Predict Crime: A comparative study between Chicago and Seattle using spatial and temporal approach

Name: Yangyang Dai

Research question:

How to predict crime using spatial and temporal approach in machine learning and how the results of this method differ between city Chicago and Seattle.

Literature Review

Past work:

There have been a lot of studies, experiments and apps done related to crime predictions both in the U.S and internationally in the last several decades.

Most of the past prediction forecasts cover the range from a decrease to an increase in crime and they are generally based on differing assumptions and the margins of error inherent in mathematical modeling (Schneider, 2002). For example, the mathematical model of Dhiri et al. (1999) offers substantially differing crime predictions. At one extreme lies their prediction that the number of recorded burglaries and thefts in 1999 and 2000 will increase by approximately 40 percent when compared to 1997. Alternatively, they project that burglary might fall below the 1997 level. Abrahamse (1996) projects homicide arrest rates in California until 2021 using pessimistic, nominal, and optimistic assumptions. Under the pessimistic assumption, by 2021 homicide arrest rates will nearly double the 1994 rate; under the nominal assumption, homicide arrest rates will be about 28 percent higher in 2021 than in 1994; and under the optimistic assumption, homicide arrest rates in 2021 will be about 14 percent below 1994 levels.

Apart from the crime rates prediction, large datasets have also been used to analyze information such as location and the type of crimes and helped people follow law enforcements by Almanie, Tahani, Rsha Mirza, and Elizabeth Lor (2015). Existing methods have used these databases to identify crime hotspots based on locations. However, even though crime locations have been identified, there is no information available that includes the crime occurrence date and time along with techniques that can accurately predict what crimes will occur in the future (2015). On the other hand, the previous related work and their existing methods mainly identify crime hotspots based on the location of high crime density without considering either the crime type or the crime occurrence date and time (2015). For example, related research work containing a dataset for the city of Philadelphia with crime information from year 1991 - 1999. It was focusing on the existence of multi-scale complex relationships between both space and time(2015). Another research titled “The utility of hotspot mapping for predicting spatial patterns of crime” looks at the different crime types to see if they differ in their prediction abilities (2015). Other existing works explore relationships between the criminal activity and the socio-economics variables such as education, ethnicity, income level, and unemployment (2015).

The newly utilized methodology by Almanie, Tahani, Rsha Mirza, and Elizabeth Lor (2015) focusd on using spatial and temporal approach to compare and predict the crime patterns of Denver, Colorado and Los Angeles, California by discovering both hotspots and crime types at a specific time and location.

Predictive methods and technologies:

Apriori is one of the basic algorithms for mining frequent patterns. It scans the dataset to collect all item sets that satisfy a predefined minimum support. The goal of using this model is to find all possible crime frequent patterns regardless of the committed crime type. And the algorithm is implemented on location and time features and excluded the crime type feature. Additionally, to obtain more frequent patterns, constraint-based mining is applied by restricting the extraction process on the frequent patterns having this formula of three specific item sets (Location, Day, Time) (2012).

Multinomial Naïve Bayes is used for multinomial distributed data that conforms to the categorical features in our datasets. The crime features contain (month, day, time, location) of the crime while the crime type is selected to represent the class label. And then the same classifier is trained on the training data for each of Denver and Los Angeles datasets to obtain two different models ready for crime type prediction in each of the two cities (2015).

Decision Tree classifier is another supervised learning algorithm to create a model to predict the class label values by learning simple decision rules implied from the data features.

Point Process Model is a new method being employed by Brown (2001) and his colleagues based on the theory of point patterns and multivariate density estimation, and can best be described as a point process model (Brown, 2001). The modeling is akin to neural networks in that there is training involved, and past data are used to predict future events. In essence, this approach glues multivariate models together and uses notions from kriging and density estimation (Brown, 2001).

Artificial Neural Networks: In essence, the network is trained by feeding it past data and adjusting the weights assigned to the input units; when the network is processed, the error signal is fed back, or "backpropogated" through the network to adjust the weights until, ultimately, the error signals are minimized (Olligschlaeger, 1997). GIS was employed in conjunction with this model in order to process spatial and temporal data, including data aggregation and determination of spatial and temporal lags. This was accomplished by overlaying a grid and summing the data points that fall into each cell, as well as employing contiguity measures.

My contributions:

Since past spatial and temporal studies are mainly focused on Los Angeles and Denver, I would like to explore more about cities in the mid-west and west coast such as Chicago and Seattle. The potential findings could add on to the understanding of the differences of crime patterns in those two regions in the U.S. And furthermore, more similar analysis could be done on other cities in other regions as well.

Secondly, the past works are mainly building models on one dataset and apply on another city’s dataset, hence the accuracy is not very good. I am aiming to build different models on different datasets for Chicago and Seattle (doing training within the datasets and test the accuracy of the model) because each city’s demographic characteristics are very different.

The last but not least, I would like to employ more machine learning and classification algorithms to predict the crime type and test which one is the best based on past studies.

Difficulties:

Since quantitative models suffer from the inability to anticipate unpredictable future events, including unforeseen technological advances, economic vicissitudes, social trends, and advances made in law enforcement and private security technology (2002), we do not expect perfectly accurate models and results.

References:

Almanie, T., Mirza, R., & Lor, E. (2015). Crime prediction based on crime types and using spatial and temporal criminal hotspots. *arXiv preprint arXiv:1508.02050*.

Groff, Elizabeth R., and Nancy G. La Vigne. "Forecasting the future of predictive crime mapping." *Crime Prevention Studies*13 (2002): 29-58.

Adams-Fuller, T. (2001). "Historical Homicide Hot Spots: The Case of Three Cities." Doctoral dissertation, Howard University, W ashington, DC.

Anderson, D., S. Chenery and K. Pease (1995). *Biting Back: Tackling Re- peat Burglary and Car Crime.* (Home Office Crime Prevention Unit Series, No. 58.) London, UK: Home Office.

Anselin, L. (1999). Spatial Data Analysis with SpaceStatTM and ArcView®. Ann Arbor, MI: SpaceStat (http://www.spacestat.com).

Sivan Aldor-Noiman, Lawrence D Brown, Emily B Fox, and Robert A Stine. Spatio- temporal low count processes with application to violent crime events. arXiv preprint arXiv:1304.5642, 2013. 1, 5

Luc Anselin, Jacqueline Cohen, David Cook, Wilpen Gorr, and George Tita. Spatial anal- yses of crime. Criminal justice, 4:213–262, 2000. 2.1

Jeremy Atherton, University of Chicago Library Map Collection, and Christopher Sicil- iano. Community areas of Chicago, Illinois. 2011. http://en.wikipedia.org/wiki/File:US- IL-Chicago-CA.svg. 5

Adrian Baddeley and Rolf Turner. Spatstat: an r package for analyzing spatial point pat- terns. Journal of statistical software, 12(6):1–42, 2005. 2.1

Joel M Caplan, Leslie W Kennedy, and Joel Miller. Risk terrain modeling: brokering criminological theory and gis methods for crime forecasting. Justice Quarterly, 28(2): 360–381, 2011. 1

Jacqueline Cohen, Wilpen L. Gorr, and Andreas M. Olligschlaeger. Leading indicators and spatial interactions: A crime-forecasting model for proactive police deployment. Ge- ographical Analysis, 39(1):105–127, 2007. ISSN 1538-4632. doi: 10.1111/j.1538-4632. 2006.00697.x. URL http://dx.doi.org/10.1111/j.1538-4632.2006.00697.x. 1

Noel Cressie. The origins of kriging. Mathematical Geology, 22(3):239–252, 1990. 2.5

Noel Cressie and Hsin-Cheng Huang. Classes of nonseparable, spatio-temporal stationary covariance functions. Journal of the American Statistical Association, 94(448):1330–1339, 1999. 1

Peter Diggle. A kernel method for smoothing point process data. Applied Statistics, pages 138–147, 1985. 2.1

David Duvenaud, James Robert Lloyd, Roger Grosse, Joshua B Tenenbaum, and Zoubin Ghahramani. Structure discovery in nonparametric regression through compositional kernel search. arXiv preprint arXiv:1302.4922, 2013. 2.4