Rotational grid, PAI-maximizing crime forecasts

George Mohler
Indiana University - Purdue University Indianapolis
PredPol

Michael D. Porter
University of Alabama
PASDA

Abstract

Crime forecasts are sensitive to the spatial discretizations on which they are defined. Furthermore, while the Predictive Accuracy Index (PAI) is a common evaluation metric for crime forecasts, most crime forecasting methods are optimized using maximum likelihood or other smooth optimization techniques. Here we present a novel methodology that jointly i) selects an optimal grid size and orientation and ii) learns a scoring function with the aim of directly maximizing PAI. Our method was one of the top performing submissions in the 2017 NIJ Crime Forecasting challenge, winning 9 of the 20 PAI categories under the name of team PASDA. We illustrate the model on data provided through the competition from the Portland Police Department.

1 Introduction

A number of statistical models have been proposed for "predicting" or "forecasting" the locations of crime hotspots including multivariate regression based models [19, 13, 11, 5], kernel density estimation [2, 4, 9, 8, 6] and spatio-temporal point processes [17, 15, 16]. In a typical crime forecast, geographical space is divided into sub-regions that are scored and/or ranked

over a forecasting time window. Forecasting models estimate a probability or crime rate within a given spatial region using predictors derived from past crime history (times and locations) or other spatial data defined in the spatial sub-regions requiring a forecast (e.g. census variables, locations of liquor stores and other crime attractors, parolee information). The implications for policing are both short-term, where spatial regions are ranked by a forecast and patrols and other interventions can be directed to the highest ranked spatial regions, and long-term, where the estimated crime rate may help inform resourcing decisions and the design of patrol beats.

The National Institute of Justice (NIJ) hosted a "Real-time crime forecasting challenge" in 2017 aimed at spurring further interest and research in this domain. The Portland Police Department provided crime data from March 2012 up to end of February 2017 and participants where asked to forecast crime hotspots for four types of incidents (burglary, motor vehicle theft, street crime, and all calls for service) over the months of March, April and May of 2017. In particular, participants were asked to define a grid subject to area and geometrical constraints and to rank grid cells for each crime type over several forecasting windows. Unlike forecasting research focusing on retrospective analysis, this competition was a true prospective forecasting test given that the validation data was not yet generated at the time of submission. Forecasts were made for 1-week, 2-week, 1-month, 2-month and 3-month time windows and scored on the basis of the PAI accuracy metric (which we define below).

In this article we provide an overview of the competition and present our method that won 9 out of 20 PAI categories in the large business division of the competition (and was the top performing solution in terms of PAI across all three competition divisions). The method jointly i) selects an optimal rectangular grid cell size and orientation and ii) learns a scoring function with the aim of directly maximizing PAI. The outline of the paper is as follows. In Section 2, we provide details on the contest including the data used, the submission guidelines and the evaluation metrics. In Section 3, we present our Rotational Grid PAI-Maximizing (RGPM) methodology and also outline the feature engineering and models we used within the RGPM framework. In Section 4, we analyze the results of the competition and the accuracy of the RGPM model and in Section 5, we include a discussion of the competition and some directions for future research.

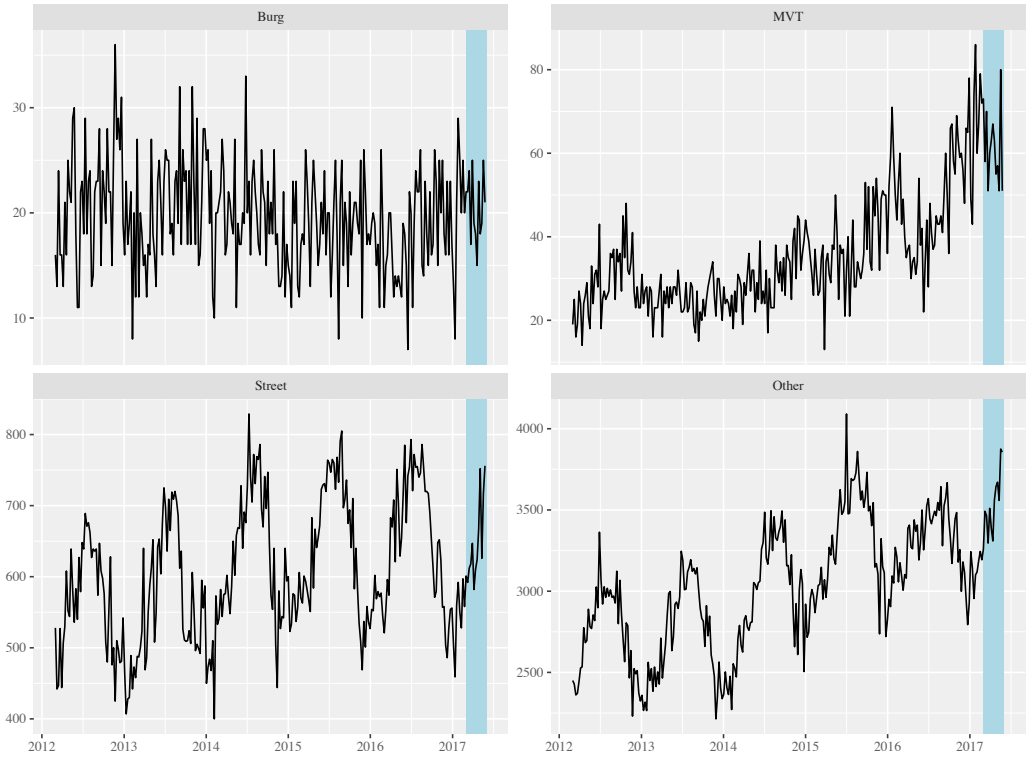

Figure 1: Weekly crime counts during the study period. The predictions were evaluated during March 1, 2017 - May 31, 2017 (highlighted in blue).

2 Data and Contest Details

The NIJ contest was based on forecasting the spatial locations for crime related call for service in Portland, OR. Specifically, the contestants were given event data comprising projected geographic coordinates, date, and category (burglary, street crime, theft of auto, other) for the period of March 1, 2012 through February 28, 2017. The weekly event counts during the training and evaluation periods are given in Figure 1.

Separate forecasts were made for 4 event types: burglary (Burg), street crime (Street), theft of auto (MVT), and all calls for service (ACFS) and 5 forecast horizons: 1 week (March 1-7), 2 weeks (March 1-14), 1 month (March 1-31), 2 months (March 1-April 30), and 3 months (March 1-May 31). The submitted forecast was specified to be a set of regular grid cells that covered all of the study region with some of the cells flagged as a "hotspot".

The grid cells were required to be a regular tessellation of the Portland, OR administrative region in which all grid cells must have the same size, shape, and orientation. Rectangles, triangles, and hexagons were the permitted grid shapes. Furthermore, the grid cells were required to have an area between 62,500 ft^2 and 360,000 ft^2 with the smallest dimension being at least 125 ft. The cells flagged as hotspots were required to have aggregate area between 0.25 mi^2 - 0.75 mi^2, but there was no requirement that the hotspot cells be connected.

Forecast evaluation was based on[1] the Prediction Accuracy Index (PAI) [4]. Given a set of k predicted hotspot cells, the PAI is determined by computing the ratio of the proportion of crime captured in the hotspots relative to the porportion area of the city flagged as hotspots. Specifically, defining H to be the union of the hotspot cells (which does not need to be connected) and S the spatial region of interest (e.g. Portland, OR), the PAI is defined as

$$\text{PAI}(H) = \frac{N(H)}{|H|}\frac{|S|}{N(S)}$$

where $N(H)$ is the number of events in H over the forecasting window and $|H|$ is the size of the hotspot region $H \subset S$. Letting $\lambda(H) = N(H)/|H|$ be the estimated intensity of events in region H and $\bar{\lambda} = N(S)/|S|$ be the total intensity of events in the region of interest, the PAI becomes

$$\text{PAI}(H) = \frac{\lambda(H)}{\bar{\lambda}} \propto \lambda(H)$$

which is only a function of $\lambda(H)$ because $\bar{\lambda}$ is not dependent on the hotspot region. Thus, PAI can also be interpreted as the average rate of crime in the hotspots relative to the average crime rate in the city.

On February 28, 2017 the forecasts (i.e., grid cells with hotpot indicator) for each crime type and forecast period (20 total) were submitted to NIJ.

3 Methodology

The RGPM methodology is designed for jointly learning an optimal grid and scoring function for the purpose of maximizing PAI in crime forecasts.

[1]The Prediction Efficiency Index (PEI) [7] was also used to evaluate forecasts, but we did not attempt to compete in that category. We include some remarks on PEI in the discussion in Section 5.

In particular, we assume a grid of equally sized rectangles and fix the grid cell size to be the minimum allowed in the competition, A_{min}. We then parametrize the grid with three parameters: cell height h, a grid translation parameter γ and a rotation angle θ. The overall procedure is captured in Algorithm 1.

Algorithm 1 Optimal rotational grid PAI maximizing methodology

1: **procedure** PAI(h,θ,γ,$\vec{x}_i$,t_i,ω,A_{min})

a. Set up grid with cell height h, cell area A_{min}, grid angle θ, and offset γ.

b. Calculate event based features on grid using crime locations $\vec{x}_i$ and times t_i.

c. Fit a supervised model $\mathcal{M}$, using tuning parameters ω, on event features defined on the training set.

d. Predict $\mathcal{M}$ on test data features and output PAI.

return PAI

2: **procedure** OPTIMIZEGRID($\vec{x}_i$,t_i,ω,A_{min})

Run simplex method to maximize PAI$(h, \theta, \gamma, \vec{x}_i, t_i, \omega, A_{min})$ over h, θ, and γ.

return h, θ, and γ.

In the outer loop of the algorithm, procedure 2, a simplex method is used to optimize the RGPM with respect to the grid parameters, given that PAI is a non-differentiable function. In the inner loop of the algorithm, procedure 1, PAI is calculated on a test data set given a grid parameter set, supervised learning model $\mathcal{M}$ and a set of features computed on the grid. In the competition, we used two different supervised learning algorithms for $\mathcal{M}$ that are outlined below.

3.1 RGPM: random forest

The first model we consider utilizes a regression framework and a random forest to map features to a target variable and we refer to this model as RGPM-RF. In particular, given a spatial discretization of the city, a number of features are defined within each spatial grid cell along with a target variable that the model attempts to predict. In the competition we used the following event count based features defined in each cell:

1. Number of target crime incidents 0 to 2 months before the forecasting window
2. Number of target crime incidents 3 to 5 months before the forecasting window
3. Number of target crime incidents 6 to 14 months before the forecasting window
4. Number of target crime incidents more than 14 months before the forecasting window
5. Number of target crime incidents before the forecasting window in March through May
6. Features 1 through 5 defined for each of the other three event categories (leading indicators)

The target variable was then the logarithm of the number of crimes of the target event type in a given cell over the forecasting time window and the distribution was assumed to be Gaussian.

Given the features and response variables defined on a fixed grid using historical data, we then estimated a random forest model [3] to map the features to the response. Random forests are a type of ensemble learning algorithm where many individual decision trees are estimated on bootstrap samples of the training data. While each decision tree over-fits a particular sample, when averaged together they comprise a random forest with good variance reduction properties.

3.2 RGPM: sparse logistic regression with point process features

The second model for $\mathcal{M}$ we consider is a sparse logistic regression, RPGM-GLM, where the features are determined by marked point process kernels. Let $(t_i, \vec{x}_i, m_i)$ denote an event i in the history leading up to a forecasting window where t_i is the event time, $\vec{x}_i$ is the 2-dimensional event spatial location, and m_i is the crime type of event i. Given a specific grid configuration $\mathcal{C}$, we can find the grid cell containing each event, which we denote with c_i.

Instead of modeling the event rate in each cell, we model the probability that the event rate (or equivalently the number of events) exceeds a threshold.

The threshold is set so that if the event count in the cell reaches the threshold then the cell would be part of the *optimal* hotspot region. Using the historical event data, we found the threshold, $\phi(\tau, m)$, that a grid cell would need to be part of the optimal hotspot region (subject to the minimum size constraints) for a forecast period of length τ and crime type m.

Let $p_j(\tau, m) = P(n_j(\tau, m) \geq \phi(\tau, m))$ be the probability that the number of events, n_j, in grid cell j exceeds the hotspot threshold. A multivariate/marked Hawkes process [15] or exciting point process is used to model the probability that a grid cell will be included in the hotspot.

For a given τ and m, we model the logit of the probability that cell j is in the hotspot using the features from a marked Hawkes process [15]:

$$\begin{aligned} \text{logit}(p_j(\tau, m)) &= \alpha + \sum_{i:c_i=j} h_{m,m_i}(t - t_i; \vec{\beta}, \vec{\theta}) \\ &= \alpha + \sum_{i:c_i=j} \sum_{k=1}^{K} \beta_k(m_i, m) g(t - t_i; \theta_k) \\ &= \alpha + \sum_{l} \sum_{k} \beta_k(l, m) Z_{jl}(\theta_k) \end{aligned} \tag{1}$$

where t is the start of the forecast period. The decay function $h_{m,m_i}(u) = \sum_{k=1}^{K} \beta_k(m_i, m) g(u; \theta_k)$ is a mixture of K geometric pmfs (i.e., $g(u; \theta) = \theta(1-\theta)^{u-1}$), We used the pre-specified values $\vec{\theta} = .001, .005, .01, .02, .1$ to provide a range of decay effects. The last line of (1) shows that this is in the form of a GLM model (logistic regression) where $Z_{jl}(\theta_k) = \sum_i 1(c_i = j, m_i = l)\, g(t - t_i; \theta_k)$ are the covariates formed from geometric point process kernels evaluated at historical event times. The model parameters $\vec{\beta}$ are estimated using an elastic net penalty [20] with $\alpha = 0.8$, a mix between ridge and lasso penalties which can provide a sparse solution. The correct penalty strength was determined from 10-fold cross-validation. The resulting fitted decay functions, $\hat{h}_{m,l}(\cdot)$, are shown in Figure 2.

4 Results

To construct the 2017 forecasts, we first trained the RGPM-RF and RGPM-GLM models on data up to the spring of 2016 and then evaluated the models

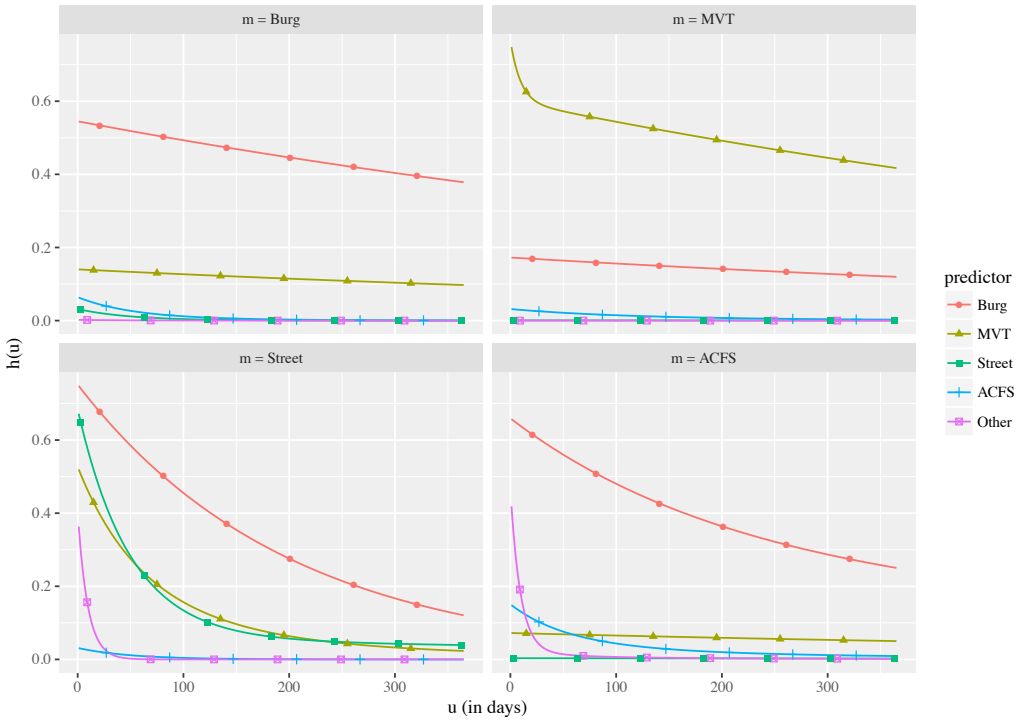

Figure 2: The decay functions $h_{m,l}(\cdot)$ for predicting crime type m using the past events of type l.

with the spring 2016 data to perform model comparison and select the best model and grid for each category. We then used the date up to February 28, 2017 to build the forecast models and generate the hotspot estimates. Table 1 displays the performance of the models, for the 20 PAI categories, during the contest period. The table also shows the difference in PAI scores between the two models and which model was submitted to NIJ. The random forest (RF) tended to do better for the long-range forecasts (2MO and 3MO), whereas the sparse GLM scored better for the short term forecasts. However, while both models produced similar PAI scores, they did so using different grid configurations.

In Table 2 we show the feature importances of the RGPM-RF for each target event category. For high volume incident types (street crime and all), we note that leading indicators play less of a role and the 5 best features are the count features for that particular event category. However, for lower volume incident types leading indicators are more important as they serve to reduce variance. The features created from the *Other* crime type were useful in predicting burglary and motor vehicle theft, for example. In Table 3 we show the feature importances of the RGPM-GLM. There are some similarities, for example the most recent data (e.g., large values of θ) are not always the most important and the *Other* and *ACFS* leading indicators are useful for forecasting the lower volume crime types. Here we see some differences, for example burglary is a stronger predictor of burglary than in the random forest.

In Table 4 we compare the relative PAI values when using fixed vs. rotational grids. For this purpose we use the RGPM-RF, though we note similar improvements are observed for the RGPM-GLM. The PAI values increase by 2-6 when employing a rotating grid. In Figure 3 we provide an example of the final street crime grid used in the competition, which uses rectangular cells aligned with the NE-SW direction.

Finally, in Table 5 we include overall competition results illustrating the accuracy of the RGPM approach. In the table we list the number of overall (across the three divisions) 1st, 2nd and 3rd place PAI finishes for teams having placed at least once. We note that the RGPM tied for the most 1st and 2nd place finishes and had the most 3rd place finishes across the crime type categories and forecasting windows. We also include in Table 5 the total number of finishes (3rd place and higher) within our division (large business) and overall, in both cases the RGPM method had the most finishes.

Table 1: RGPM-RF vs RGPM-GLM model performance during the competition period. n is the number of events in the hotspot, N is the total number of events during the time frame, δ_{PAI} is the difference in PAI between the two models, and an asterisk next to the model indicates the model which was submitted to the competition.

Crime	Time	n	N	PAI	Model	δ_{PAI}
Burg	1WK	1	20	29.5	RF	0.1
Burg	2WK	2	41	28.7	GLM	14.3
Burg	1MO	4	93	25.3	*GLM	6.3
Burg	2MO	6	175	20.2	RF	3.4
Burg	3MO	8	268	17.6	RF	0.1
MVT	1WK	3	71	24.9	*GLM	0.1
MVT	2WK	12	135	52.2	*RF	4.1
MVT	1MO	21	273	45.4	*GLM	0.2
MVT	2MO	41	543	44.5	*GLM	5.6
MVT	3MO	55	805	40.3	GLM	3.1
Street	1WK	99	629	92.8	GLM	1.1
Street	2WK	185	1205	90.6	GLM	1.7
Street	1MO	405	2680	89.1	GLM	1.3
Street	2MO	771	5352	84.8	*RF	1.2
Street	3MO	1253	8480	87.0	*RF	2.9
ACFS	1WK	392	3876	59.5	GLM	1.6
ACFS	2WK	814	8021	59.7	GLM	0.1
ACFS	1MO	1805	17873	59.4	GLM	0.2
ACFS	2MO	3556	35770	58.5	*RF	0.9
ACFS	3MO	5570	55744	58.8	*RF	1.9

Table 2: Feature importances for the RGPM-RF for each of the four target crime types (top 5 in bold)

Feature	Street	Burg	MVT	ACFS
count1_other	1195.1	34.8	80.6	**59621.0**
count2_other	1839.3	**40.1**	94.0	**80035.2**
count3_other	2177.2	**50.9**	**126.9**	**97239.9**
count4_other	2259.8	**60.1**	**149.6**	**87000.8**
count5_other	1836.7	**47.4**	**121.0**	**76231.5**
count1_street	**2586.2**	16.4	37.7	15026.5
count2_street	**2833.1**	20.5	49.6	16556.2
count3_street	**5167.8**	32.9	78.1	30617.3
count4_street	**4379.3**	**44.4**	101.8	27493.2
count5_street	**3191.6**	28.5	66.9	18651.6
count1_mvt	81.8	2.2	47.7	677.2
count2_mvt	91.6	5.6	96.0	1008.9
count3_mvt	174.7	9.0	99.2	5234.8
count4_mvt	277.9	12.4	**185.5**	5152.1
count5_mvt	144.3	5.2	**157.7**	2381.2
count1_burglary	33.7	2.7	2.4	279.7
count2_burglary	119.1	4.8	4.6	587.9
count3_burglary	127.6	9.7	13.2	1007.2
count4_burglary	214.9	10.7	18.6	1777.3
count5_burglary	85.7	4.5	9.9	778.5

Table 3: Feature importances for the RGPM-GLM for each of the four target crime types. The importance score is the estimated coefficient using standardized covariates.

Feature (crimetype_θ)	Street	Burg	MVT	ACFS
ACFS_0.1	-	-	-	-
ACFS_0.02	0.07	0.11	-	0.19
ACFS_0.01	-	0.06	0.08	0.12
ACFS_0.005	-	-	0.11	0.21
ACFS_0.001	-	-	-	0.20
other_0.1	0.15	-	-	0.17
other_0.02	-	-	-	0.03
other_0.01	-	-	-	-
other_0.005	-	-	-	0.04
other_0.001	-	-	-	0.02
street_0.1	-	-	-	-
street_0.02	0.31	0.02	-	-
street_0.01	-	-	-	-
street_0.005	-	-	-	-
street_0.001	0.38	-	-	0.02
mvt_0.1	-	-	0.01	-
mvt_0.02	0.02	-	-	-
mvt_0.01	0.03	-	-	-
mvt_0.005	0.01	-	-	-
mvt_0.001	0.01	0.06	0.27	0.03
burglary_0.1	-	-	-	-
burglary_0.02	-	-	-	-
burglary_0.01	-	-	-	-
burglary_0.005	0.07	-	-	0.03
burglary_0.001	-	0.11	0.04	0.06

Table 4: PAI values for RGPM-RF vs fixed grid random forest

Method	Street	Burglary	MVT	All
RGRF	84.99	18.12	34.90	61.26
RF	78.75	13.59	32.57	59.12

Table 5: Aggregate number of 1st, 2nd and 3rd place PAI finishes across divisions along with total number of overall 3rd and higher finishes (A) and number of 3rd and higher finishes within division (B).

Name	1st	2nd	3rd	A	B
PASDA	4	5	4	13	20
TAMERZONE	4	5	2	11	15
GRIER	1	4	0	5	8
JeremyHeffner	2	0	3	5	9
ANDY_NIJ	1	2	1	4	9
KUBQR1	0	1	3	4	7
pennaiken	2	0	2	4	10
Codilime	3	0	0	3	7
MURRAYMIRON	0	1	2	3	6
MARUANALSHEDIVAT	0	1	1	2	7
BATESANALYTICS	1	0	0	1	1
DYLANFITZPATRICK	0	0	1	1	1
GRANTHAM	0	0	1	1	7
Intuidex	0	1	0	1	3
TEAM_Kernel_Glitches	1	0	0	1	2
WARREN	1	0	0	1	20

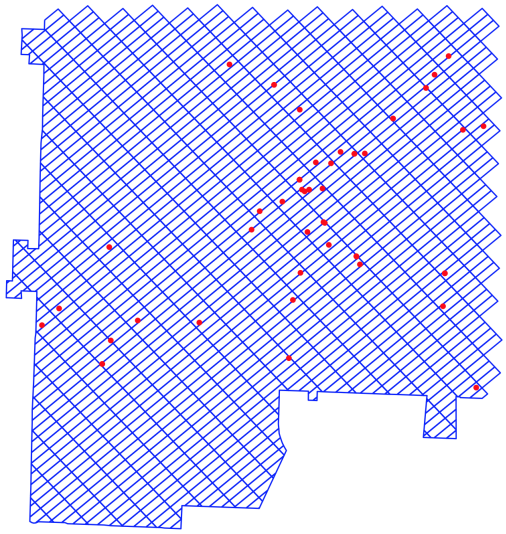

Figure 3: Optimal street crime grid with street crime events plotted for the 2017 forecasting window.

5 Conclusion

We provided a summary of the 2017 NIJ Real-time crime forecasting challenge and our top performing approach. Several take aways include i) the spatial units and grid on which predictions are defined are a key component to maximizing PAI ii) directly maximizing PAI can also improve accuracy, especially for large volume incident types. These observations are in line with several recent studies [18, 14] that move beyond the use of arbitrary grids in crime forecasting and prediction. While our focus was on optimizing PAI, PEI was another metric used in the competition. PEI is the ratio of PAI to the maximum possible PAI a model could have achieved in the competition. There appears to be some trade-off between the two metrics, as the other top performing team, Kernel Glitches, in the large business category won a number of the PEI categories. We believe a method like the one outlined here can be tuned to optimize a particular metric of interest, including PEI. In practice, given a fixed number of hotspots and police resources, PAI and PEI are equivalent as they only diverge if the size of hotspot region can vary. Thus the trade-off is not often of practical consequence. However, other accuracy metrics (for example precision, the percentage of hotspots having a

crime each day) may be useful to consider, as false positives can have the effect of decreasing officer buy-in to predictive policing.

While the methodology introduced here improved accuracy of crime forecasts, greater improvements in the future may be possible with the greater prevalence of sensor data provided by the internet of things in smart and connected cities. For example, images and video of city locations may provide features related to "broken windows" crime patterns [10] and some recent work has shown that such imagery may help estimate crime rates [12]. Pedestrian and vehicle count data provides real time information on the density of targets for various types of violent and property crime and other recent research has shown that this type of data can also improve predictive models of crime [1]. The methods we have introduced in this paper may be used in combination with new sensor streams to improve crime forecasts in the future.

References

[1] Andrey Bogomolov, Bruno Lepri, Jacopo Staiano, Nuria Oliver, Fabio Pianesi, and Alex Pentland. Once upon a crime: towards crime prediction from demographics and mobile data. In *Proceedings of the 16th international conference on multimodal interaction*, pages 427–434. ACM, 2014.

[2] Kate J Bowers, Shane D Johnson, and Ken Pease. Prospective hotspotting the future of crime mapping? *British Journal of Criminology*, 44(5):641–658, 2004.

[3] Leo Breiman. Random forests. *Machine learning*, 45(1):5–32, 2001.

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

[5] Jacqueline Cohen, Wilpen L Gorr, and Andreas M Olligschlaeger. Leading indicators and spatial interactions: A crime-forecasting model for proactive police deployment. *Geographical Analysis*, 39(1):105–127, 2007.

[6] Matthew Fielding and Vincent Jones. 'disrupting the optimal forager': predictive risk mapping and domestic burglary reduction in trafford, greater manchester. *International Journal of Police Science & Management*, 14(1):30–41, 2012.

[7] Joel M Hunt. *Do crime hot spots move? Exploring the effects of the modifiable areal unit problem and modifiable temporal unit problem on crime hot spot stability*. PhD thesis, American University, 2016.

[8] Shane D Johnson. *Prospective crime mapping in operational context: Final report*.

[9] Shane D Johnson, Kate J Bowers, Dan J Birks, and Ken Pease. Predictive mapping of crime by promap: accuracy, units of analysis, and the environmental backcloth. In *Putting crime in its place*, pages 171–198. Springer, 2009.

[10] Kees Keizer, Siegwart Lindenberg, and Linda Steg. The spreading of disorder. *Science*, 322(5908):1681–1685, 2008.

[11] Leslie W Kennedy, Joel M Caplan, and Eric Piza. Risk clusters, hotspots, and spatial intelligence: risk terrain modeling as an algorithm for police resource allocation strategies. *Journal of Quantitative Criminology*, 27(3):339–362, 2011.

[12] Aditya Khosla, Byoungkwon An An, Joseph J Lim, and Antonio Torralba. Looking beyond the visible scene. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3710–3717, 2014.

[13] Hua Liu and Donald E Brown. Criminal incident prediction using a point-pattern-based density model. *International journal of forecasting*, 19(4):603–622, 2003.

[14] G. Mohler. Learning to rank spatio-temporal event hotspots. 2017.

[15] George Mohler. Marked point process hotspot maps for homicide and gun crime prediction in chicago. *International Journal of Forecasting*, 30(3):491–497, 2014.

[16] George O Mohler, Martin B Short, Sean Malinowski, Mark Johnson, George E Tita, Andrea L Bertozzi, and P Jeffrey Brantingham. Randomized controlled field trials of predictive policing. *Journal of the American Statistical Association*, 110(512):1399–1411, 2015.

[17] GO Mohler, MB Short, P Jeffrey Brantingham, FP Schoenberg, and GE Tita. Self-exciting point process modeling of crime. *Journal of the American Statistical Association*, 106(493):100–108, 2011.

[18] Gabriel Rosser, Toby Davies, Kate J Bowers, Shane D Johnson, and Tao Cheng. Predictive crime mapping: Arbitrary grids or street networks? *Journal of Quantitative Criminology*, pages 1–26, 2016.

[19] Xiaofeng Wang, Donald E Brown, and Matthew S Gerber. Spatio-temporal modeling of criminal incidents using geographic, demographic, and twitter-derived information. In *Intelligence and Security Informatics (ISI), 2012 IEEE International Conference on*, pages 36–41. IEEE, 2012.

[20] Hui Zou and Trevor Hastie. Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 67(2):301–320, 2005.