EECS 349 Project Proposal

Predicting Crimes In Chicago From Weather

Alan Fu 2015-04-08

In this solo project I will investigate the relationship between crimes and local weather in Chicago City by building and training a machine learning program. The motivation for this project emerges from my experience with Chicago's weather since my attendance at Northwestern. I have noticed significance effect of weather on my mood and behavior, so have my close friends in Northwestern. I wish to further explore this effect in a more quantified and rigorous manner, and I decided to use crimes as the specific measurement of human behavior (specifically, how violent human turn to each other). I choose Chicago as the location of study because of its persistently high crime rate and its volatile (and sometimes extreme) weather. The input of my program will be the daily weather summary report of Chicago and the crime archive of Chicago in the last 15 years, and the output will be a trained algorithm that takes in weather parameters and output predicted crime parameters (the specification of the parameters will be discussed later).

Important applications may arise if the program can predict crimes in Chicago in reasonable accuracy from weather. The effect of weather on crime can be a valuable addition to the existing crime-predicting tools such as *PREDPOL*[1] to achieve greater accuracy in predicting crimes, enabling law enforcement agencies to stop crimes better, faster, or even before they happen. The prediction made by the program can also be used in other psychological or sociological studies on weather and human behaviors to gain further insights.

The crime data I will be using is from *City of Chicago Data Portal*[2], and the weather data comes from *Iowa Environmental Mesonet*[3]. I have already obtained both data. First the irrelevant parameters in the data (such as "arrested" in crime data and station name in weather data) will be discarded, and the rest will be stored in appropriate data structures in Python. The data will then be randomly divided into training set (70%) and testing set (30%). My first goal is to predict the number of crimes in a day given the average temperature in that day. Then, I will add precipitation, humidity and wind speed parameters into weather input and update the model. If time permits, I will also look into how those weather parameters affect the frequencies of different types of crimes committed. For now, I plan to use decision tree to build the model (although our input and output has continuous parameters, they can easily be categorized). At the end, the accuracy of the model will be tested on the 30% testing set of weather and crime data.

---

[1] https://www.predpol.com/technology/

[2] https://data.cityofchicago.org/Public-Safety/Crimes-2001-to-present/ijzp-q8t2

[3] http://mesonet.agron.iastate.edu/request/download.phtml?network=IL_ASOS