Timmy Huynh, Pennsylvania State University
Brian Swedberg, Pennsylvania State University
PAA 2015 Submission

Crime Prediction in Chicago: Using a Statistical Learning Approach to Create a Spatiotemporal Vulnerability Surface of Violent Crime

Short Abstract

Much research has analyzed spatial, demographic, or weather-related explanators for violent crime in cities. In our analysis, we combined all these factors into a single time-dependent surface of the likelihood of violent crime occurrences in the Near West Side neighborhood in the city of Chicago. Since we are able to spatially link together data from multiple sources (e.g., City of Chicago, U.S. Census, and NOAA), each data point in our analysis contains information for all our predictor variables. However, because our independent variable (violent crime occurrence) happens at a fixed point and is spatially-sparser than our other data, we used a distance-decay formula to interpolate the violent crimes over our surface. We then trained and tested a random forest classifier to create a final violent crime vulnerability surface over our selected neighborhood. Our accuracy of violent crime prediction proved to be fairly good for a social science problem.

Keywords: crime, spatial demography, GIS, statistical learning

Extended Abstract

Homicide rates in Chicago reached an all-time low in October 2014, after having decreased for eight consecutive quarters (Gorner, 1 October 2014). However, despite these drops, reported shootings have increased by 6% from the previous year. These conflicting statistics showcase the complex nature of violent crime rates, and therefore, we aim to begin to understand the explanators of this tragic phenomenon. In particular, our project is interested in predicting the likelihood of all violent crimes given certain spatiotemporal conditions. Using historical violent crime data, we created several spatiotemporally-sensitive surfaces as input, in addition to the inclusion of other empirically-informed weather and demographic variables as input, for the predictive capabilities of a random forest classification. Whereas the term "violent crime" is generally loosely-defined, for our research, we used the FBI Uniform Crime Reporting (UCR) Program's (2012) definition of violent crime. The UCR Program defines violent crime as an act that involves force or the threat of force. Furthermore, they delineate four types of violent crime: homicide (first and second degree), criminal sexual assault, robbery, and aggravated assault or battery.

The main benefit in being able to predict crime is to increase public safety. If we can predict where crime is likely to occur with data mining techniques, police or investigators can better allocate their time and resources by focusing on those particular areas (Chen et al., 2004). However, this is a challenging task because spatiotemporal data add additional complexity. More specifically, conditions such as weather, demographics, and other criminal activity are constantly changing throughout the city landscape. Our goal is to provide a vulnerability surface of violent crime in which a decision-making official could task police officers in concentrations comparable to crime likelihoods. Our project may then be a prototype for a violent crime prediction model that could take into account new violent crime occurrences, weather events, and updated demographic data in real-time.

Certainly, there is a vast amount of research on crime and correlated variables. However, current research does not examine a holistic, systems-based approach for crime analysis. Criminological research has typically focused only on effects of weather on crime rates, only on socio-demographic predictors of crime, or only on the spatial patterns of crime. Likewise, data mining research exploring social phenomena typically does not take into account the variables that social science research has found to be important. Therefore, our project faces the challenge of combining multiple sources and dimensions of data, adding nuanced complexity to the interpretability of the end result.

For our application of a random forest classification, we first defined the features, basing feature inclusion on past research. The dependent variable was created from a binary classifier of the incidence of violent crime occurrences. As prior research has explored the effects of singular spatio-temporal attributes (namely, weather and sociodemographic variables), we similarly brought in comparable predictor variables as controls for our predictions of crime in order to have higher accuracy and to account for the other factors that may lead to future violent crime occurrences. For example, research on weather and crime has found positive correlations between heat and violent crimes, while other criminological studies have found higher rates of crimes within marginalized groups (e.g., people in poverty or in racial minority groups). Spatially, these high crime rates are also clustered in neighborhoods reflective of these aforementioned variables, as prior research has shown. Temporally, crime also parallels different scales of periodicity, such as peaks on weekends or peaks in the summer. These predictor variables were then linked to our dependent variable of violent crime incidence: demographic variables were spatially joined to violent crime occurrences, weather variables were temporally linked, and an intermediate analysis was performed for the spatiotemporal distance of each crime occurrence to other crimes within a set spatial and temporal distance. Based on our resulting accuracy scores, our prediction rates (upwards of 60 percent) are fairly good for social science research. Our analysis of feature importance found that the time-based weather features ranked higher than the spatially-oriented demographic features, while the spatiotemporal distances to other crimes ranked as the most important.

By merging these three types of data (i.e., weather, demographics, and spatiotemporal distance to other crime), we were able to expose a more holistic view of Chicago crime. Through a random forest classification, we plotted a spatio-temporal prediction surface of violent crime. Ultimately, we hope that our methods will be used to better allocate police forces throughout the city (Lauritsen et al., 2014) and serve to better describe the urban landscape in terms of crime.

Selected References

Chainey, S., Tompson, L., & Uhlig, S. (2008). The utility of hotspot mapping for predicting spatial patterns of crime. *Security Journal*, 21(1), 4-28.

Chen, H., W. Chung, W., J. J. Xu, G. Wang, Y. Qin, and M. Chau. 2004. "Crime Data Mining: A General Framework and Some Examples." *Computer* Vol. 37, Issue 4: 50-56.

Chicago Data Portal, City of. 2014. City of Chicago. Retrieved September 5, 2014, from https://data.cityofchicago.org/.

Eck, J., Chainey, S., Cameron, J., & Wilson, R. (2005). Mapping crime: Understanding hotspots.

Federal Bureau of Investigation. (2012, January 1). *Crime in the United States 2012: Violent Crime.* Retrieved October 2, 2014, from http://www.fbi.gov/about-us/cjis/ucr/crime-in-the-u.s/2012/crime-in-the-u.s.-2012/violent-crime/violent-crime

Gorner, J. (2014, October 1). 2014 Chicago homicides on decline, but shootings on rise. Retrieved October 2, 2014, from http://www.chicagotribune.com/news/ct-chicago-homicides-shootings-met-20141001-story.html

Huang, B., Wu, B., & Barry, M. (2010). Geographically and temporally weighted regression for modeling spatio-temporal variation in house prices. *International Journal of Geographical Information Science*, *24*(3), 383-401.

Lauritsen, J. L., White, N., Bureau of Justice Statistics, US Dept of Justice, Office of Justice Programs, & United States of America. (2014). Seasonal Patterns in Criminal Victimization Trends.

Pellow, D. N., and Brehm, H. N. July 2013. An Environmental Sociology for the Twenty-First Century. *Annual Review of Sociology* Vol. 39: 229-250

South, Scott J., and Steven F. Messner. 2000. Crime and Demography: Multiple Linkages, Reciprocal Relations. *Annual Review of Sociology* Vol. 26: 83-106.

Townsley, M., Homel, R., Chaseling, J. 2003. "Infectious burglaries. A test of the near repeat hypothesis." *British Journal of Criminology* 43(3):615–633.