandi (Attributes of crime Ex Operating time or day, tools used, targeting race or age ..etc). Series Finder looks both the common characteristics of all patterns and the unique aspects of each specific pattern. As the pattern grows the modus operandi (M.O.) becomes well defined.

We use P = { C1, C2, C3,…Cp} to denote a true pattern of crime, where each of the Ci’s represents a crime. Only a seed of a few crimes from P are known. Series Finder uses the seed to grow a set of discovered crimes P, in hopes that P will eventually be similar to the underlying (and unknown) set P. The pairwise similarity ϒ measures how similar crimes Ci and Ck are in a pattern set P. We model it in the following form: ϒp(Ci, Ck) = λj ηp,j Sj(Ci, Ck) where { λj}j are “pattern-general” weights, and { ηp,j }j are “pattern-specific” weights and Sj is the similarity measure for the jth attribute. High ϒp(Ci, Ck) indicates two crimes Ci and Ck have similar attributes that fit in the crime pattern. A crime fits into a patter if it has ether high crime-crime similarity or pattern-crime similarity. The series finder algorithm is as shown below.



IBM i2 COPLINK:

IBM i2 COPLINK is scalable and flexible and supports various needs of law enforcement IBM i2 COPLINK enables nationwide information sharing across departments, agencies and states and enables faster searches. Through COPLINK software we can connect to other COPLINK repositories. With the help of COPLINK software police officer can quickly search the database for suspects with their partial names and aliases. COPLINK provides various interfaces through which police officers can perform various crime analytics such as face reorganization, Computer Statistics( detecting patters) and Incident Analyzer. COPLINK Streamlines investigations by automating repetitive searches. COPLINK Active Agent notifies users via system alert or email whenever new information is available or when others are conducting similar searches.

Challenges:

In law enforcement domain Big data flows in numerous forms like gunshot sensors, surveillance video, social media, Images, blood spatters and the like make it difficult to process, cleanse and analyze data. Vast dataset made up of still images, surveillance video, HTML and rich text make it impossible to interpret data without right tool. Sexual abuse of children which has been inherently multination ally serious crime, involves gigabytes and sometimes terabytes of data when they arrest or raid someone who runs a service that trades in images of child sex abuse. The sheer amount of data makes it difficult to process. It makes it even more difficult when is data is traded in network. In this situation Police officials need to find the root of crime in order to demolish this network. The situation is complicated by the fact that images and videos of abuse are widely traded. Without big data analytics police officers could spend a lot of time literally retracing the steps of other forces that have already worked out who was behind one set of images or who they depict.

References:

1. <http://www.policeone.com/police-products/software/Data-Information-Sharing-Software/articles/6396543-How-Big-Data-is-helping-law-enforcement/>
2. <http://www.govtech.com/public-safety/Predicting-Crime-Using-Analytics-and-Big-Data.html>
3. <http://www.innovative-analytics.com/analytics-big-data-promise-challenge-fighting-preventing-crime/>
4. <https://datafloq.com/read/los-angeles-police-department-predicts-fights-crim/279>
5. <http://theinstitute.ieee.org/technology-focus/technology-topic/the-future-of-crime-prevention>
6. <http://www.bbc.com/news/business-26520013>
7. <https://ored.rutgers.edu/content/predicting-crime-big-data-affordably>
8. <http://www.forbes.com/sites/howardbaldwin/2014/11/17/big-data-as-superhero-fighting-crime-and-boosting-security/>
9. <http://www.geekwire.com/2014/big-data-meets-crime-fighting-seattle-police-launch-seastat-quickly-pinpoint-crime-hotspots/>
10. [riskterrainmodeling.com/](http://www.riskterrainmodeling.com/)
11. Detecting Patterns of Crime with Series Finder By Tong Wang, Cynthia Rudin, Daniel Wagner and Rich Sevieri
12. <http://www-03.ibm.com/software/products/en/coplink>