Spatio-Temporal Modeling of Criminal Activity

Michail Misyrlis
University of Southern California
misyrlis@usc.edu

Chung Ming Cheung
University of Southern California
chungmin@usc.edu

Ajitesh Srivastava
University of Southern California
ajiteshs@usc.edu

Rajgopal Kannan
US Army Research Lab-West
rajgopal.kannan.civ@mail.mil

Viktor Prasanna
University of Southern California
prasanna@usc.edu

ABSTRACT

Accurate crime forecasting can allow law enforcement to more effectively plan their resource allocation such as patrol routes and placements. We study the effectiveness of traditional regression approaches in forecasting crime occurrences in Portland, Oregon. We divide the area of interest into equally spaced cells and investigate the spatial autocorrelation between the crime occurrence rates of neighboring cells. We also attempt to use neighboring cells' information in the regression models along with the cell's own time series to enhance the forecast results. Our results show that regression is a promising method that outperforms a moving window averaging method, especially when the future horizon to be predicted increases. However, addition of neighborhood cells decreased the quality of predictions, suggesting that spatial correlation in crime is more complex than geographical neighborhood. We also explore a possibility of connection of criminal activities and popularity of crime incidents in Portland on the Web, and discuss future directions we will take to improve crime prediction.

CCS CONCEPTS

•**Mathematics of computing** →**Time series analysis;** *Multivariate statistics;* •**Information systems** →**Spatial-temporal systems;** •**Human-centered computing** →*Social engineering (social sciences);*

KEYWORDS

Crime Prediction, Time Series, Spatial Correlation, Urban Sensing, Social Media Crime Anycasting

ACM Reference format:
Michail Misyrlis, Chung Ming Cheung, Ajitesh Srivastava, Rajgopal Kannan, and Viktor Prasanna. 2017. Spatio-Temporal Modeling of Criminal Activity. In *Proceedings of SocialSens'17, Pittsburg, PA, USA, April 21 2017,*, 6 pages.
DOI: http://dx.doi.org/10.1145/3055601.3055613

1 INTRODUCTION

Crime forecasting is an area of interest to both police agencies and academic researchers. Improving the accuracy of crime prediction will allow the limited policing resources to be allocated appropriately so as to render communities safer. Thus, it is pivotal to gain further insight into any spatial and temporal patterns that criminal activities follow.

This work is supported by U.S. National Science Foundation Eager award under Smart and Connected Communities program, that aims to perform data-driven modeling of crime for crime prediction and to develop crime prevention strategies. We view this problem as that of urban sensing, where data generated by calls for service, connectivity of locations, socio-economic factors, occurrence of major events, and popularity of criminal activity reports on social media, all come into play. This paper presents some initial results on our spatio-temporal analysis to predict crime, and future directions we wish to take in this project.

Here we focus on prediction of the location where crimes will occur and the time when that will happen using the history of criminal activities in the region of interest. We abstract the problem as a one-step forecasting problem. A one-step forecasting problem is defined as the problem of predicting the value at a certain time given a certain number of immediate historical values, i.e., a look back window of certain length. We solve the problem of predicting the number of crimes that will happen in a location in a week given the number of crime occurrences in the past. We can then use the predictions to identify crime hotspots in the region.

Hotspot mapping [2, 3] refers to identifying areas of increased criminal activity. An established popular technique for mapping hotspots is known as kernel density estimation (KDE) [1–3]. It produces a smooth map in which the height and/or color gradient reflects crime count and has been popularized due to its wide availability. A different set of approaches leverages clustering methods. In [7] a clustering method is proposed to learn the boundaries between regions of different crime trends, while in [13] the authors develop a clustering and boosting algorithm that learns a global prediction pattern based on local ensemble patterns. Instead of using hotspots, we apply well-known regression methods to obtain accurate predictions.

The rest of the paper is organized as follows. In Section 2 we discuss the methods that are considered in this work. In Section 3.1 we briefly introduce the dataset used in our experiments, while in Section 3.2 we explain the our evaluation metrics. In Section 3.3 we present the results of our experiments. In Section 4 we discuss our plans on how to extend this work by leveraging additional information in order to get a higher accuracy of prediction. Finally in Section 5 we provide a brief conclusion of our work.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
SocialSens'17, Pittsburg, PA, USA
© 2017 ACM. 978-1-4503-4977-2/17/04…$15.00
DOI: http://dx.doi.org/10.1145/3055601.3055613

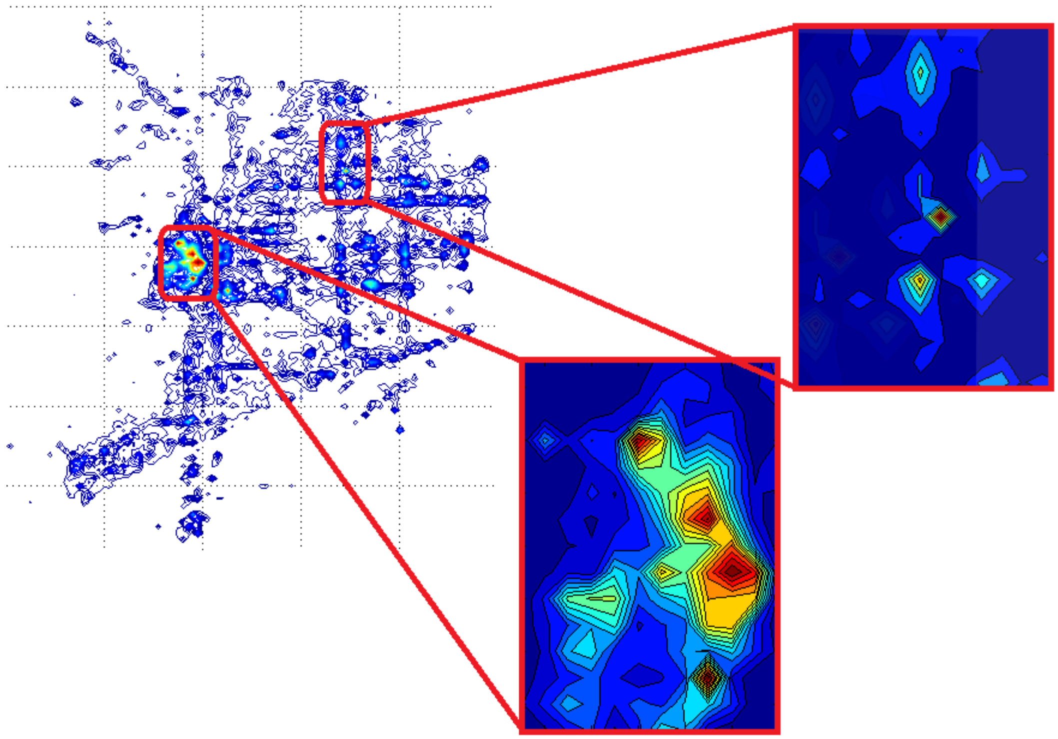

Figure 1: Crime density plot for the city of Portland. Two regions of interest with high density of crime have been shown: Downtown and North East Portland (from left to right)

2 METHODOLOGY

In this section we introduce the methods that we used in our experiments. Let $y_{i,t}$ be the number of incidents of crime at location i at time t. We denote by $X_t \in R^{m \times n}$ the matrix with m samples and n features that represents historical time series values for each location, i.e., $X_{i,j} = y_{i,t-j}$. The number of incidents at time t is modeled as the vector $\mathbf{y_t} = \{y_{i,t}\}$ in the form of a linear regression model:

$$\mathbf{y_t} = \beta X_t \tag{1}$$

In regression model the goal is to obtain a vector of weights β that minimizes the linear combination of a loss function, which is typically squared error loss, and a regularizer term that penalizes some norm of the weights:

$$\beta = \arg\min_{\beta} \|\mathbf{y} - \beta X_t\|_2^2 + \lambda J(\beta) \,. \tag{2}$$

This assumes that $y_{i,t}$ is a linear function of $y_{i,t-1}, y_{i,t-2}, \ldots, y_{i,t-n}$. n is the size of the time interval used for prediction. We consider two different regularizers. First, Ridge regression [4], for which $J(\beta) = ||\beta||_2$. The ℓ_2 norm imposes a constraint that encourages shrinkage of the weights. Second, we employ Lasso regularization [10], for which $J(\beta) = ||\beta||_1$. In this case, penalizing the ℓ_1 norm encourages both shrinkage and sparseness of weights, leading to some features (historical values) being zeroed out, and thus not contributing to the predictions at all. Finally, we also consider a moving window averaging method in which the predicted value is simply the mean value of the past n values of the time series

$$y_{i,t} = \frac{1}{n} \sum_{j=1}^{n} y_{i,t-j} \,. \tag{3}$$

We model the dataset of crime occurrences over a region as grids of equally sized cells. Each cell holds the number of crimes that have occurred in it over a window of time (n). We train a separate regression model for each cell in the data grid to predict the next value, given a historical window of values. Samples are obtained by shifting a moving window of a certain size (n) across the entire time series of crime occurrences for each grid so that we have chunks of time series of size n. The target value to be predicted for each of these chunks is the $(n + 1)^{th}$ value.

We also investigated the benefit of including neighboring cells' information in the regression by appending the number of crimes among the 8 neighboring cells at $t - 1$. The spatial autocorrelation between immediate neighboring cells in the dataset can be measured by using the Moran's I measure [8]. We measure the spatial autocorrelation value of the two regions we experimented on, and assess whether this additional method is able to leverage this correlation in order to make more accurate predictions. The model along with spatio-temporal information is

$$y_{i,t} = \sum_{k} \beta_k y_{i,t-k} + \sum_{l \in Neighbor(i)} \alpha_l y_{l,t-1} \,. \tag{4}$$

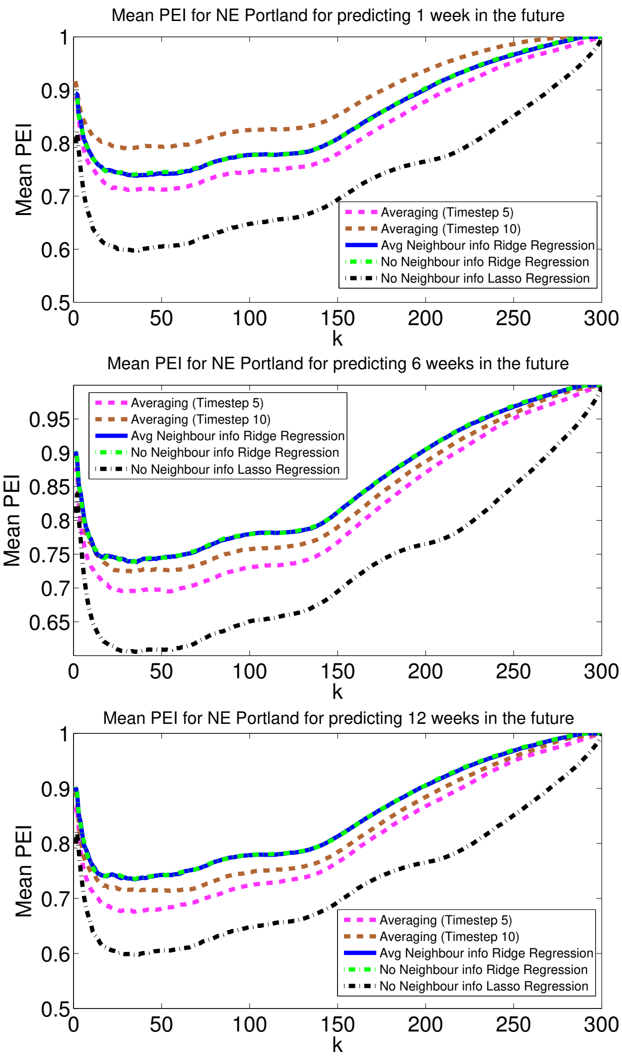

Figure 2: Comparison of PEI for varying k for various methods when predicting for an increasing number of weeks into the future in NE Portland

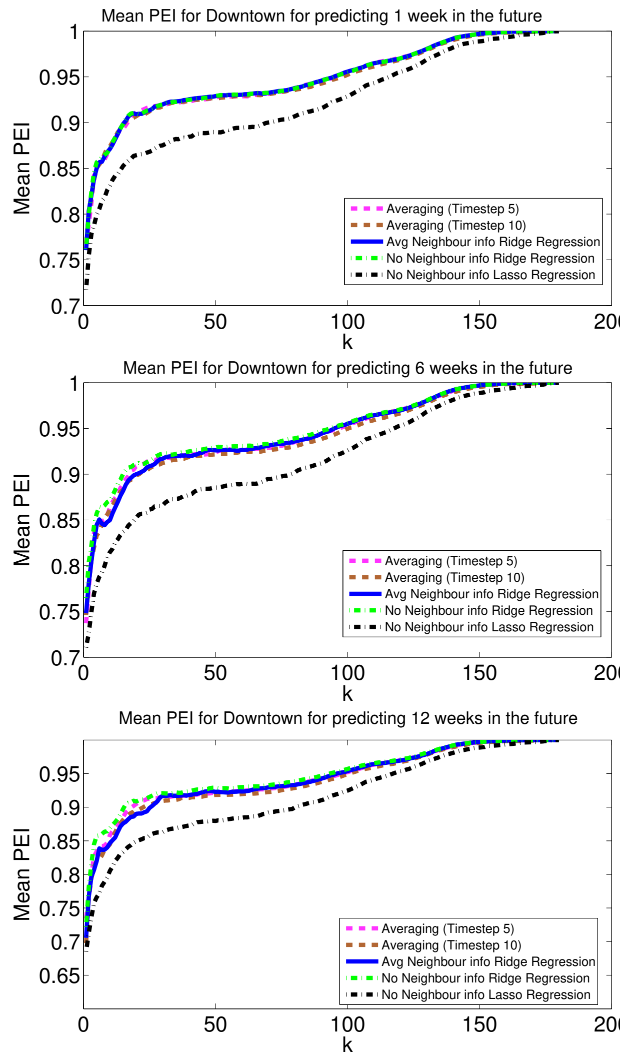

Figure 3: Comparison of PEI for varying k for various methods when predicting for an increasing number of weeks into the future in Downtown Portland

3 EXPERIMENTS

3.1 Dataset

We evaluated the discussed methodologies on a public dataset of crime occurrences in Portland, Oregon reported from March 2012 to December 2016 [9] provided by the National Institute of Justice. The dataset contains geolocation, crime type, date and time of the call for service. The area of interest is divided into equally spaced square cells. We count the number of occurrences of crimes in each cell in intervals of one week, which forms one data point $y_{i,t}$.

Since cells which have few crimes are not very informative, we focus our analysis on areas of Portland that have the highest crime rates. We plotted a contour graph of the total crime occurrences over the entire Portland within the time period of the dataset, and picked two regions in the map that had high crime rate, as shown in Figure 1. We extracted a 12×15 grid in the Downtown Portland region and a 15×20 grid in the North East Portland (NE Portland) region for our experiments. For each cell in both of these grids, we have a time series of length 252. From each of these time series we extract windows (look back) of length n, where $n = 5, 10$.

3.2 Evaluation Metrics

The algorithms are evaluated using two measures - Mean Squared Error (MSE) and Predictive Efficiency Index (PEI) [5].

MSE is a common error measure for regression problems. It is computed by mean of all the squared error of the predicted values against the ground truth. Mathematically,

$$MSE = \sqrt{\sum_i (y_i - \hat{y}_i)^2}, \tag{5}$$

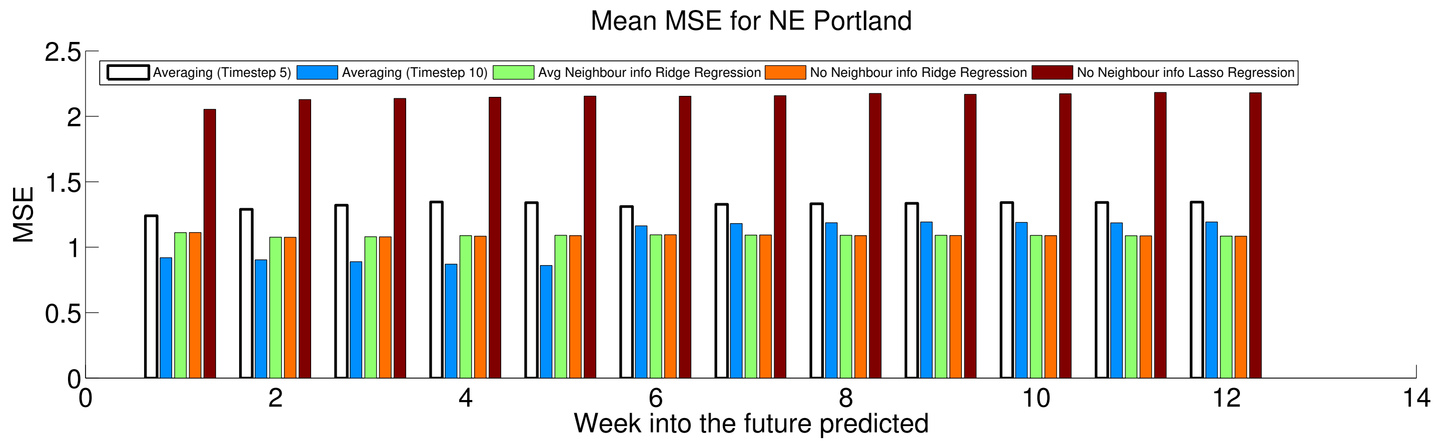

Figure 4: Comparison of MSE for varying k for different algorithms when predicting for different numbers of weeks in the future in NE Portland

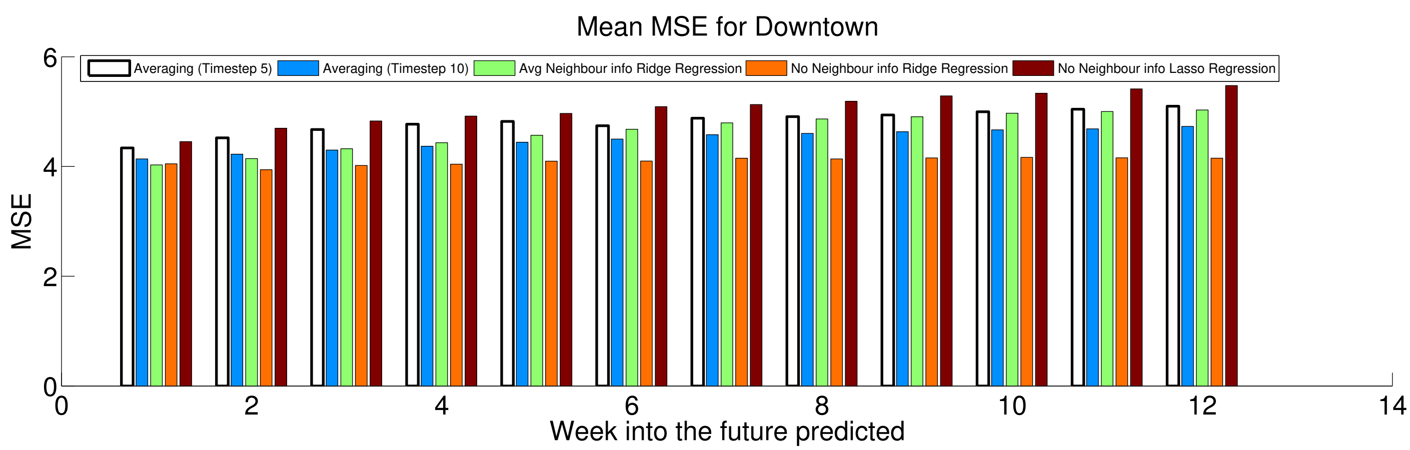

Figure 5: Comparison of MSE for varying k for different algorithms when predicting for different numbers of weeks in the future in Downtown Portland

where y_i is the target value for the i-th data point and $\hat{y}_i$ is the corresponding predicted value.

PEI is an index that measures how well the algorithm's predictions are able to capture the regions of crime hotspots compared to the optimal solution. It makes use of the Prediction Accuracy Index (PAI) [1] to measure the effectiveness of the forecast information in capturing hotspot information. Assume that the k cells with highest forecast crime occurrences are determined as hotspots, the PAI is then the ratio between the crime density in the forecast hotspot region over the overall crime density in the entire region of interest.

$$PAI = \frac{n}{ka} * \frac{A}{N}, \tag{6}$$

where n is the number of crimes in the forecast hotspot region, N is the total number of crimes within the prediction week, a is the area of one cell, A is the total area of the region of interest.

Then, PEI is the ratio of PAI of the forecast result to the maximum PAI that can be achieved for the given k. Thus,

$$PEI = \frac{PAI}{PAI^*}, \tag{7}$$

where PAI* denotes the maximum possible PAI. Alternatively it can be written as

$$PEI = \frac{n}{n^*}, \tag{8}$$

where $n*$ is the number of crimes in the hotspot region for obtaining the PAI* value.

To further test the effectiveness of the methods, we evaluate the one-step forecast models to predict for more than one week in the future. To achieve this, we use the given chunk of time series $(y_{i,t-1}, y_{i,t-2}, \ldots, y_{i,t-n})$, to obtain a prediction $\hat{y}_{i,t}$, and then use $(\hat{y}_{i,t}, y_{i,t-1}, \ldots, y_{i,t-n+1})$ to get $\hat{y}_{i,t+1}$, and so on. In other words, for each week predicted, the latest predicted value is appended to the data point with the oldest week taken out, the new data point is then used to predict the following week. We present the results for predictions up to 12 weeks by using the described chaining method.

3.3 Results and Discussion

We performed experiments to compare the mean PEI values acquired by the averaging method and regression methods as k increases. The mean PEI value is given by the mean of the PEI values for the predictions over all the weeks we have data for, when k is set to be a specific value.

For the averaging method, we present the results for look back window size of 5 or 10. For ridge and lasso regression, we used grid search to find the best value for the regularization term λ and plot the curve for the best parameter with window size $n = 5$. We also conducted experiments with regression on window size 10, but those results are omitted as regression with $n = 5$ always performed better. This could be due to overfitting of the model. We also investigated if including neighboring cells' information helped enhance the results. We computed the Moran's I for the Downtown Portland and North East Portland region. Their values are 0.2925 and 0.0565 respectively, implying that Downtown shows some autocorrelation among immediate neighbors while neighbors in NE Portland are almost uncorrelated.

Figures 2 and 3 show the results of the experiments for predicting 1, 6, and 12 weeks into the future in north east Portland and Portland Downtown respectively.

We noticed that averaging works better as the look back window increases, as this makes the method more rigorous against high variance data. Lasso Regression does not perform well in any cases. For all methods except both ridge regression methods in NE Portland, and ridge regression with no neighbor information in Downtown, we see a decrease in performance as the number of weeks to predict into the future increases. For ridge regression, no neighbor information approach remains to be the best performing algorithm in both regions, while the performance of using neighbor information drops as number of weeks in the future to predict increases for Downtown but not NE Portland. This suggests that either the regression model is unable to capture the neighbor information effectively to improve the prediction accuracy or the autocorrelation between neighbors are small and is misleading to use for predictions. These results are also supported by looking at the MSE of various methods for varying future weeks to predict in Figure 4 and 5. However, we do not observe such a performance deterioration for NE Portland, which has a much lower correlation between neighbors. This means that regression is able to simply ignore the unnecessary extra information.

4 FUTURE DIRECTIONS

4.1 Capturing Neighborhood Information

We observed that including neighborhood information into the regression model did not improve prediction accuracy. Although, in criminology literature [6] "near-repeat victimization" has shown to occur, it may depend on the the geography of the region (homogeneous vs heterogeneous housing, etc.) as pointed out by [11]. This motivates a definition of a more generic neighborhood compared to only physical neighborhood, which is based on relatedness of crime at two locations. These "neighbors" of a given location may be geographically farther than some other locations, yet perhaps due to easier accessibility may have more impact on occurrence of crime.

4.2 Connection with Social Media

Ideally, we expect the crime rate to decrease immediately after a major crime incident in that location. This is due to the fact that a news related to that incident would spread rapidly over social media, bringing attention to the area, and thus decreasing the opportunity of crime. As the first step to test this hypothesis, we compared the time-series of total number of violent crime in Portland with popularity of keyword search "Portland crime" on Google Trends[1] News. We assume that number of searches in Google News is a good indicator of popularity of crime incidents on the web. From Figure 6, we observe that there is a decrease in crime, on and after a week where there is a peak in popularity in the news. This partially aligns with our hypothesis. We expect, having "finer grained" data such as daily keyword popularity instead of weekly, would show more relationships. Moreover, we plan to find location specific popularity on the Web and compare that to the time-series of the cell containing that location. Finally, we plan to incorporate popularity on Twitter and other social media [12]. Including "popularity" information should improve the prediction of crime.

4.3 Other Factors Affecting Crime

Instead of relying solely on past criminal activity, we plan to develop scalable data-driven algorithms that will adapt with the constantly evolving state of criminal activity by continuously learning from a rich set of spatial and demographic features, including traffic, spatial attributes (e.g., proximity to the nearest business, residential areas, and main streets), socio-economic characteristics of neighborhoods, and current time, as well as context (e.g., large social events). The rationale is that such a rich set of features will lead to better crime index estimation in known crime-prone areas as well as better crime forecasts for previously "crime-free" regions, as a direct result of the correlations between, for example, burglary and business proximity.

5 CONCLUSION

We investigated the use of traditional regression methodologies for a one step forecasting problem in the context of crime occurrence predictions. Our experiments were performed on a dataset of call of service histories in Portland, Oregon from March 2012 to December 2016. The results demonstrated that ridge regression outperformed the baseline method of using the mean of historical crime occurrence data as the prediction. The method was also able to maintain high prediction accuracy even as the number of weeks into the future to be predicted increases. However, it was unable to leverage the information provided by immediate neighboring regions that were in the neighborhood of the target cell.

We plan to investigate better methodologies to find locations with related crimes as we observe that physical neighborhood does not seem to capture "near-repeat victimization". We have observed some connection of crime rate at certain points in time with popularity of crime reports on the Web. We plan to explore the effect

[1] www.google.com/trends/

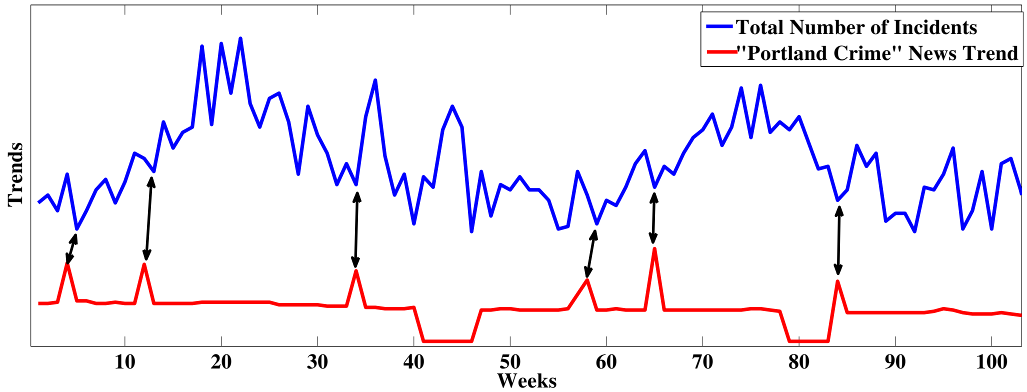

Figure 6: Comparison of trends in crime and popularity on Google News. Arrows suggest some relationship between popularity of "Portland Crime" on the news and dropping of crime rate.

of Social Media on crime in more depth. We also intend to incorporate factors such as traffic and socio-economic characteristics of neighborhoods to improve crime forecasting. We will also study the effect of these factors in predictions of different categories of crime by learning distinct models. Finding similarities among these models may demonstrate which categories of crime are related, thus not only improving prediction results but also providing insights into the nature crime.

ACKNOWLEDGMENTS

This work is supported by U.S. National Science Foundation under EAGER Award No.:CNS-1637372. The authors would like to thank Dr. Animesh Pathak for his insights and feedback.

REFERENCES

[1] Spencer Chainey, Lisa Tompson, and Sebastian Uhlig. 2008. The utility of hotspot mapping for predicting spatial patterns of crime. *Security Journal* 21, 1-2 (2008), 4–28.

[2] John Eck, Spencer Chainey, James Cameron, and R Wilson. 2005. Mapping crime: Understanding hotspots. (2005).

[3] Timothy Hart and Paul Zandbergen. 2014. Kernel density estimation and hotspot mapping: examining the influence of interpolation method, grid cell size, and bandwidth on crime forecasting. *Policing: An International Journal of Police Strategies & Management* 37, 2 (2014), 305–323.

[4] Arthur E Hoerl and Robert W Kennard. 1970. Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics* 12, 1 (1970), 55–67.

[5] Joel M Hunt. 2016. *Do crime hot spots move? Exploring the effects of the modifiable areal unit problem and modifiable temporal unit problem on crime hot spot stability.* Ph.D. Dissertation. AMERICAN UNIVERSITY.

[6] Shane D Johnson and Kate J Bowers. 2004. The burglary as clue to the future: The beginnings of prospective hot-spotting. *European Journal of Criminology* 1, 2 (2004), 237–255.

[7] M Vijaya Kumar and C Chandrasekar. 2011. Spatial clustering simulation on analysis of spatialtemporal crime hotspot for predicting crime activities. *International Journal of Computer Science and Information Technologies* 2, 6 (2011), 2864–2867.

[8] Patrick AP Moran. 1948. The interpretation of statistical maps. *Journal of the Royal Statistical Society. Series B (Methodological)* 10, 2 (1948), 243–251.

[9] National Institute of Crime. Real-Time Crime Forecasting Challenge. 2016. (2016).

[10] Robert Tibshirani. 1996. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)* (1996), 267–288.

[11] Michael Townsley, Ross Homel, and Janet Chaseling. 2003. Infectious burglaries: A test of the near repeat hypothesis. *The British Journal of Criminology* (2003), 615–633.

[12] Xiaofeng Wang, Matthew S Gerber, and Donald E Brown. 2012. Automatic crime prediction using events extracted from twitter posts. In *International Conference on Social Computing, Behavioral-Cultural Modeling, and Prediction*. Springer, 231–238.

[13] Chung-Hsien Yu, Wei Ding, Ping Chen, and Melissa Morabito. 2014. Crime forecasting using spatio-temporal pattern with ensemble learning. In *Pacific-Asia Conference on Knowledge Discovery and Data Mining*. Springer, 174–185.