

Crime prediction using administrative big data and machine learning

Gyeongseok Oh

Sam Houston State University, USA

It is indisputable that machine learning techniques and large data analysis became the most topics in most discipline of science and industry during the past decade. Concurrently, numerous governments within the world are collecting enough amounts of administrative data which will be analyzed by machine learning techniques to research the causes of social phenomena and to enhance the efficiency of public administration. Despite the info analytic techniques and therefore the capability of knowledge storage are remarkably improved, an outsized number of students within the field of science hold conservative perspective on applying machine learning and large data analysis to explaining social phenomena. The goal of this study is to fill the void by providing empirical evidence. This study will plan to examine the validity of using administrative big data to predict crime incidents. Records of involves service through 311 mayor's hotline system in Houston, Texas and therefore the reform the official crime reports of Houston local department were examined to assess whether signs of physical decay and the presence of social nuisance predict the crime incidents at neighborhood level. The results of this study will corroborate the Broken Windows Theory and present new windows to explore the causes of crime. Several policy implications for state and police administrators are going to be developed and discussed.

The Federal Bureau of Investigation (FBI) defines a violent crime as an offense which involves force or threat. The FBI's Uniform Crime Reporting (UCR) program categorizes these offenses into four categories: murder, forcible rape, robbery, and assault. The FBI UCR program defines each of the offenses as follows: (i) Murder - The willful (non-negligent) killing of 1 person by another. The UCR doesn't include deaths caused accidentally, suicide, negligence, justifiable homicides and attempts to murder or assaults to murder (which are scored as aggravated assaults), during this offense classification. (ii) Forcible Rape - Rape may be a sexual attack on a female against her will. Though attempts or assaults to commit rape by threat or force are considered crime under this category, carnal abuse (without force) and other sex offenses are excluded. (iii)

Robbery - The taking or attempting to require anything useful from the care, custody, or control of an individual or persons by force or threat of force or violence and/or by putting the victim in fear. (iv) Assault - it's the unlawful attack conducted by one person upon another to inflict severe or aggravated bodily injury. The UCR program specifies that an assault usually involves the utilization of a weapon or other means to supply death or great bodily harm. Attempted aggravated assaults that involves the utilization of guns, knives and other weapons are considered to belong to the present category because if the assault were completed, it might have led to serious personal injury. An offense that involves both assault and larceny-theft occurring together, the offense is taken into account to belong to the category of robbery. Unfortunately, these sort of crimes seem to possess become common place within the society. Enforcement officials have turned to data processing and machine learning to assist within the fight of crime prevention and enforcement . during this research, we implemented the rectilinear regression , Additive Regression, and Decision Stump algorithms using an equivalent finite set of features, on the communities and crime un normalized dataset to conduct a comparative study between the violent crime patterns from this particular dataset and actual crime statistical data for the state of Mississippi that has been provided by neighborhoodscout.com. The crime statistics used from this site is data that has been provided by the FBI and had been collected for the year 2013. a number of the statistical data that was provided by neighborhoodscout.com like the population of Mississippi, population distribution by age, number of violent crimes committed, and therefore the rate of these crimes per 100K people within the population also are features that are incorporated into the test data to conduct analysis.