Learning to rank spatio-temporal event hotspots

George Mohler
Department of Comuter and Information Science
Indiana University - Purdue University Indianapolis
Email: gmohler@iupui.edu

***Abstract*—Crime, traffic accidents, terrorist attacks, and other space-time random events are unevenly distributed in space and time. In the case of crime, predictive policing algorithms aim to focus limited resources at the highest risk crime hotspots in a city. A crucial step in the implementation of these strategies is the construction of scoring models used to rank spatial hotspots. While these methods are evaluated by area normalized Recall@k (called the Predictive Accuracy Index), models are typically trained via maximum likelihood or rules of thumb that may not prioritize model accuracy in the top k hotspots. Furthermore, current algorithms are defined on fixed grids that fail to capture risk patterns occurring in neighborhoods and on road networks with complex geometries. We introduce CrimeRank, a learning to rank boosting algorithm for determining a crime hotspot map that directly optimizes the percentage of crime captured by the top ranked hotspots. The method also employs a floating grid combined with a greedy hotspot selection algorithm for accurately capturing spatial risk in complex geometries. We illustrate the performance using crime and traffic incident data provided by the Indianapolis Metropolitan Police Deparment and IED attacks in Iraq.**

I. Introduction

Many types of events related to human activity cluster in space and time, forming event "hotspots." For example, in Figure 1 we plot Improvised Explosive Device (IED) attacks in Baghdad from 2004-2009. These events cluster along road networks and at major intersections within the spatial geography of the city. IED attacks also tend to cluster in time (1) due to self-excitation and exogenous effects. In the case of burglary, offenders are known to replicate success at nearby (or identical locations) to previous crimes (2). Hotspot policing is a strategy for deterring crime where police resources are directed to the highest volume crime areas of a city. Experimental studies indicate that elevated policing presence in hotspots comprising a relatively small area of the city can lead to statistically significant crime rate reductions (3). The standard approach for determining hotspots consists of dividing a city into geographic sub-regions, often grid cells, and scoring hotspots based upon historical crime counts over a specified time window (4). More recently, point processes have been introduced for ranking crime hotspots (5) and have been shown to lead to further crime rate reductions in field trials over traditional hotspot mapping (6). Other approaches for ranking crime hotspots include generalized linear models (7; 8), generalized additive models (8), and random forests have been applied to the problem of ranking offenders (9). Space-time models for event prediction also have been applied to conflict (10) and terrorism (11) datasets.

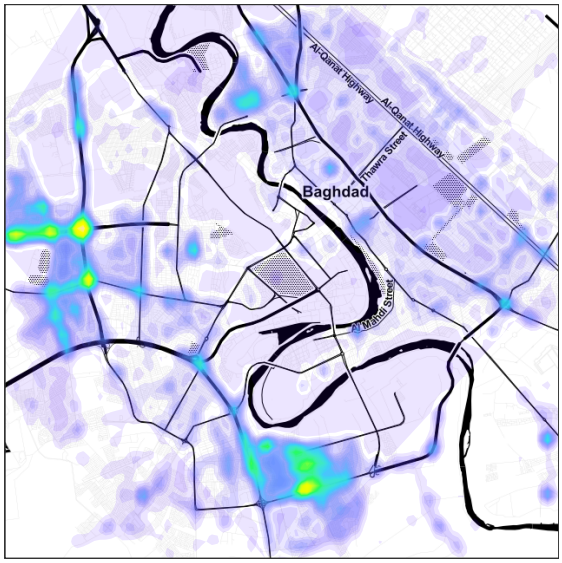

Fig. 1. IED attack hotspots in Baghdad from 2004 to 2009.

Since the goal of hotspot and predictive policing is crime rate reduction, the standard metric for assessing a given scoring procedure is simply the percent of crime captured inside the top ranked hotspots in the absence of police intervention. The predictive accuracy index (PAI) (4; 6; 12),

$$\text{PAI} = \frac{\text{crime in k hotspots}}{\text{total crime}} \cdot \frac{\text{total area}}{\text{area of k hotspots}}, \qquad (1)$$

measures the percent of crime predicted in the top k hotspots normalized so that spatially random predictions have a PAI value of 1. In practice, the value of k is chosen to correspond to policing resources and realistic values may correspond to an area on the order of 1% of a city (6).

Similar loss functions, such as NDCG@k, Prec@k and Recall@k, are used in information retrieval (13) to measure the effectiveness of scoring algorithms aimed at producing a high percentage of relevant documents in the top k documents returned from a query. The mathematical formulation of the two problems is similar, where the analog of a query is the time unit (window) for which crime hotspot predictions are made, the analog of a document is a single spatial unit (grid cell, neighborhood, block, street corner, etc.) in the city, and the analog of relevance is a binary or integer variable indicating whether or not a crime occurred inside the spatial unit and time window (or how many crimes occurred). We therefore

use the notation PAI@k to denote the PAI value when the top k hotspots are flagged for police intervention. Learning to rank algorithms attempt to directly optimize the loss function of interest and have been shown to out-perform regression and likelihood based algorithms that optimize a smooth surrogate loss function (13; 14).

In this paper we develop a learning to rank algorithm, CrimeRank, for space-time event hotspot ranking. A general overview of the algorithm is as follows. Features are defined for each hotspot in a city at a particular time unit and then those features are used to define a score that is used to rank hotspots according to risk over the next time unit (in the future). Similar to LambdaMart (14), we introduce a pseudo-derivative for PAI@k and then perform gradient ascent boosting to maximize PAI. At each iteration we use decision trees as the weak learner to model the derivative of PAI as a function of the features in each hotspot. At prediction time we compute a score over a collection of over-lapping spatial grids and then perform a greedy sort to select the top k hotspots so that they do not overlap. We apply the method to several space-time event data sets to illustrate the improvement in PAI of CrimeRank over existing methodologies. The outline of the paper is as follows: in Section 2 we provide details on the CrimeRank algorithm and in Section 3 we include results for the CrimeRank algorithm on several data sets including crime and traffic incidents in Indianapolis and IED attacks in Baghdad.

II. CrimeRank: An algorithm for ranking crime and other space-time event hotspots

A. Feature selection

Given a data set of space time event locations up to the present day, our goal is to flag a set of k spatial areas that have the highest risk for event occurrence in the near future, e.g. the next day, week, month, etc. In this paper we will consider a regular grid for dividing a city into sub-areas, though our methodology applies to more general polygons and other subdivisions.

In the case of crime, algorithms typically fall into one of two broad categories for ranking spatial areas, namely nonparametric methods utilizing only event data (kernel hotspot maps and point processes are common methods) or multivariate models that explicitly incorporate additional variables such as demographics (8), income levels (15), distance from crime attractors (8; 15; 7), and leading-indicator crimes (16; 17).

Because the focus of this paper is on the optimization method used to train a hotspot ranking model, rather than feature selection, we restrict our attention to univariate modeling where features are derived from the event data alone. Our methodology would easily extend to other types of features. For univariate feature extraction we compute a 52-week time series consisting of the event count in each grid cell for the 52 weeks leading up to the present. Our learning task is then to rank grid cells in the subsequent week in the future such that the highest k cells have the largest number of events in the next week.

A training data set is then created over a historical time period by computing the 52 dimensional feature set for each cell and each week in the historical period, where the label is the number of events in the next week. While each row in the data set is a grid cell-week pair, we note that rows are not independent in the ranking setting and all rows corresponding to the same week must be considered simultaneously to compute PAI. The analog of a week in the information retrieval setting is a query. However, regression based methods will treat all rows as independent during training.

B. Optimization of PAI@k

Next we describe our optimization method for maximizing PAI@k, the area normalized fraction of crime in the top k event hotspots. Let i index all grid cell-week pairs and let z_i denote the feature vector for cell-week i. We will denote the score for z_i as s_i and the label (number of events in next week) as y_i.

We first note that PAI is non-smooth as a function of s_i. In particular, consider fixing the scores except for two grid cells in the same week indexed by i and j and assume $y_i > y_j$. Then PAI will be piecewise constant as a function of $s_i - s_j$ and will have a jump discontinuity at $s_i = s_j$. Therefore PAI has no derivative for performing gradient ascent. However, we follow the approach of (14) and introduce a pseudo-derivative λ_i,

$$\lambda_i = \sum_{j:y_i>y_j} \frac{|\Delta_{PAI}|}{1+e^{s_i-s_j}} - \sum_{j:y_j>y_i} \frac{|\Delta_{PAI}|}{1+e^{s_j-s_i}}, \qquad (2)$$

that models the gradient of PAI at cell-week i. Here the term Δ_{PAI} denotes the change in PAI resulting when swapping the ranking of cell-weeks i and j (leaving all other rankings fixed). The first summation is over all hotspot pairs where i should be ranked higher than j and thus is positive in order to increase the score s_i and thus increase the PAI. The logistic term evaluated at $s_i - s_j$ is introduced to add regularization and in (14) the authors find that it has the effect of adding a margin. The second term is over hotspot pairs where i should be ranked lower than j and thus has the effect of lowering the score s_i (and therefore increasing PAI).

We note that the computational cost of λ_i over all i is quadratic, however in practice the performance is approximately linear. First, only grid cells in the same week need to be considered when computing λ. Second, within a given week for many event data sets a small percentage of cell labels y_i will be non-zero. When computing λ, the two sums in Equation 2 only need to be computed for $y_i \neq y_j$, which is $O(M_0 M_1)$ where M_1 is the number of non-zero labels for a given week and M_0 is the number of zero label cells.

Given the model for the derivative λ of PAI@k, we then use decision tree based gradient boosting to optimize the loss function. We call our method CrimeRank and provide pseudocode in Algorithm 1. Starting with an initial guess for scores s_i, we then perform boosting iterations where 1. the pseudo-derivative λ_i is computed using the current score guess, 2. a regression tree is fit to the derivative λ_i as a function of

Algorithm 1 CrimeRank

Input: features z_i, labels y_i for $i = 1, ..., N$. Number of trees M, learning rate η. Initial guess score $s_i = 0$.
for $l = 1, ..., M$ **do**
  for i=1,...,N **do**
    $\lambda_i = \sum_{j:y_i>y_j} \frac{|\Delta_{PAI}|}{1+e^{s_i-s_j}} - \sum_{j:y_j>y_i} \frac{|\Delta_{PAI}|}{1+e^{s_j-s_i}}$
  end for

  $\Gamma_l \leftarrow RegTree(\{z_i, \lambda_i\}_{i=1}^N)$

  for i=1,...,N **do**
    $s_i = s_i + \eta\ \Gamma_l(z_i)$
  end for
end for

the features z_i, and 3. the score s_i is updated by a gradient ascent step. In practice we find that using stochastic gradient ascent performs better where a random subset of λ_i are used to estimate the regression tree Γ at each iteration. In Figure 2 we plot an example of boosting iterations for robbery incidents in Indianapolis. Empirically we find that the pseudo-derivative is effective in maximizing the PAI (proportional to the fraction of crime predicted) on training data. We provide more results in Section 3.

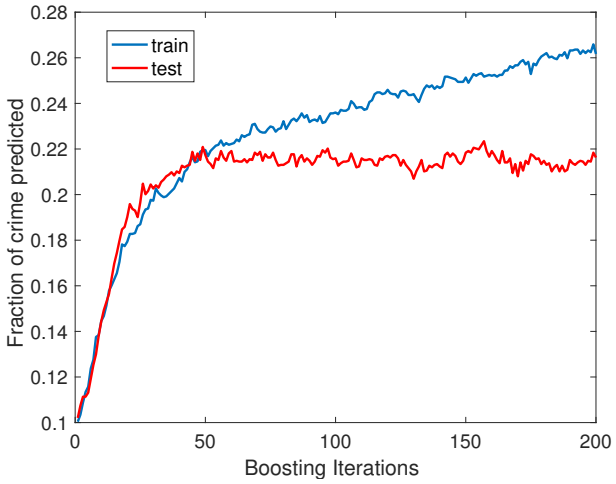

Fig. 2. CrimeRank boosting iterations using stochastic gradient ascent for robbery hotspot ranking in Indianapolis over 2013-2015 (split for training and testing).

C. Offgrid space-time ranking

The second component of CrimeRank is an "offgrid" approach that we introduce for dealing with complex geometries that are associated with event patterns along road networks and other urban structures. In Figure 3 we provide an illustration of the problem that arises with fixed grids used in spatial hotspot ranking. Here four events are plotted over a regular grid (thick black lines) and we let $k = 2$. Then four grid cells each have one event, the others have zero, so that the maximum

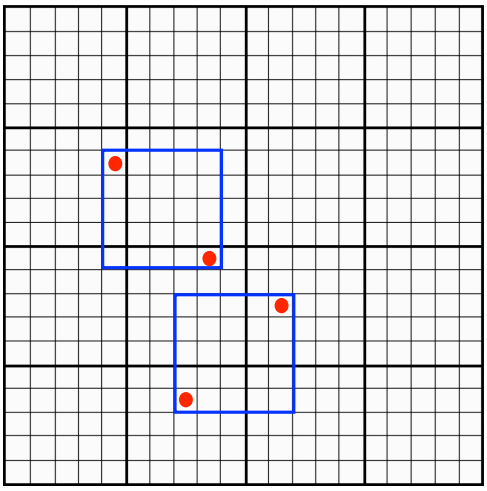

Fig. 3. 4x4 grid scenario. Maximum PAI@2 on the fixed grid is 4 (1/2 of crime captured divided by 2/16 of area flagged). Maximum PAI@2 on a floating grid is 8.

possible PAI@2 is four (two crimes out of four predicted area normalized by two cells out of sixteen). However, cells chosen without respect to a regular grid can achieve a PAI@2 of eight even with the same size and shape.

We introduce a simple heuristic for moving to an offgrid approach while taking advantage of the CrimeRank algorithm introduced in Section 2b. In particular, we train CrimeRank on a fixed regular grid and then introduce a greedy sort algorithm at prediction time applied to a collection of over-lapping grids. We use $g \times g$ over-lapping grids identical to the original fixed grid except that they are offset by a multiple of $\Delta x/g$ from the fixed grid where Δx is the length of the side of a grid cell. In Figure 3 we provide an illustration with $g = 5$.

We score every cell in each over-lapping grid using CrimeRank and at the time of prediction we then select k hotspots greedily (see Algorithm 2). First we select the cell with the highest score over all grids. Second we select the cell with the next highest score such that it does not overlap with the first cell. We continue on in this fashion, where the jth cell is selected with the highest score such that it does not overlap with cells $1, ..., j-1$. In practice we find that $g = 10$ works well in balancing accuracy and storage/computational costs.

Algorithm 2 Greedy sort for hotspot selection

Input: number of hotspots k, grid cell scores s_j, grid cell polygons U_j, hotspot cell set V=$\emptyset$
while Area$(V) < k$·Area(U_1)
  j $\leftarrow \arg\max\{s_j : U_j \cap V = \emptyset\}$
  V $\leftarrow V \cup U_j$
end while

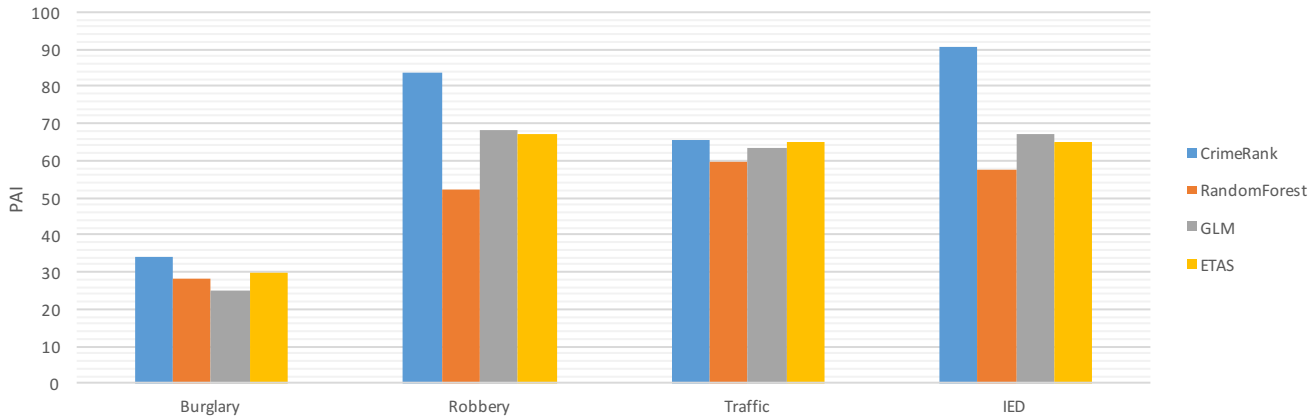

Fig. 4. PAI results of CrimeRank vs. three methods of comparison for ranking hotspots of burglary, robbery, traffic incidents, and IED attacks.

III. Results

A. Indianapolis crime hotspot ranking

In our first example we test the CrimeRank methodology using crime incident data provided by the Indianapolis Metropolitan Police Department from 2012-2015. In the data set there are 35,225 burglary incidents, 13,135 robbery incidents, and 40,066 traffic related incidents including investigations and arrests. We make weekly predictions and following (6) we use grid cells of size $150m \times 150m$. We compare CrimeRank to several existing methods including a random forest and generalized linear model (GLM) using identical features, along with an Epidemic Type Aftershock Sequence (ETAS) point process (6; 5). For the ETAS model, we follow (6) where the conditional intensity $f_i(t)$ of events in grid cell i at time t is determined by,

$$f_i(t) = \mu_i + \sum_{t_i^j < t} \theta\omega e^{-\omega(t-t_i^j)}, \tag{3}$$

where t_i^j are the times of events in cell i in the history of the process. The ETAS model has two components, one modeling place-based environmental conditions that are constant in time and the other modeling dynamic changes in risk. The parameters μ_i, θ and ω are estimated using an Expectation-Maximization algorithm (6).

We use the time period 1/1/2013 to 6/31/2014 for training and we evaluate the methods over the time period 7/1/2014 to 12/31/2015. For CrimeRank we use a max leaf size of 500 for the regression trees and subsample 1/4 of the training data when constructing each tree. We use $k = 200$ grid cells for evaluation, comprising approximate 0.4% of the city, on the same order of magnitude as realistic predictive policing deployments (6).

In Figure 4 we plot the PAI results for the four methods applied to crime incident data in Indianapolis. For all three crime incident types CrimeRank outperforms the other methodologies. In the case of burglary, CrimeRank captures 21%, 36% and 14% more crime in the top 200 hotspots compared to the random forest, GLM, and ETAS models respectively. The largest improvement is in the case of robbery, where CrimeRank improves by over 20% crime captured compared to the next best method. For traffic incidents, CrimeRank is marginally better than ETAS (not statistically significant). An explanation for these results is that in the case of robbery, crime is highly clustered on street networks and CrimeRank is able to adapt to the geometry of the network (see Figure 5). Burglary is more spatially disaggregated and thus there is less room for improvement, whereas traffic investigations and arrests may be more random than robbery and burglary.

B. Improvised Explosive Device (IED) attacks in Baghdad, Iraq

In our second example we test the CrimeRank methodology using IED incident data from central Baghdad, including date, latitude and longitude of attacks, during the Iraq War from 2004 to 2009. In the data set there are 16,495 IED attacks. We again make weekly predictions and use grid cells of size $150m \times 150m$. We compare CrimeRank to the same three methods as in Section 3a including a random forest and generalized linear model (GLM) using identical 52 week time series features, along with the Epidemic Type Aftershock Sequence (ETAS) point process. We use the time period 1/1/2006 to 6/31/2007 for training and we evaluate the methods over the time period 7/1/2007 to 12/31/2008. We again use $k = 200$ grid cells for evaluation, comprising approximate 0.4% of the central area of Baghdad.

In Figure 4 we plot the PAI results for the four methods applied to IED incident data in central Baghdad. Similar to robbery, CrimeRank outperforms the other methodologies by 57%, 35% and 39% compared to the random forest, GLM, and ETAS model respectively. In Figure 5 we provide an example of the CrimeRank hotspot distribution on a given week in the testing period for a section of central Baghdad. We note that grid cells are able to align to intersections and diagonal roads in a manner such that the corners of the grid cell are aligned

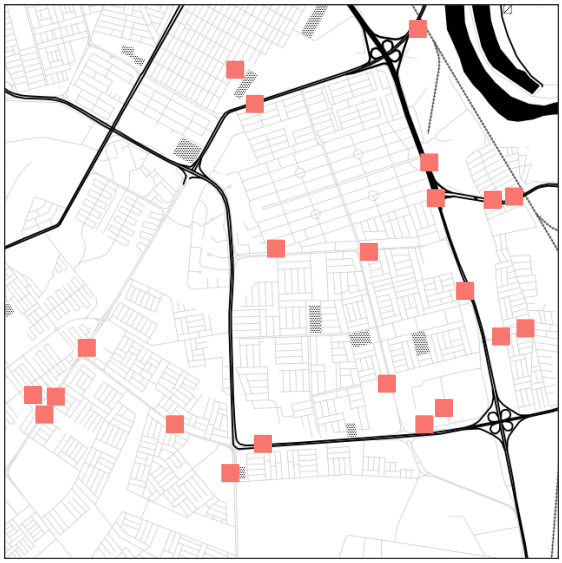

Fig. 5. CrimeRank determined IED hotspots for a week in 2008 in an area of central Baghdad. Hotspots align to road network and certain intersections to maximize PAI.

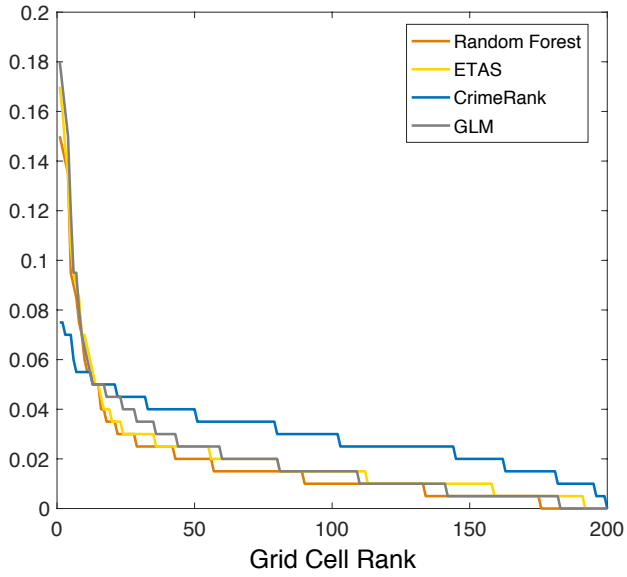

Fig. 6. Average number of incidents captured in the top k grid cells.

with the street, thus maximizing PAI (for example the left most cluster of four cells illustrate this effect).

In Figure 6 we plot the average number of IED incidents captured in the top k grid cells (as a function of k). One interesting effect to note is that the highest grid cells of CrimeRank contain less incidents compared to the other three methods. This is likely due to the fact that PAI is not changed by a re-ordering of the top grid cells ranking, but instead is sensitive to cells either being inside or outside of the top k. After the top 10 cells, CrimeRank cells contain significantly more incidents than the other three methods, explaining the overall improvement in PAI.

IV. Conclusion

We developed a spatial-temporal learning to rank algorithm, CrimeRank, for identifying high risk "hotspots" in human activity data. The method directly optimizes the PAI@k loss function from criminology using gradient boosting. Although the loss function is non-smooth, a pseudo derivative is used in the boosting algorithm that empirically maximizes PAI. CrimeRank also deals with the geometry of hotspots in urban environments using a novel greedy sorting algorithm at the time predictions are made. We show that CrimeRank improves the % of events captured in hotspots by up to 35% compared to existing methods for crime and IED event data.

V. Acknowledgements

This work was supported in part by NSF grants SES-1343123 and CCF-1659488. The author is a co-founder of PredPol, a company offering predictive policing services to law enforcement agencies and serves on the board of directors.

References

[1] E. Lewis and G. Mohler, "A nonparametric em algorithm for multiscale hawkes processes," *preprint*, 2011.

[2] M. Short, M. D'Orsogna, P. Brantingham, and G. Tita, "Measuring and modeling repeat and near-repeat burglary effects," *Journal of Quantitative Criminology*, vol. 25, no. 3, pp. 325–339, 2009.

[3] A. Braga, "The effects of hot spots policing on crime," *The ANNALS of the American Academy of Political and Social Science*, vol. 578, no. 1, pp. 104–125, 2001.

[4] S. Chainey, L. Tompson, and S. Uhlig, "The utility of hotspot mapping for predicting spatial patterns of crime," *Security Journal*, vol. 21, no. 1, pp. 4–28, 2008.

[5] G. Mohler, M. Short, P. Brantingham, F. Schoenberg, and G. Tita, "Self-exciting point process modeling of crime," *Journal of the American Statistical Association*, vol. 106, no. 493, pp. 100–108, 2011.

[6] 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, no. 512, pp. 1399–1411, 2015.

[7] L. Kennedy, J. 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, no. 3, pp. 339–362, 2011.

[8] X. Wang and D. Brown, "The spatio-temporal modeling for criminal incidents," *Security Informatics*, vol. 1, no. 1, pp. 1–17, 2012.

[9] R. Berk, L. Sherman, G. Barnes, E. Kurtz, and L. Ahlman, "Forecasting murder within a population of probationers and parolees: a high stakes application

of statistical learning," *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, vol. 172, no. 1, pp. 191–211, 2009.

[10] A. Zammit-Mangion, M. Dewar, V. Kadirkamanathan, and G. Sanguinetti, "Point process modelling of the afghan war diary," *Proceedings of the National Academy of Sciences*, vol. 109, no. 31, pp. 12 414–12 419, 2012.

[11] P. Gao, D. Guo, K. Liao, J. J. Webb, and S. L. Cutter, "Early detection of terrorism outbreaks using prospective space–time scan statistics," *The Professional Geographer*, vol. 65, no. 4, pp. 676–691, 2013.

[12] "https://nij.gov/funding/pages/fy16-crime-forecasting-challenge.aspx," 2017.

[13] T.-Y. Liu *et al.*, "Learning to rank for information retrieval," *Foundations and Trends® in Information Retrieval*, vol. 3, no. 3, pp. 225–331, 2009.

[14] C. J. Burges, "From ranknet to lambdarank to lambdamart: An overview," *Learning*, vol. 11, no. 23-581, p. 81, 2010.

[15] H. Liu and D. E. Brown, "Criminal incident prediction using a point-pattern-based density model," *International journal of forecasting*, vol. 19, no. 4, pp. 603–622, 2003.

[16] 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, no. 1, pp. 105–127, 2007.

[17] W. L. Gorr, "Forecast accuracy measures for exception reporting using receiver operating characteristic curves," *International Journal of Forecasting*, vol. 25, no. 1, pp. 48–61, 2009.