

A Case Study of Spatio-Temporal Crime Prediction in Chicago

Abhijnan Chattopadhyay
Department of Probability and Statistics
Michigan State University
chattop3@msu.edu

I. INTRODUCTION

Crime data and predictive analysis regarding crime data has long been addressed for the sake of social safety and creating a protective environment related to crime. For law enforcement departments, the study of crime data is highly beneficial to fight crime and for general citizens, it has become a national cause to create safe and healthy society for the future. With advent of big data and efficient data analysis tools, one can increase the predictive power of analysis for the true random features explaining crime propagation. To fight rising crime activities, large consultant companies like IBM now offers technological solutions that can predict the patterns existing in crime data. Evidence shows that, Manchester Police Department has implemented the IBM solution and able to reduce robberies by 12 percent, burglaries by 21 percent and the theft from motor vehicles by 32 percent [8]. It is quite natural that law enforcement is completely based on the vast amount of data resources available to the various Police departments and with the science based technology capable to effectively handle them, one can develop proper predictive policing and strategic forecast. Many police departments now invest in data science to translate trend reports, correlation analysis or aggregated behavioral modeling to identify criminal hot-spots in order to gain potential advantage against crime.

However, creating efficient and effective predictive models in this context involve a lot of challenges. Both Statistics and Computer Science intellects have developed different ways to address this issue. Theoretically models deal with a large scale approximation for the predictive purpose. But, for crime prediction, small scale variations is usually most important as large scale approximation tends to smooth out the time and space dependencies incorporated for the actual prediction [1]. Another potential issue is the data itself, in the sense, developing an algorithm which is simple to understand as well as dynamically efficient to harness spatial and temporal dimension is a difficult task. Finally, implementation of such models with stable results and yet dynamically updating itself can be an area of concern. In this project, I want to dive deep into the spatio-temporal aspect of the crime data to predict which category of crime is likely to take place. The dataset used for this project is the crime data for the city of Chicago

which are available from 2001 onwards. For the sake of simplicity I only used a part of the data ranging from 2012 to 2016.

II. PROBLEM DESCRIPTION

In the era of big data, specially for Crime detection, numerous crimes are recorded daily by the data officers working alongside the law enforcement authorities throughout the United States. Many cities in the United States have signed the Open Data initiative, thereby making this crime data, among other types of data, accessible to the general public. The City of Chicago provides the city data (including the crime dataset) for public use through the Chicago Data Portal at <http://data.cityofchicago.org/>. The crime dataset spans from a time period of 2001 to present and has record of all criminal activities happened in Chicago city except for last seven days as the data is updated every week. Usually the difficulties for this daily based data arises due to their size and also complexity in understanding the spatio-temporal dependence. In this project the objective is to explore the Dynamic Spatio-Temporal models and Spatio-Temporal Neural Networks to forecast possible location and time of crime in Chicago.

Chicago is renowned as one of the violent crime cities in the world. The Brennan Center of Justice projects a 47.1 percent increase in murder compared to 2015. [9] Though the national rate of homicide has an expected increase of 13 percent in 2016, majority of this is from Chicago alone [9]. As of November 2016, the homicide count in Chicago was 616, more than New York and Los Angeles combined [10].

In our project, we use the spatio temporal feature in the database to try out multiple classification algorithm for multi class prediction: Naive Bayes, Random Forest, Support Vector machine etc. Most part of this work will be focused to identify hot-spots of crime in Chicago.

Moreover, I wanted to address the so called Next Check in Problem in this data aspect. This means, we try to predict the next crime activity a location will encounter given the historical data in that spatial location. We would rank all the potential target categories in the prediction region,

so that the actual target category is ranked as high as possible.

III. RELATED WORK

There is in-numerous amount of literature pertaining to this area. People have talked about large datasets, performed multidimensional time series analysis, and even incorporated spatial demographic variables to better train the model. Those existing models has different potentials in varied geographic regions. So in this project we are looking for the related models focused to the Chicago region.

Related research work containing a dataset for the city of Philadelphia with crime information from 1991 to 1999. This study [5] by Bogomolov et al, used the multi scale-complex relationship between both space and time to identify crime hot spots based on the location of high crime density. Chung-Hsien Yu et al has classified different neighborhoods of a city as hot-spots of residential burglary with the help of a variety of classification algorithms, such as, Support vector machine, Naive Bayes and Neural Networks [6]. Also, for the city of Philadelphia, Toole et al shows that significant spatio-temporal correlation exists and identified clusters of neighborhood whose crime rates are affected simultaneously by external forces [7]. More recently, Kaggle hosted San Francisco Crime Classification competition, which has also been groundbreaking for the literature.

IV. DATASET DESCRIPTION

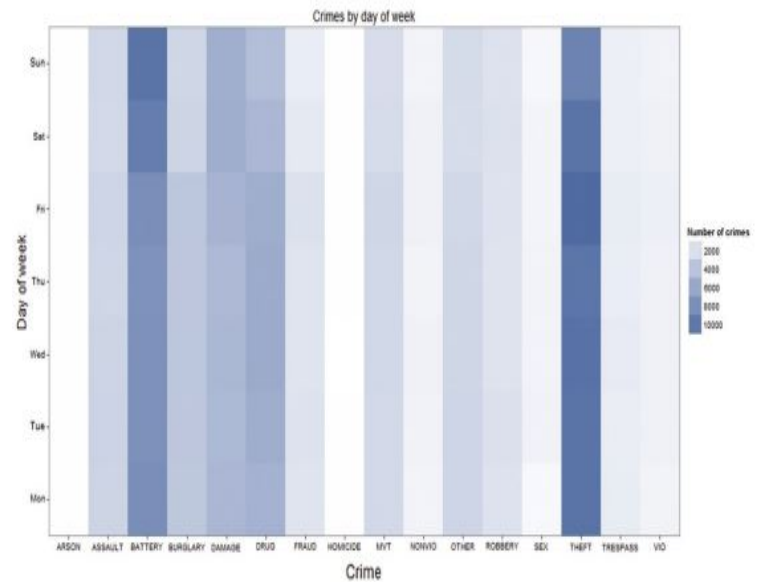
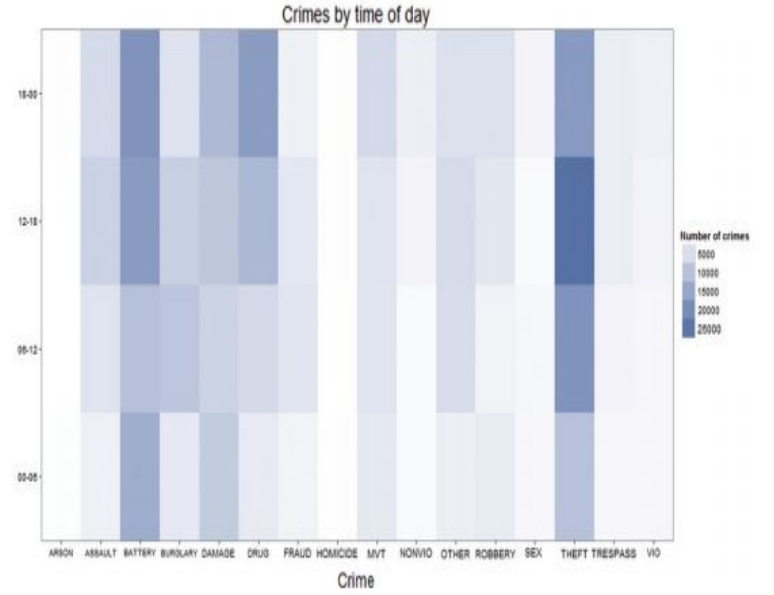
Chicago crime data are downloaded from the City of Chicago data portal. The dataset contains reported incidents of crime (with the exception of murders where data exists for each victim) that occurred in the City of Chicago from 2001 to the present (Crimes - 2001 to present, 2016). For each reported crime, the dataset contains the date and time, block identification, type of crime, crime description, latitude and longitude, and FBI code. [11].

In this project I only worked on the time frame 2012 to 2016. For the training set and the test set, crime activity from 2012 to 2014 is taken into the former class, same for 2015 and 2016 is treated as the testing set. In the each set, we have the following variables:

- **Timestamp**- Time of occurrence for the crime incident, in the format of YYYY-MM-DD HH:MM:SS.
- **Category**- Category of the crime incident. This is our target variable.
- **Day of week**- Which day in the week the incident took place.
- **District**- Name of the police department district the incident was reported in.
- **Resolution**- The way crime was resolved(arrest or not).
- **Address**- The approximate street of the incident.
- **Latitude**

• Longitude

Before taking the classification and the prediction task, we performed a bunch of explanatory data analysis given the dataset. At first, we omitted the arrest variable, description of the crime and the approximate address of the incident taking place from our dataset. Which later found out to be pivotal. Since we are dealing with a number of categorical variables it was important to encode those categories as dummy variable. Also from the time stamp from the dataset, it was important to decompose them into Year, Month(1,2,...,12), Date(1,2,...,31), Hour(0,1,...,23) and minutes(0,1,...,59). From the primary findings, we can see how does the different crime pattern changes with different times of day and different times of the week.



V. METHODS

Given the original database there can be a number of analysis that can be performed on the dataset. One can perform a number of explanatory data analysis, have visual representations of the dataset, create a classification problem based on the features and also make predictions about the number of crimes. Due to lack of time during this class project, I focused myself mostly on the classification framework and also for a statistical modeling, I added a spatio temporal prediction based on a number of transformed features. Due to extensive literature of this topic, I spent most of this time referring to different literature for Spatio-temporal sequence forecasting. I will mostly refer to [12], [13] for this work.

Below are the two methods I used to perform the classification task:

- **Naive Bayes Classifier**

Naive Bayes Classifier is a supervised learning algorithm which is both easy and effective while tackling multi class problem. This is indeed a statistical model that predicts class membership probabilities based on Bayes' theorem. We applied the multi variate version of the classifier to model the target variable(Category). Since the Category of crime is not overlapped we can model the events independent of each other. So the problem is now a multi-class classification problem where the target variable Y can take any one of the 33 class, represented by number 1 to 33. we try to model the probability ϕ_y using multinomial distribution

$Y \in 1, 2, \dots, 33$ and $Y \sim \phi_y(\text{Multinomial})$

Spatial information like latitude and longitude are not taken into consideration. All the remaining categorical features are taken as predictor variables. Each of the features has different range of values, for example the variable Month can be between 1 to 12, Days of the week can be between 1 to 7 and so on. So, we have to model each feature by multinomial distributions, $X_i \in 1, 2, \dots, s_i$ and X_i given $Y = j \sim \phi_{i|y=j}(\text{Multinomial})$. So we have jointly modeled the parameters $\phi_y, \phi_{i|y=j}$ by using the laplace smoothed equations.

- **Random Forest**

A collection of decision trees is constructed from K random samples drawn from the training set with replacement. Each decision tree is grown by repeating the following steps for each node in the tree until we reach the limits on the size of the tree: Randomly select m out of the p feature variables. Split the decision tree on the variable that maximizes the Gini impurity associated with the variable. So Broadly speaking, Random Forest is ensemble learning algorithm which aggregates a family

of classifiers and each member of the family works as a classification tree. So over here each tree works as a separate decision tree, until the data is totally partitioned or we reach the maximum allowed depth.

VI. RESULTS

A. Classification on the 33 Classes

In this section we comment about the result based on the classification done on the 33 class labels for the initial models. Next, We confined ourselves to the top 4 crime categories and re perform the analysis. With the original target category as 33 classes, Naive Bayes classifier gave 40 percent accuracy on the training set and around 20 percent accuracy on the testing set. Similar trend was observed when performed over varying training size. Accuracy did not improve significantly even there as well. But what we can comment from here is that we did very poorly while classifying on the testing set. But still we get a better result than random guessing.

Next we implemented The random forest algorithm on the same setup. We implemented Random forest on the training set with a accuracy over 80 percent but performing cross validation on 5 folds, we receive a test error of 70 percent, which is pretty high compared with the training error and also indicating high variance. So our classifiers work very poorly while predicting the proper classes given the feature vectors. One possible comment based on our observation might be that usually classifiers like this are modeled based on numerical features, But for our dataset we largely had categorical predictors with dummy codings. This might be one of the reasons for poor results.

B. Classification on the 4 Classes

Since our classifier worked very poorly on the multi class classification, we thought to collapse the classes and perform the analysis on the basis of a smaller set of class. One intuitive idea is to take the top 4 crime categories and try to classify them based on the training and testing sample. So we figured out the top 4 crime categories, which are : Theft, Battery, Criminal Damage and Narcotics. We also put thought on this categories so that they are not biased towards any of the existing class. Since, If there is a minority class, classifier tends to give false positive classification probability towards the majority class. We perform the random forest algorithm with proper tuning of the parameters and were able to improve the precision power. Below is the confusion matrix provided by the random forest classifier.

variable	Theft	Battery	criminal damage	Narcotics
Theft	0.978404624	0.021505371	0.000000000	0.000000000
Battery	0.005524862	0.988950276	0.000000000	0.005524862
criminal damage	0.000000000	0.027777778	0.916666667	0.055555556
Narcotics	0.009527710	0.019098619	0.000000000	0.971542857

Since our goal was to predict the probability of a crime belongs to a certain category based on its time and location,

we thought it can be useful to inspect this as a multi class logarithmic loss, given by,

$$\text{logloss} = -\frac{1}{N} \sum_{i=1}^N \sum_{j=1}^M y_{ij} \log(p_{ij})$$

Where N is the Number of samples in the test cases, M is the class labels, y_{ij} is the observation corresponding to the value 1 if the observation i is in class j and 0 otherwise, Also, p_{ij} is the predicted probability that observation i belongs to class j, log is the natural logarithm function. We obtain that the multiclass log loss of train set is 1.9278 and on test set 2.3410 .

VII. SPATIO TEMPORAL PREDICTION

For the prediction purpose we actually processed the data quite a bit. Since the objective is to get a good predictive model based on the spatial information and time information, we actually collapsed the classes and counted the number of crime activities based on a location and time. We indicate $Y_{s,t}$ as the number of crime incidents on the beat s and corresponding to the day t . Since we are working with a count variable, in order to model the count data, we have to use the generalized linear model for the regression setup with the response being modeled as either Poisson or Negative Binomial regression. One principal assumption of the Poisson random variable is the equality of mean and variance. Since in our model, we did not have this condition to be satisfied we tried to model this based on the negative binomial regression. For the first part of the model we neglect the spatial structure in the response and constructed a generalized linear model with negative binomial regression.

The mathematical formulation is given by:

$$Y_i \sim \text{Negbin}(\mu_i, \alpha)$$

$$\log(\mu_i) = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_k x_{ik}$$

Where μ_i is the conditional mean which is applied to a log link, β_i are the regression coefficient and x_i are the predictors. Below is the results obtained from the regression.

variable	estimate	p value
(Intercept)	-0.17349312	2.00E-16
past.crime.1	0.012163721	2.00E-16
past.crime.7	0.001914812	0.049199
past.crime.30	0.01013589	2.00E-16
policing	-0.231141	2.00E-16
crime.trend	1.33744581	2.00E-16
dayMon	-0.055998421	1.04E-14
daySat	-0.049561245	1.06E-11
daySun	-0.10095315	1.98E-09
dayThu	-0.043147585	0.000859
dayTue	-0.024131177	0.00185
dayWed	-0.022313666	4.05E-08
seasonspring	0.030796798	2.39E-15
seasonsummer	0.044990464	0.297866
seasonwinter	-0.005747906	0.001276

Afterwards we try to incorporate the spatial structure and the temporal structure in the model. We have assumed a Conditional Autoregressive structure in the spatial regime. The weight matrix is computed from a different shape file. This structure guarantees that neighboring beats(locality) assume

similar pattern, whereas non-neighboring beats are dissimilar. We assume a structured random walk for the temporal dependence. For the sake of simplicity, we did not consider space and time interaction. We have a total of 297 beats and 359 days where we have performed the regression. One of the important prospect of the analysis was to showcase that the variable Policing can decrease the number of criminal activities in the area. we consider the following model setup:

$$Y_{st} \sim \text{Pois}(\mu_{st})$$

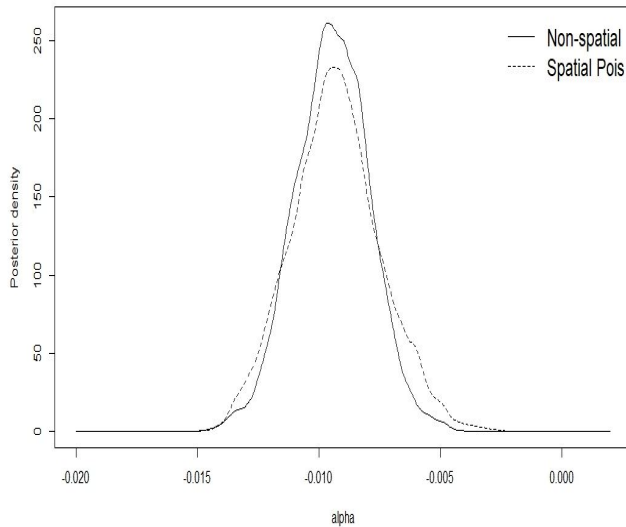
$$\log(\mu_{st}) = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_k x_{ik} + w_s + \epsilon_t.$$

w_s is random effect term that captures possible correlation among the response variable, Y, taking into account the spatial effects of neighbouring states. In statistical literature this type of model is known as Areal data model and one can model this random effect term by a Conditional Auto-Regressive(CAR) structure, that is:

$$w_s | \{w_{s'}; s \neq s', s' \in A(s)\}, \zeta_w \sim N(0, \frac{\zeta_w^2}{n_s})$$

where $A(s)$ denotes the set of neighbors corresponding to province s , $n(s)$ denotes the number elements in $A(s)$, and ζ_w is an unknown parameter. Also ϵ_t is the time dependent variable assume to have a random walk stochastic structure.

For the implementation of this task, we took Bayesian approach and performed the task on JAGS. We implemented MCMC gibbs sampling with 1000000 burn in period and another 1000000 to capture the posterior samples. We also guaranteed the posterior convergence. Coming back to our research question to whether policing helps to decrease the number of criminal activities. Just like the non spatial model we obtain a negative coefficient for the posterior mean which suggests that our claim is true. Below is the posterior density to suffice our result. We calculated the BIC to judge for the model improvements, we obtain a BIC of 42265 for the Nonspatial model and 41009 for the spatial model. So the model was not improved significantly in spite of the model complexity. The spatio-temporal models in this paper are Bayesian hierarchical models for Poisson with spatial and temporal random effects. The main challenge for these models is computational efficiency, since the typical Markov Chain Monte Carlo methods are constrained by model complexity and database dimension. A solution to this challenge is provided by the R package INLA, which uses the Laplace approximation instead of MCMC. The Laplace approximation is a good approximation and highly efficient.



VIII. DISCUSSION AND FUTURE WORK

So from our project we have a few good results and a few poor results. The initial task of classifying the 33 different crime categories was indeed challenging. Our naive Bayes and Random forest classifier faired poorly in view of this problem. The possible reason behind this might be the categorical or factored dummy variable setup of the predictors. Though our training and testing errors were pretty high, it was still better than random guess. One possible remedy, for this problem was to collapse the classes and perform the multi class classification problem based on a lower number of output classes. What we obtain is that Random forest classifier way better work in this case as can be confirmed from the confusion matrix. For the prediction purpose, we implemented two models, One without the spatial components and other with the spatial components. Both models give mostly similar results, but for the spatial model, the model complexity was much more higher than the spatial one. One of the findings of this project was that, proper application of policing(number of arrest in a region) can significantly lower the number of crimes.

For the possible extension for our project can be the multiclass classification with a neural network setup. We did not find any suitable way to address the spatial information in the dataset, which can be treated as a latent factor while classifying. Another possible extension is to rank the different crime categories in a location such that we don't classify the top crime in a locality but predict a ranking, say, the first 5 crime category for each locality given a specific time stamp. I guess, this has a more applied significance for the city police or the government as a whole.

REFERENCES

- [1] Clark, Nicholas J.; Dixon, Philip M. Modeling and estimation for self-exciting spatio-temporal models of terrorist activity. *Ann. Appl. Stat.* 12 (2018), no. 1, 633–653. doi:10.1214/17-AOAS1112. <https://projecteuclid.org/euclid.aoas/1520564487>

- [2] Wikle, Christopher. (2015). Modern perspectives on statistics for spatio-temporal data. Wiley Interdisciplinary Reviews: Computational Statistics. 7. 10.1002/wics.1341.
- [3] Ziat, Ali, Edouard Delasalles, Ludovic Denoyer and Patrick Gallinari. Spatio-Temporal Neural Networks for Space-Time Series Forecasting and Relations Discovery. 2017 IEEE International Conference on Data Mining (ICDM) (2017): 705-714.
- [4] Dash, Saroj Kumar et al. Spatio-temporal prediction of crimes using network analytic approach. 2018 IEEE International Conference on Big Data (Big Data) (2018): 1912-1917.
- [5] Bogomolov, Andrey and Lepri, Bruno and Staiano, Jacopo and Oliver, Nuria and Pianesi, Fabio and Pentland, Alex. 2014. Once upon a crime: Towards crime prediction from demographics and mobile data, Proceedings of the 16th International Conference on Multimodal Interaction
- [6] Yu, Chung-Hsien and Ward, Max W and Morabito, Melissa and Ding, Wei. 2011. Crime forecasting using data mining techniques, pages 779-786, IEEE 11th International Conference on Data Mining Workshops (ICDMW)
- [7] Toole, Jameson L and Eagle, Nathan and Plotkin, Joshua B. 2011 (TIST), volume 2, number 4, pages 38, ACM Transactions on Intelligent Systems and Technology
- [8] Williams, Paula. "6 simple ways to help fight crime with analytics." IBM Big Data and Analytics Hub, 5 May 2016.
- [9] Mock, Brentin. "Where Crime Is Rising in 2016." CityLab, The Atlantic, 20 Sept. 2016.
- [10] Gerner, Jeremy. "August most violent month in Chicago in nearly 20 years." Edited by Peter Nikeas and Elvia Malagon, Chicago Tribune, 29 Aug. 2016.
- [11] "Crimes - 2001 to present." City of Chicago Data Portal, City of Chicago. <https://data.cityofchicago.org/Public-Safety/Crimes-2001-to-present/ijzp-q8t2/data>
- [12] Anastasios Noulas, Salvatore Scellato, Neal Lathia, and Cecilia Mascolo. 2012. Mining User Mobility Features for Next Place Prediction in Location-Based Services. In Proceedings of the 2012 IEEE 12th International Conference on Data Mining (ICDM '12). IEEE Computer Society, Washington, DC, USA, 1038-1043. DOI=<http://dx.doi.org/10.1109/ICDM.2012.113>
- [13] Shi, Xingjian and Yeung, Dit-Yan. (2018). Machine Learning for Spatiotemporal Sequence Forecasting: A Survey.