

A simple crime hotspot forecasting algorithm

Robert Bogucki

University of Warsaw

ul. Banacha 2, 02-097 Warsaw, Poland

and

deepsense.ai,

Al. Jerozolimskie 162A, 02-342 Warsaw, Poland

Email: robert@deepsense.ai

Jan Kanty Milczek, Patryk Miziula

deepsense.ai,

Al. Jerozolimskie 162A, 02-342 Warsaw, Poland

Email: {jan.milczek, patryk.miziula}@deepsense.ai

Abstract—Crime hotspot forecasting is an important part of crime prevention and reducing the delay between a 911 call and the physical intervention. Current developments in the field focus on enriching the historical data and sophisticated point process analysis methods with a fixed grid. In the paper we present a simple spatio-temporal point process allowing one to perform exhaustive (literal) grid searches. We then show that this approach can compete with more complex methods, as evidenced by the results on data collected by the Portland Bureau of Police. Finally, we discuss the advantages and potential implications of the new method.

I. INTRODUCTION

SPATIO-TEMPORAL crime forecasting is a field that grabs the attention of both scientists and practitioners. Many academic researchers have published results based on time series analysis ([1]), regression methods ([2], [3], [4]), kernel density estimation ([5], [6], [7], [8], [9], [10]) or self-exciting point processes ([11], [12], [13], [14], [15], [16], [17]). Moreover, the US Government appreciates the impact predictive policing has on society (see [18]).

In a typical crime prediction task, the forecast area is fixed and divided into small sub-regions, called cells. The cells are then scored separately over a given future time window. The ones with the highest rate are chosen as the most dangerous areas and called hotspots. In this article we present a point of view for hotspot forecasting that differs from those which can be found in the literature. We emphasise the simplicity and efficiency of our algorithm for a fixed grid to get an opportunity to check as many grids as possible. We place those attributes over sophisticated methods, with state-of-the-art results in practice. Our models won eight categories of the Real-Time Crime Forecasting Challenge conducted by the National Institute of Justice ([19]).

The rest of the paper is organized as follows. In section II we explain our approach in detail. Section III contains a comprehensive description of case study of our method – the Real-Time Crime Forecasting Challenge. Further comments and summary are placed in section IV.

II. THE MODEL

A. The choice of grid

There is a vast literature available about crime forecasting for a given grid of cells based on past crimes committed

(see references in the Introduction). In such a setup, more or less sophisticated methods are applied to predict which fixed parts of the investigated region will experience the highest future rate of crime. Clearly, changing the grid changes the entire task as well and may lead to completely different predictions with different levels of effectiveness in the real world. However, as far as we know, whenever the cell division is not imposed in advance, searching for a good grid is in practice reduced to *grid search*, *random search* (see [20]) or another primitive method of walking among parametrizations of possible tessellations. The reason there is a lack of 'smarter' grid choosing techniques may be that spatial distributions of crimes committed in urban areas are 'weird': they contain atoms with very high crime rates (related to, for example, large-area stores or shelters for the homeless). Therefore, using the same data-driven algorithm for even very similar grids can cause a huge discrepancy in the qualities of the predictions obtained. Hence, grid optimization cannot be neglected.

Taking into consideration the massive number of grids worth checking we concluded there was a need for a very fast but still well performing supervised model for a fixed grid, one that would simply execute a random search on a rich space of grid parameterizations to find the 'optimal' grid. This would yield a better final result than a more sophisticated, but slower algorithm applied to a random set of grids that would be too small to contain any decent tessellation.

B. Fast algorithm for a given grid

The main idea behind our algorithm for a fixed grid is simple: count the past crimes in every cell and mark the cells with 'the worst past' as hotspots. In other words, we assume that if many crimes occurred somewhere, more are likely to happen. This principle may strike some as naive and outdated, but we believe that it is both accurate enough and fast. Up-to-date crime registries are freely available for several US cities. They form the main dataset in data-driven crime forecasting algorithms. One can search for any external data which could affect future crimes, but have not left a trace on those crimes that have already been committed. We are aware that weather, demographics and even social media information (see [4]) are sometimes used in similar contexts. Unfortunately, they significantly increase the model's

complexity, often without a guarantee of noticeably improved accuracy. Keeping computations as simple as possible, by using merely historical crime data, enables us to spend more time on selecting the right grid.

We refine the raw algorithm by taking care of data aging and seasonality. Namely, we assign weights to all the past crimes and then sum up the weights of all the crimes in consecutive cells to find the hotspots. The weight of an event decreases exponentially as a function of age (in days) of a crime. The intensity of the decrease is a hyperparameter. Also, we boost the weights of crimes committed on the same days of the year as those in the forecasted time span. The power of boosting is a hyperparameter as well.

Moreover, we introduce a primitive 'spatial radiation' of past crimes. For each data point, we put eight of its copies with reduced weights in the corners and in the center of the sides of the rhombus around it (see Figure 1). In this way, a 'part' of an event that has occurred close to the cell border could fall into a neighboring cell. We chose to use a rhombus because it reflects the Manhattan metric, a reasonable match for North-South-oriented axis grid street plans, of which there are many in US cities. The size of the rhombus and reduction of weights of added copies are hyperparameters.

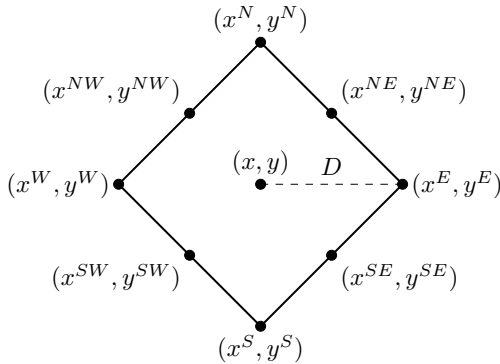


Fig. 1. Points on the rhombus around given point.

The approach presented here can be expressed in the language of spatio-temporal point processes (cf., e.g., [14] and references therein). Consider a counting process $N(t, x, y)$ characterized by its conditional intensity function λ . In our case we define lambda as (1), where

- $\mu_{\text{age}}(t_j, t) = e^{-A(t-t_j)}$,
- $\mu_{\text{seas}}(t_j) = 1 + B \cdot \nu(t_j)$ for $\nu(t_j) = 1$ if t_j is the same day of year as those placed in the forecasted timespan and $\nu(t_j) = 0$ elsewhere,
- $\varphi(x_j, y_j, x, y) = \begin{cases} 1, & \text{if } (x_j, y_j) \text{ and } (x, y) \\ & \text{are in the same cell} \\ 0, & \text{elsewhere,} \end{cases}$
- $\varphi_{\text{blur}}(x_j, y_j, x, y) = C \cdot \sum_{\{i \in \mathcal{D}\}} \varphi(x_j^i, y_j^i, x, y)$ for $\mathcal{D} = \{N, NE, E, SE, S, SW, W, NW\}$ and (x^i, y^i) defined as in Figure 1, i.e., $(x^N, y^N) = (x, y + D)$, $(x^{NE}, y^{NE}) = (x + D/2, y + D/2)$, $(x^E, y^E) = (x + D, y)$, etc.

We sum through all the past events (t_j, x_j, y_j) with $t_j < t$. A, B, C, D are hyperparameters. Our lambda is much simpler than those found in the literature. We need neither smoothing nor symmetry properties. Also, for a fixed t , $\lambda(t, x, y) = \lambda(t, x', y')$ for every (x, y) and (x', y') lying in the same cell. Hence, we can think about λ as of the intensity of the entire cell and simply choose cells with the greatest values of λ as hotspots.

C. Validation

To find the best grid and hyperparameter values, we split the dataset into training, validation and test parts in the following way: the last period becomes the test set, the second-to-last is treated as the validation set and all the earlier events make up the training set. Then we generate hotspots for different grids and hyperparameters using training data and compare them on validation data to choose the best settings. Finally, we compute the hotspots for the best model once more - this time with use of both training and validation data - and obtain the ultimate score using test data.

In classic crime forecasting, the score functions taken from the binary classification – ROC/AUC, sensitivity, etc. – are used (see [6]). There are also two newer functions on the market: predictive accuracy index (PAI, [6]) and prediction efficiency index (PEI, [21]) given by $PAI = \frac{n/N}{a/Ar}$, $PEI = \frac{n}{n^*}$, respectively, where:

- n - the number of future crimes in k proposed hotspots,
- n^* - the number of future crimes in k 'worst' cells,
- N - the number of all future crimes in the entire area,
- a - the total volume of k proposed hotspots,
- Ar - the volume of the entire area,

assuming that k cells were indicated as hotspots. They all have their disadvantages. Binary classification-based functions are inconvenient if the area of the hot-spots to be forecast is a very small fraction of the investigated jurisdiction, which is typical. As for other functions, PAI favors smaller single cell areas while PEI likes as great a single cell area as possible. For this reason it is impossible to maximize both PAI and PEI with the same grid, which casts doubt on the validity of using either of them. Moreover, PEI is bounded by 1 from above whereas the range of PAI is a positive half line, so they are not directly comparable. Nevertheless, our approach is metric-agnostic, therefore any reasonable score function can be applied here.

III. CASE STUDY

A. The competition

In September 2016, the National Institute of Justice in the US announced the *Real-Time Crime Forecasting Challenge*. The goal was to predict future crimes in Portland, Oregon. Contestants were asked to divide the area under Portland police jurisdiction (an area roughly 15 by 20 miles) into a grid of small cells (i.e., 250 by 250 feet) and indicate the cells that would have the highest future crime rate - hotspots. Several restrictions on the cells' shape and the total volume of hotspots were imposed.

$$\lambda(t, x, y) = \sum_{\{j: t_j < t\}} \mu_{\text{age}}(t_j, t) \cdot \mu_{\text{seas}}(t_j) \cdot [\varphi(x_j, y_j, x, y) + \varphi_{\text{blur}}(x_j, y_j, x, y)], \quad (1)$$

Four different categories of crime were considered separately: all crimes, burglaries, car thefts and street crimes (including assaults, robberies, shots fired). Five future time spans (starting in March 2017) were involved: one week, two weeks, a month, two months and three months. Hence, there were 20 type/time categories. In each of them, the predictions were compared against the actual state of affairs in Portland using both PAI and PEI. Thus, the competition consisted of $4 \cdot 5 \cdot 2 = 40$ separate sub-competitions in total. Only the best submission was awarded in each of them. Three independent tracks of the challenge were run simultaneously: intended for large businesses, small businesses and students, respectively. Each track had the same rules and goals, but separate contestants, winners and prizes.

B. Data

The NIJ delivered historical data on all the crimes registered in Portland between March 2012 and February 2017. Almost 1,000,000 records were provided in total. Each of them contained the day the crime was committed, coordinates (with accuracy to one foot) and the type of crime committed. There were no data gaps.

The distribution of data between crime categories was highly imbalanced: burglaries, car thefts and street crimes were only 0.5%, 1%, and 16.5% of records, respectively. Thus, we expected a huge discrepancy in the numbers of crimes committed between particular type/time categories between March and May 2017. That was true, two extreme cases were: all the crimes between March and May 2017 - 65,000 records, and burglaries in the first week of March 2017 - only 20 events.

Distributions of crimes in all the categories with a big enough number of events had similar characteristics: they consisted of the 'dense' part looking like a sample from a continuous distribution and the 'discrete' part made from atoms. It seems that although the accuracy of the coordinates of crimes committed was in general one foot, police officers tended to 'discretize' some areas like stores or shelters to a single spatial point next to the entrance to the building/area.

C. Computations

The first attempts showed that in each of the 20 type/time categories the PAI metric was maximized by a lot of small hotspots whereas PEI behaved best for a small number of large hotspots. Hence it was clear that we should not attempt to satisfy both metrics simultaneously. Since each metric formed an independent sub-competition with a separate prize, it was better to have a good score for one metric than mediocre results for both. So, for each of the 20 type/time categories we had to decide which metric to focus on in our further work. The metrics were incomparable, scores between the categories were incomparable and we did not know other competitors

and their results. Thus, we did not have any hitching point that would help us to choose a metric. Moreover, our approach was metric-agnostic. Hence, to choose a metric, we just tossed a coin for each of 20 type/time categories.

During the competition we were examining parallelogram, triangular and hexagonal grids. No shape proved noticeably better than other ones. We ultimately decided to only use unrotated rectangular grids, parameterized by cell height, width, horizontal and vertical shift. The number of predicted hotspots was also a hyperparameter. We optimized the grid and our model hyperparameters for each of 20 type/time categories separately.

D. Results

Table I gathers information about seven categories with the largest numbers of crimes committed during the test periods. Our predictions proved the most accurate in all of them in the contest track for large businesses. Moreover, all of those predictions remained on the top after comparing results from the competition's three tracks (for large business, small businesses and students). This was the best result among all the competitors, while the runner-up achieved four across-track wins.

TABLE I
COMPETITION CATEGORIES WITH THE LARGEST NUMBERS OF CRIMES COMMITTED.

category	number of crimes	metric used	metric value
all, 3 months	55744	PAI	60.53
all, 2 months	35770	PEI	0.989
all, 1 month	17873	PAI	61.37
street, 3 months	8480	PEI	0.967
all, 2 weeks	8021	PEI	0.957
street, 2 months	5352	PEI	0.940
all, 1 week	3876	PAI	62.35

One more category for which our prediction was the most effective in the large business competition (but not in the total rank) was for burglaries between March and May 2017. However, in our opinion the number of crimes committed, 268, was so low that no model would be able to credibly predict them, so our success was just a matter of luck. We would conclude the same about seven more categories: the other time periods for burglaries (175, 93, 41, and 20 crimes) and car thefts in a one-month period and less (273, 135, and 71 incidents).

The results allowed us to conclude that for both the PAI and PEI metrics we were able to find grids and hotspots with quality competing with predictions obtained by authors of more complicated methods described in the literature (cf. [22], [23]). Our approach proved especially effective in categories with the biggest number of crimes committed.

Since different competitors submitted different grids, we are unable to compare algorithms for a fixed grid created by particular contestants. Therefore, we cannot judge whether the good performance of our models was an effect of thoroughly scouring potential grids or the power of simplicity of our algorithm for a fixed grid, or perhaps both.

IV. DISCUSSION

The comparative case study on crime data from Portland, OR, shows that our computation time-oriented approach can compete with more sophisticated crime forecasting methods existing in the literature. This result is somewhat surprising. One may conclude that the spatio-temporal distribution of crimes committed is too complicated to be estimated well enough with the use of parametric methods. Or maybe the choice of the proper grid matters much more than it seems. Moreover, we have no reason to claim that the good performance of our algorithm is a one-shot success valid only for Portland since our model contains no part priorly adapted to any particular city. Unfortunately, we did not have the opportunity to compare the quality of crime forecasts done with use of different methods (including our own) for the same fixed grid. Such research would shed more light on this field.

The advantage of our algorithm for cases with thousands or more crimes to forecast can be attributed to two possible factors: a specific spatial distribution of crimes or computational simplicity. As stated above, for most statistical parametric methods it may be intractable to cover a distribution containing both a continuous and a discrete part. Comparing the performance of different models for a fixed grid would bear this out. On the other hand, sophisticated algorithms can paradoxically struggle to find the optimal grid and hotspots when presented with large volumes of training data. A time-consuming training procedure for a fixed grid does not allow one to check a sufficient number of potential grids. This problem may be addressed by more efficient algorithms' implementations and significantly increasing computing resources. Also, adding more constraints on the admissible grid shapes clearly solves the problem, though it also makes it less universal.

Finally, we note that in the perspective of maintaining and updating the crime forecasting system, using only the historical crime data seems to be a good solution. It is hard to find any non-constant external factor which can both influence future crimes and be easier to predict than crimes themselves. Besides, the impact of any hidden important feature is ultimately reflected in the historical data. Moreover, changes in the spatial crime distribution caused by system-driven preventive police activities may be not easy to manage when external data sources are used for forecasting. At the same time, a forecasting system based on merely historical data is able to simply retune to the current crime distribution.

REFERENCES

- [1] W. Gorr, A. Olligschlaeger, and Y. Thompson, "Short-term forecasting of crime," *International Journal of Forecasting*, vol. 19, pp. 579–594, 2003.
- [2] J. Cohen, W. L. Gorr, and A. M. Olligschlaeger, "Leading indicators and spatial interactions: A crime-forecasting model for proactive police deployment," *Geographical Analysis*, vol. 39, pp. 105–127, 2007.
- [3] L. W. Kennedy, J. M. Caplan, and E. Piza, "Risk clusters, hotspots, and spatial intelligence: Risk terrain modeling as an algorithm for police resource allocation strategies," *Journal of Quantitative Criminology*, vol. 27, pp. 339–362, 2011.
- [4] X. Wang, M. S. Gerber, and D. E. Brown, "Automatic crime prediction using events extracted from twitter posts," in *Social Computing Behavioral - Cultural Modeling and Prediction*. Springer Berlin Heidelberg, 2012, pp. 231–238.
- [5] K. J. Bowers, S. D. Johnson, and K. Pease, "Prospective hot-spotting: The future of crime mapping?" *The British Journal of Criminology*, vol. 44, pp. 641–658, 2004.
- [6] S. Chainey, L. Tompson, and S. Uhlig, "The utility of hotspot mapping for predicting spatial patterns of crime," *Security Journal*, vol. 21, pp. 4–28, 2008.
- [7] M. Fielding and V. Jones, "Disrupting the optimal forager: Predictive risk mapping and domestic burglary reduction in trafford, greater manchester," *International Journal of Police Science & Management*, vol. 14, pp. 30–41, 2012.
- [8] W. L. Gorr and Y. Lee, "Early warning system for temporary crime hot spots," *Journal of Quantitative Criminology*, vol. 31, pp. 25–47, 2015.
- [9] M. D. Porter and B. J. Reich, "Evaluating temporally weighted kernel density methods for predicting the next event location in a series," *Annals of GIS*, vol. 18, pp. 225–240, 2012.
- [10] M. A. Boni and M. S. Gerber, "Automatic optimization of localized kernel density estimation for hotspot policing," in *Proc. 15th IEEE Int. Conf. Machine Learning and Applications (ICMLA)*, Dec. 2016, pp. 32–38.
- [11] H. Liu and D. E. Brown, "Criminal incident prediction using a point-pattern-based density model," *International Journal of Forecasting*, vol. 19, pp. 603–622, 2003.
- [12] M. A. Taddy, "Autoregressive mixture models for dynamic spatial poisson processes: Application to tracking intensity of violent crime," *Journal of the American Statistical Association*, vol. 105, pp. 1403–1417, 2010.
- [13] G. Rosser and T. Cheng, "Improving the robustness and accuracy of crime prediction with the self-exciting point process through isotropic triggering," *Applied Spatial Analysis and Policy*, pp. 1–21, 2016.
- [14] G. O. Mohler, M. B. Short, P. J. Brantingham, F. P. Schoenberg, and G. E. Tita, "Self-exciting point process modeling of crime," *Journal of the American Statistical Association*, vol. 106, pp. 100–108, 2011.
- [15] G. Mohler, "Marked point process hotspot maps for homicide and gun crime prediction in chicago," *International Journal of Forecasting*, vol. 30, pp. 491–497, 2014.
- [16] G. O. Mohler, M. B. Short, S. Malinowski, M. Johnson, G. E. Tita, A. L. Bertozzi, and P. J. Brantingham, "Randomized controlled field trials of predictive policing," *Journal of the American Statistical Association*, vol. 110, pp. 1399–1411, 2015.
- [17] C. Loeffler and S. Flaxman, "Is gun violence contagious? A spatiotemporal test," *Journal of Quantitative Criminology*, 2017.
- [18] W. Perry, B. McInnis, C. Price, S. Smith, and J. Hollywood, "Predictive policing: The role of crime forecasting in law enforcement operations," RAND Corporation, Santa Monica, Tech. Rep., 2013.
- [19] National Institute of Justice, "Real-Time Crime Forecasting Challenge," <https://www.nij.gov/funding/Pages/fy16-crime-forecasting-challenge.aspx>, 2017. [Online]. Available: <https://www.nij.gov/funding/Pages/fy16-crime-forecasting-challenge.aspx>
- [20] L. A. Rastrigin, "About convergence of random search method in extremal control of multi-parameter systems," *Automation and Remote Control*, vol. 24, pp. 1467–1473, 1963.
- [21] J. 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," Ph.D. dissertation, 2016.
- [22] G. Mohler and M. D. Porter, "Rotational grid, PAI-maximizing crime forecasts," 2018, to appear in *Statistical Analysis and Data Mining*.
- [23] S. Flaxman, M. Chirico, P. Pereira, and C. Loeffler, "Scalable high-resolution forecasting of sparse spatiotemporal events with kernel methods: a winning solution to the nij "Real-Time Crime Forecasting Challenge," 2018.