Predict Crime Type: A Comparative Study between Chicago and Seattle

Yangyang Dai

University of Chicago
daiy@uchicago.edu

Jun 6, 2018

Abstract

This project focuses on predicting crime types using different machine learning methods. It analyses two different real-world crimes datasets for Chicago, IL and Seattle, WA and provides a comparison between the two datasets. The project uses Random Forest, Gradient Boosted Tree methods and Neural Network Multiclass Classifier to predict potential crime types in both cities. To further analyze crimes' datasets, the project also applies GIS tools to explore the spatial and temporal aspects of the data.

KEYWORDS

Crime type prediction, crime classification, spatial patterns, Chicago and Seattle

1. INTRODUCTION

This paper aims to explore the characteristics of crime patterns for cities in the mid-west and west coast such as Chicago and Seattle. It uses different machine learning models to predict the possibility of different crime types that will could occur on a given area and a given time frame.

The potential findings could add on to the understanding of the differences of crime patterns in those two regions in the U.S. More similar analysis could be done on other cities in other regions as well. Furthermore, the past works mainly build models on one dataset and apply on another city's dataset, hence the accuracy is not very good. In this project, different models on different datasets for Chicago and Seattle will be used, which is doing training and testing within the two datasets and evaluate the accuracy of different models, because each city's demographic characteristics are very different. The last but not least, this project employs different machine learning and classification algorithms to predict the crime type and test which one is the best based on past studies.

2. RELATED WORK

There have been a lot of studies, experiments, and computer and phone applications 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. 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) focuses 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.

3. DATASETS

3.1 Chicago crime data:

The first Chicago crime data set is extracted from the Chicago Police Department's CLEAR - Citizen Law Enforcement Analysis and Reporting system from https://data.cityofchicago.org/Public- Safety/Crimes-2001-to-present/ijzp-q8t2. The data is recorded from 2001 to 2018 with 22 attributes including 'ID', 'Case Number', 'Date', 'Block', 'IUCR', 'Primary Type', 'Description', 'Location Description', 'Arrest', 'Domestic', 'Beat', 'District', 'Ward', 'Community Area', 'FBI Code', 'X Coordinate', 'Y Coordinate', 'Year', 'Updated On', 'Latitude', 'Longitude', 'Location'. There are in total 6.57M instances of crimes.

3.2 Seattle crime data:

The second Seattle crime dataset is obtained from all the Police responses to 9-1-1from Seattle police department https://data.seattle.gov/Public- Safety/Seattle-Police Department- 911-Incident-Response/3k2p- 39jp?category=Public Safety&view name=Seattle-Police- Department-911-Incident-Response. The data ranges from 2010 to 2017 and contains 19 attributes, 1.47M instances in total. The variables include 'CAD CDW ID', 'CAD Event Number', 'General Offense Number', 'Event Clearance Code', 'Event Clearance Description', 'Event Clearance SubGroup', 'Event Clearance Group', 'Event Clearance Date', 'Hundred Block Location', 'District/Sector', 'Zone/Beat', 'Census Tract', 'Longitude', 'Latitude', 'Incident Location', 'Initial Type Description', 'Initial Type Subgroup', 'Initial Type Group', 'At Scene Time'.

4. METHODOLOGY

4.1 Data Preparation

Chicago data:

1. Target crime types:

   There are originally twenty different primary crime types, which is a very high dimensional prediction feature. I narrow the twenty types to four main types based on the crime type descriptions, which are ‘violent crimes’, ‘property crimes’, ‘crime involving weapons’, and ‘other incidents’.

2. Time interval during a day

   Time data from 12am to 12pm are transformed into eight categories of time slot, which is represented as ‘t’ column in the new data frame. In specific, T1: 12am-3am, T2: 3am-6am, T3: 6am-9am, T4: 9am-12pm, T5:12pm-3pm, T6: 3pm-6pm, T7: 6pm-9pm, T8: 9pm-12am.

3. Density

   The density feature is created using Chicago community and population data from the Chicago city government and then incorporated into the crime data frame. Furthermore, since the scale of the density value is very big relatively to other variables, so I used feature scaling in Python to transform the range of the values to 0 to 1.

4. Final features used to build the model

   ‘beat_num’, ‘Primary Type’, ‘District’, ‘Ward’, ‘Location Description’, ‘Community Area’, ‘Year’, ‘Month’, ‘Day’, ‘time’, and ‘density’ are the eleven final features that are used to build the final models. Among them, ‘Year’, ‘Month’, ‘Day’, ‘time’ are extracted from the original ‘date’ variable. ‘beat_num’, ‘Primary Type’, ‘District’, ‘Ward’, ‘Location Description’, ‘Community Area’ are originally categorical data types, but are changed to integer type in order to build models.

Seattle data:

1. Target crime types:

   There are originally forty-five different Event Clearance types, which is a very high dimensional prediction feature. I narrow the forty-five types down to four main types based on the crime type descriptions, which are ‘violent crimes’, ‘property crimes’, ‘crime related to traffic’, and ‘other incidents’.

2. Time interval during a day

   Time data from 12am to 12pm are transformed into eight categories of time slot, which is represented as ‘t’ column in the new data frame. In specific, T1: 12am-3am, T2: 3am-6am, T3: 6am-9am, T4: 9am-12pm, T5:12pm-3pm, T6: 3pm-6pm, T7: 6pm-9pm, T8: 9pm-12am.

3. Final features used to build the model

   ‘District/Sector’, ‘Event Clearance Group’, ‘Zone/Beat’, ‘Year’, ‘Month’, ‘Day’, ‘Day of week’, and ‘Hour’ are the eight final variables that I use into building the final model. Among them, ‘Year’, ‘Month’, ‘Day’, ‘Day of week’, and ‘Hour’ are extracted from the original ‘Event Clearance Date’ variable. ‘District/Sector’, ‘Event Clearance Group’, ‘Zone/Beat’ are originally categorical data types, but are changed to integer type in order to build models.

4.2 Model Building

In this project, three machine learning methods, random forest, gradient boosted trees and artificial neural network, are used to predict crime types in Chicago and Seattle.

Random Forest

Many past studies have used decision tree algorithms to predict the crime types. This project chooses to use random forest methods because Random decision forests correct for decision trees' habit of overfitting to their training set. In specific, Random forests or random decision trees are an ensemble learning method for classification in our project, that operate by constructing a multitude of decision trees at training time and outputting the class that is the mode of the class for our classification. In python, package sklearn provides a nice RandomForestClassifier method that we can directly use to build our random forest model.

Gradient Boosted Trees

Gradient boosting is a machine learning technique that can also be used in classification problems. It builds prediction models based on the ensemble of weak prediction models, typically decision trees. It builds the model in a stage-wise fashion like other boosting method, and then generalizes them by optimizing an arbitrary differentiable loss function. In specific, gradient boosting combines weak learners into a single strong learner in an iterative fashion. At each stage of gradient boosting, it may be assumed that there is some imperfect model F, and the gradient boosting algorithm improves on F by constructing a new model that adds an estimator to provide a better model. A generalization of the idea to loss functions other than squared error and to classification problems follows from the observation that residuals for a given model are the negative gradients. This method is worth exploring in this project because in recent data competitions, the method seems to outperform a lot of other methods. In python, scikit-learn provides a nice tool to use the XGBoost to build gradient boosted tree models for our crime prediction.

Artificial Neural Network

The basic ideas behind ANN is to infer the mapping implied by the data using a cost function. A commonly used cost is the mean-squared error, which tries to minimize the average squared error between the network’s output and the target value over all the example pairs. In the project, MLP, multilayer perceptron, is used to minimize the cost for the class of neural networks by gradient descent and hence produces the backpropagation algorithm for training neural networks. The Artificial neural network tends to work well with sequential data such as hand writing and speech recognition in the past, but there are not many crime prediction works done using this method. Therefore, the paper aims at exploring the effectiveness of this method in predicting our crime patterns.

5. RESULTS

5.1 Exploratory Data Analysis

Time series data:

When put all the crime data into a time series graph, we can see the temporal crime pattern from these two datasets shown below.

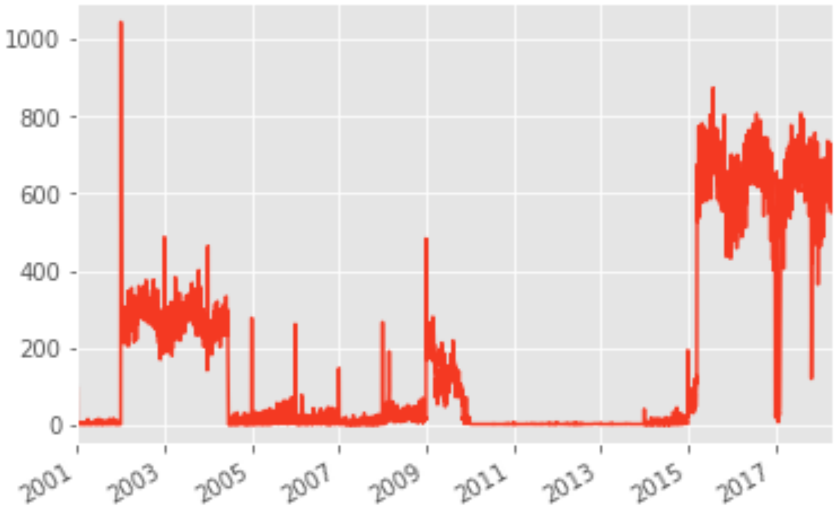

Fig1. Chicago crime time series from 2001 to 2018

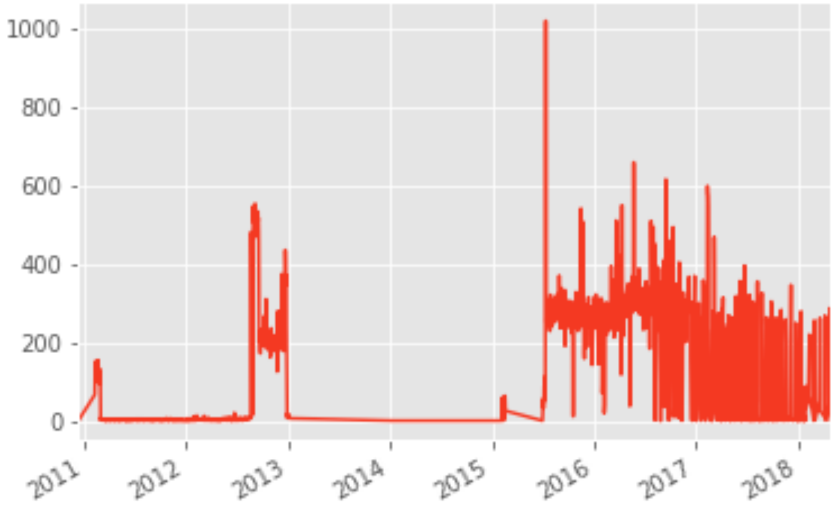

Fig2. Seattle crime time series from 2011 to 2018

For both datasets, there is significant rise in crime after 2015. This might be due to a different recording methods of crime, new allocation of police resources or different policy related to crime and social justice. Overall, the crime counts from 2015 to 2017 seem to have pretty constant seasonal volatility pattern. Although the date extends to 2018, as it is only half through the year, so 2018 is excluded in our further data analysis and model building part.

Spatial crime distribution:

The heat map obtained from google map html for both cities are shown below.

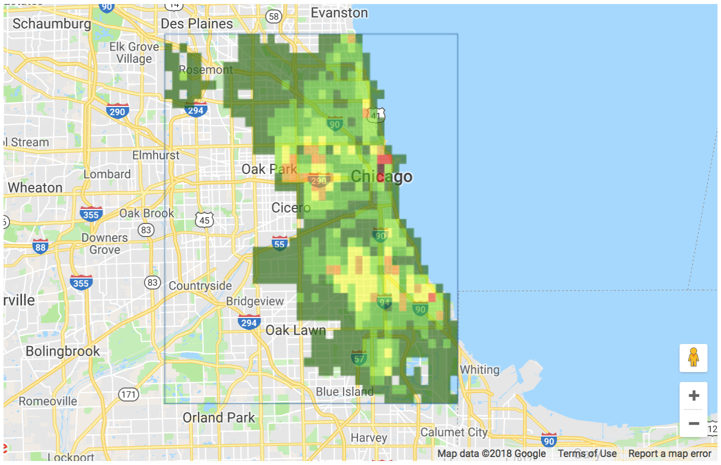

Fig3. Chicago crime heat map from 2001 to 2018

The above heat map for Chicago crime indicates the spatial distribution of crimes and the hot spots in the city. The redder the color is, it suggests the location has higher density of crime incidents. We can see from the map that the downtown area StreetVille has the highest crime density. The next crime-populated areas include west side and part of the south side, which appear to have yellow to orange color. This crime distribution makes sense as the downtown is more populated than the rest of Chicago, and there are more police patrols and also there are more targets for criminals due to the large amount of tourists.

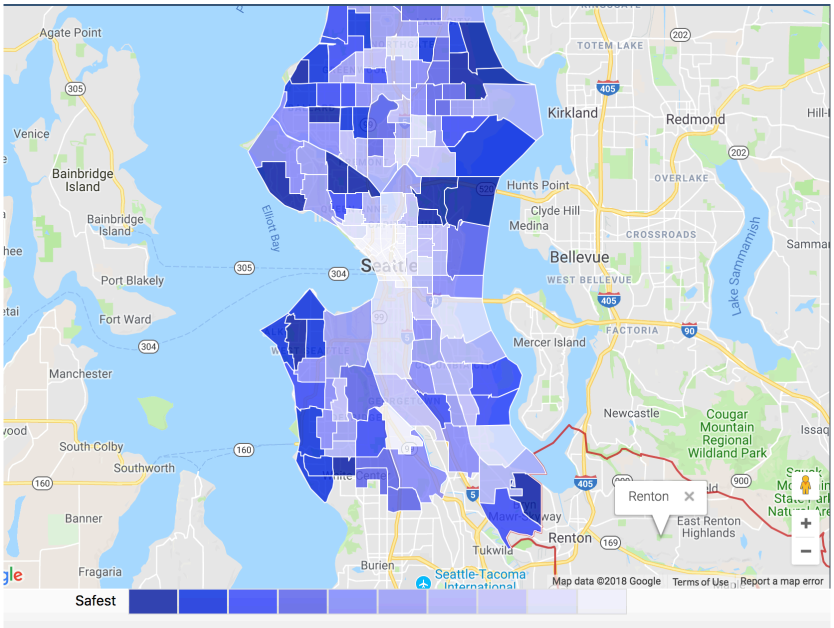

Fig4. Seattle crime heat map from 2011 to 2018

The above heat map for Seattle crime indicates the spatial distribution of crimes and the hot spots in the city. The lighter the color is, the less safe hence the higher density of crime incidents it has in that location. We can see from the map that the middle part downtown area has the highest crime density. The next crime-populated areas include several north parts and part of the southeast side, which appear to have gray color. This crime distribution makes sense as the third and first blocks in downtown areas are usually marked as dangerous and there are more criminal activities reported in public records.

5.2 Model Comparison

Since there are four crime types in total for both datasets, the baseline of model accuracy would be 25%. Any score more than this number would be considered as a success due to the complexity of our datasets. The three models for both datasets and their classification reports are shown below.

5.2.1. Random Forest

Chicago

Random Forest Classification Report				**Chicago**
	Precision	Recall	f1-score	Support
0	0.4	0.39	0.4	26640
1	0.39	0.26	0.31	27150
2	0.47	0.8	0.59	48312
3	0	0	0	24402
avg/total	0.35	0.44	0.38	126504

Seattle

				Seattle
	Precision	Recall	f1-score	Support
0	0.34	0.34	0.34	5771
1	0.4	0.44	0.42	8405
2	0.35	0.34	0.35	6931
3	0.11	0.06	0.08	1620
avg/total	0.35	0.36	0.35	22727

5.2.2. Gradient Boosted Tree

Chicago

Gradient Boosted Trees Classification Report				Chicago
	Precision	Recall	f1-score	Support
0	0.33	0.36	0.34	26640
1	0.34	0.33	0.34	27150
2	0.5	0.55	0.52	48312
3	0.25	0.18	0.21	24402
avg/total	0.38	0.39	0.38	126504

Seattle

				Seattle
	Precision	Recall	f1-score	Support
0	0.43	0.29	0.35	5771
1	0.41	0.72	0.52	8405
2	0.41	0.25	0.31	6931
3	0	0	0	1620
avg/total	0.39	0.41	0.38	22727

5.2.3 Artificial Neural Network

Chicago

Neural Network Classification Report				**Chicago**
	Precision	Recall	f1-score	Support
0	0.35	0.48	0.4	26640
1	0.36	0.14	0.2	27150
2	0.47	0.77	0.58	48312
3	0	0	0	24402
avg/total	0.33	0.42	0.35	126504

Seattle

				Seattle
	Precision	Recall	f1-score	Support
0	0.32	0.29	0.31	5771
1	0	0	0	8405
2	0.32	0.82	0.47	6931
3	0	0	0	1620
avg/total	0.18	0.32	0.22	22727

The overall model accuracy comparison is shown below:

	Chicago		**Seattle**	
	Train Accuracy	Test Accuracy	Train Accuracy	Test Accuracy
RF	93.69%	39.09%	86.29%	35.92%
Xgboost	44.45%	44.33%	41.23%	41.48
NN	42.18%	42.25%	32.59%	32.45%

6. DISCUSSION

In the classification report, four values are listed to show the comparison between different models, which are precision, recall, d1-score and support. In pattern recognition, information retrieval and binary classification, precision measures the percentage of how many selected instances are relevant, while recall measures the percentage of how many relevant instances are being selected. Both precision and recall are therefore based on an understanding and measure of relevance. The support is the number of occurrences of each class in correct target variables. The F1 score can be interpreted as a weighted average of the precision and recall, where an F1 score reaches its best value at 1 and worst score at 0. The relative contribution of precision and recall to the F1 score are equal. The formula for the F1 score is: `F1 = 2 * (precision * recall) / (precision + recall)`.

For Chicago crime datasets, for all three models, the number 2 crime type has the highest precision and recall scores among all the crime types. The reason behind this might be number 2 is crime involving weapons, which indicates that this type of crime might occur very regularly during the time of the year, month and day, so the predictive power is relatively high for this crime conditioned on other given variables.

For Seattle crime datasets, for random forest and gradient boosted trees, the number 1 crime type has the highest precision and recall scores among all the crime types. The reason behind this might be that number 1 is traffic related crime, which might occur at a regular time frame during the day, month and year since the traffic conditions can be correlated with weather, season, and light condition. Therefore, this relatively higher certainty might increase the predictive power is relatively high for this crime conditioned on other given variables in Seattle.

According to the overall accuracy score table shown above, the performances of the three models exhibit consistent patterns for both Chicago and Seattle crime datasets. In specific, Random Forest tends to fit the training data extremely well for both datasets, but it also tends to have significant overfitting problem, which results in average performance in fitting test data. Gradient boosted trees and Artificial neural network have consistent performance with related to both datasets. Overall, gradient boosted tree method achieves better results in fitting test data for both datasets, which is shown in many open data challenge competitions in recent years. The reason is mainly because gradient boosted tree methods tend to reduce overfitting problems by gradually improving the decision trees through fitting into their residuals using backpropagation algorithms.

7. FUTURE WORK

In future studies, detailed spatial distribution of crime patterns in both datasets can be further explore. For example, the spatial distribution conditioned on time, community, and special clusters within the two cities can be shown and used to conduct further deeper level analysis.

Furthermore, with related to crime, other socioeconomic and demographic factors of the two cities can also be incorporated into the study. For example, variables such as the household income, education level, and the availability of public transportations can be taken into account to the model construction part of the project.

REFERENCE

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

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.

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