RESEARCH   **Open Access**

Learning to rank spatio-temporal event hotspots

George Mohler[1*], Michael Porter[2], Jeremy Carter[1] and Gary LaFree[3]

Abstract

Background: Crime, traffic accidents, terrorist attacks, and other space-time random events are unevenly distributed in space and time. In the case of crime, hotspot and other proactive policing programs aim to focus limited resources at the highest risk crime and social harm 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.

Results: 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 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 Department, IED attacks in Iraq, and data from the 2017 NIJ Real-time crime forecasting challenge.

Conclusion: Our learning to rank strategy was the top performing solution (PAI metric) in the 2017 challenge. We show that CrimeRank achieves even greater gains when the competition rules are relaxed by removing the constraint that grid cells be a regular tessellation.

Keywords: Learning to rank, Crime hotspot, Gradient boosting, Crime forecast

Introduction

Related work

Real-time spatiotemporal crime forecasting has become a focal point of public and private sector development, with a desired end-state of crime reduction coupled with police efficiency (Perry 2013). Two large bodies of scholarly inquiry have served as the catalyst for this interest in improved crime forecasting. First, large proportions of crime events are concentrated within small proportions of micro-places in urban environments (Weisburd 2015). Many types of events related to human activity cluster in space and time, forming event "hotspots." Burglary offenders are known to replicate success at nearby, or identical, locations to previous crimes (Short et al. 2009) and space-time clusters are observed in patterns of shootings (Ratcliffe and Rengert 2008) due to retaliation and escalation. Event hotspots also occur in more extreme security settings, for example Improvised Explosive Device (IED) attacks tend to cluster in time (Lewis and Mohler 2011) due to self-excitation and exogenous effects. In Fig. 1, we plot IED attacks in Baghdad from 2004 to 2009. These events cluster along road networks and at major intersections within the spatial geography of the city.

Second, experimental studies indicate that elevated policing in a small set of high-risk crime locations, known

*Correspondence: gmohler@iupui.edu
[1] Indiana University Purdue University Indianapolis, Indianapolis, USA
Full list of author information is available at the end of the article

Mohler *et al. Crime Sci* (2020) 9:3
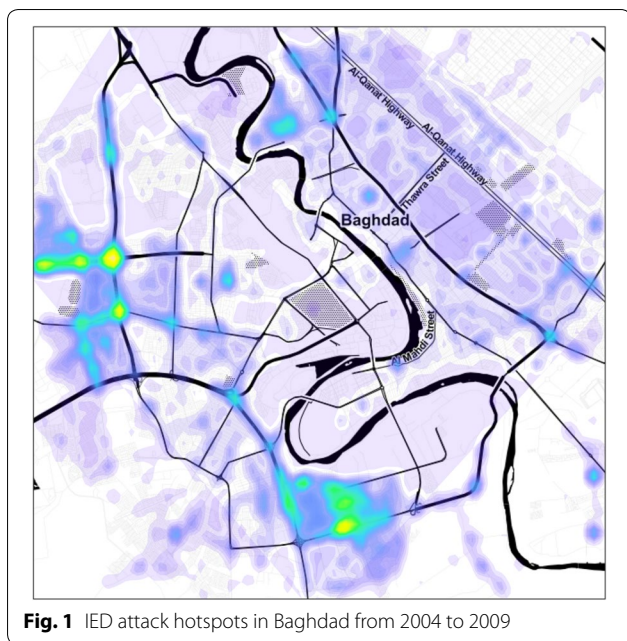
Page 2 of 12

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

as hotspots policing, can lead to statistically significant crime rate reductions (Braga et al. 2019). 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 (Chainey et al. 2008).

Despite these two empirical facts, there is much less consensus regarding the most appropriate, and most efficient, methods to estimate crime concentration and evaluate crime prediction methods. This is especially true when considering the array of event types for which police have responsibility and the variability that exists across event frequency and geographic units of analysis (Mohler et al. 2019). The discussion below of related works summarizes common approaches for crime prediction. While all existing metrics of geospatial crime concentration suffer drawbacks related to their stability over different space-time units, populations, or crime rates (Curiel 2019), forecast evaluation using concentration metrics is still a valid approach to assess the potential impact police interventions can have. Crime forecasting methods to date have taken several forms. Most common in the criminological literature are theory-driven models that account for the causes and correlates of crime, such as risk-terrain modeling (Caplan et al. 2011; Kennedy et al. 2011). These techniques rely upon environmental and structural theories of crime causation to quantify spatiotemporal crime risk. More data-driven approaches to crime prediction are prevalent across the computer science and statistics literatures. Smoothing techniques, most commonly kernel density estimation (KDE) (Gorr

and Lee 2015; Porter and Reich 2012), use historical events, rather than spatial covariates, to estimate risk. Related to KDE are log-Gaussian Cox Processes (LGCP) that model the space-time process generating crime and allow for seasonal and exogenous trends in the data. LGCPs can also detect the spatial diffusion of events, such as crime (Flaxman et al. 2018; Shirota and Gelfand 2017), violent crime (Taddy 2010), or the spread of infectious disease (Diggle et al. 2013). Self-exciting point processes are also used for ranking crime hotspots (Mohler et al. 2011) and have been shown to lead to crime rate reductions in field trials over traditional hotspot mapping (Mohler et al. 2015). Self-exciting point processes model repeat and near-repeat occurrences across space and time (Johnson et al. 2007; Piza and Carter 2018) and hotspot policing based on these models attempts to prevent this near-repeat aspect of offending. In more extreme security settings space-time point process models for event prediction have been applied to conflict (Zammit-Mangion et al. 2012) and terrorism (Gao et al. 2013) datasets and LGCPs have been combined with self-exciting point processes to predict crime and terrorism (Mohler 2013). Other approaches for ranking crime hotspots include generalized linear models (Kennedy et al. 2011; Wang and Brown 2012; Wang et al. 2016), generalized additive models (Wang and Brown 2012), and random forests have been applied to the problem of ranking offenders (Berk et al. 2009). In the past several years deep learning based approaches have also shown promise for space-time prediction of crime (Stec and Klabjan 2018; Wang et al. 2017).

Learning to rank for spatio-temporal event data

Since the goal of hotspot policing is crime rate reduction, the standard metric for assessing a given scoring procedure is the percent of crime captured inside the top ranked hotspots in the absence of proactive police intervention. The predictive accuracy index (PAI) (Chainey et al. 2008; Mohler et al. 2015); National Insititue of Justice 2017)

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

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 (Mohler et al. 2015).

Similar loss functions, such as NDCG@k, Prec@k and Recall@k, are used in information retrieval (Liu 2009) to measure the effectiveness of scoring algorithms aimed at

Mohler *et al. Crime Sci*   (2020) 9:3

Page 3 of 12

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 (Liu 2009; Burges 2010). We note that there has been some work on spatial learning to rank in the context of inferring a users location from noisy GPS data (Shaw et al. 2013), however to our knowledge no work to date has focused on the learning to rank problem in the context of crime event prediction.

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 potential hotspot in a city at a particular time unit and then used to calculate a risk score that ranks hotspots over the next (future) time unit. Similar to LambdaMart (Burges 2010), 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 the score for a collection of potentially over-lapping hotspots and then perform a greedy sort to select the top k non-overlapping hotspots. Stochastic gradient boosting has many of the advantages of random forests; the use of decision trees allows the model to capture nonlinear interactions and bootstrapping of the training data provides variance reduction. Boosting, however, has the added benefit that the loss function of interest is directly optimized.

Outline
We apply the CrimeRank method to several space-time event data sets to illustrate the improvement in PAI over existing methodologies. The outline of the paper is as follows: in "Methods" section we provide details on the CrimeRank algorithm and in "Results and discussion" section we include results for the CrimeRank algorithm on several data sets including crime and traffic incidents in Indianapolis, IED attacks in Baghdad, and data from Portland, Oregon used in the 2017 NIJ Real-time crime forecasting challenge. Our learning to rank strategy

under the team name PASDA was the top performing solution (PAI metric) in the 2017 challenge. We show that CrimeRank achieves even greater gains when the competition rules are relaxed and spatial discretizations are not required to be a regular tessellation. We discuss future directions for research in this area in "Conclusion" section.

Methods
In this section we provide the details of our algorithm. In "Feature selection" section we discuss feature selection within hotspots. In "Optimization of PAI@k" section we introduce our spatial learning to rank algorithm that models a pseudo-derivative of PAI and then performs stochastic gradient boosting. In "Offgrid space-time ranking" section we provide details on our off-grid approach to selecting event hotspot polygons.

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 rectangular grid cells for dividing a city into sub-areas, though our methodology applies to more general polygons and other sub-divisions.

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 (Wang and Brown 2012), income levels (Liu and Brown 2003), distance from crime attractors (Wang and Brown 2012; Liu and Brown 2003; Kennedy et al. 2011), leading-indicator crimes (Cohen et al. 2007; Gorr 2009), and auxiliary social sensing data (Twitter, mobile phone locations, Google street view, etc.) (Wang et al. 2012, 2016; Bogomolov et al. 2014; Khosla et al. 2014).

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 contextual features including stationary features such as census data (Kennedy et al. 2011) or more real-time data such as population density from mobile phones (Bogomolov et al. 2014). The latter is typically not available in most U.S. cities, therefore the majority of crime models use publicly available spatial covariates or are based solely on the events (e.g. univariate models). Because stationary covariates are primarily used for variance reduction in space-time crime models, they are less important in

Mohler *et al. Crime Sci*   (2020) 9:3

Page 4 of 12

learning to rank the top crime hotspots that are characterized by high volumes of events (hence variance is low). For this reason the top performing solutions in a recent NIJ forecasting competition were based on univariate modeling (Flaxman et al. 2018; Mohler and Porter 2017).

As an example using a weekly forecast window, a 52-week time series consisting of the event counts in each grid cell for the 52 weeks leading up to the present could be used as the features. Thus, the training data set would be created over a historical time period by computing the 52 dimensional feature set for each cell and each week, where the label is the number of events in the following week. The learning task is then to rank the grid cells such that the top k cells will have the largest number of events in the subsequent week. Each row in the training data is a grid cell-week pair. Because the PAI is based on all the grid cell rankings for a given time period, all rows corresponding to the same week must be considered simultaneously to compute the PAI for that week. The analog of a week in the information retrieval setting is a query. Note that regression based methods will treat all rows as independent during training.

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 $i \in \{1, 2, \ldots, N\}$ index the N grid cells and $t \in \{1, 2, \ldots, T\}$ index the T time periods in which predictions are being made. Let z_{it} denote the feature vector, s_{it} the score, and y_{it} the label for cell i at time t. Note that y_{it} is the number of events in the future time period $t + 1$. This gives a total of $N \times T$ observations.

The set of scores induce a ranking on the grid cells for each time period. Let r_{it} be the rank of score s_{it}, with a rank of one being assigned the cell with the *largest* score at time t. Then the top k cells, at time t, are $V_{kt} = \{i : r_{it} \leq k\}$. The resulting PAI is calculated separately for each time period.

We first note that PAI is non-smooth as a function of s_{it}. In particular, consider fixing the scores except for two grid cells in the same week t indexed by i and j and assume $y_{it} > y_{jt}$. Then PAI will be piecewise constant as a function of $s_{it} - s_{jt}$ and will have a jump discontinuity at $s_{it} = s_{jt}$. Therefore PAI has no derivative for performing gradient ascent. However, we follow the approach of Burges (2010) and introduce a pseudo-derivative λ_{it},

$$\lambda_{it} = \sum_{j:y_{it}>y_{jt}} \frac{|\Delta_{kt}(i,j)|}{1 + e^{s_{it}-s_{jt}}} - \sum_{j:y_{jt}>y_{it}} \frac{|\Delta_{kt}(i,j)|}{1 + e^{s_{jt}-s_{it}}}, \quad (2)$$

that models the gradient of PAI at cell-week i-t. Here the term $\Delta_{kt}(i,j)$ denotes the change in PAI if the ranking of cells i and j are swapped at time t (leaving all other rankings fixed) and can be written,

$$\Delta_{kt}(i,j) = \begin{cases} c\,(y_{it} - y_{jt})/N_t & r_i \leq k, r_j > k \\ 0 & r_i \leq k, r_j \leq k;\, r_i > k, r_j > k \\ c\,(y_{jt} - y_{it})/N_t & r_i > k, r_j \leq k \end{cases} \quad (3)$$

where $c = (\text{total area})/(\text{area of } k \text{ grid cells})$ is the PAI normalizing constant.

The first summation in (2) is over all pairs where grid cell i should be ranked higher than grid cell j and thus is positive in order to increase the score s_{it} and thus increase the PAI. The logistic term evaluated at $s_{it} - s_{jt}$ is introduced to add regularization and in Burges (2010) the authors find that it has the effect of adding a margin. The second term is over pairs where i should be ranked lower than j and thus has the effect of lowering the score s_{it} (and therefore increasing PAI).

We note that the computational cost of λ_{it} over all i is quadratic, however in practice the performance is approximately linear. First, only grid cells in the same time period need to be considered when computing $\{\lambda_{it}\}_{i=1}^{N}$. Second, for many event data sets and reasonably small grid cells only a small percentage of cells will contain non-zero counts. Because (2) only involves pairs in which $y_i \neq y_j$ the cost is $O(M_0 M_1)$ where M_1 is the number of non-zero labels for a given t and M_0 is the number of zero label cells.

Algorithm 1 CrimeRank

> **Input:** features z_{it}, labels y_{it} for $i = 1, \ldots, N$ and $t = 1, \ldots, T$. Number of trees M, learning rate η. Initial guess score $s_{it} = 0$.
> **for** $l = 1, \ldots, M$ **do**
>   **for** $t = 1, \ldots, T$ **do**
>   **for** $i = 1, \ldots, N$ **do**
>
> $$\lambda_{it} = \sum_{j:y_{it}>y_{jt}} \frac{|\Delta_{kt}(i,j)|}{1 + e^{s_{it}-s_{jt}}} - \sum_{j:y_{jt}>y_{it}} \frac{|\Delta_{kt}(i,j)|}{1 + e^{s_{jt}-s_{it}}}$$
>
>   **end for**
>   **end for**
>
> $\Gamma_l \leftarrow \text{RegTree}(\lambda, z)$
> where $\lambda = [\lambda_{11}, \ldots, \lambda_{N1}, \lambda_{12}, \ldots, \lambda_{NT}]$,
> $z = [z_{11}, \ldots, z_{N1}, z_{12}, \ldots, z_{NT}]$
>
>   **for** $i = 1, \ldots, N$ and $t = 1, \ldots, T$ **do**
> $s_{it} = s_{it} + \eta\, \Gamma_l(z_{it})$
>   **end for**
> **end for**

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 pseudo-code in Algorithm 1. Starting with an initial guess for scores s_{it}, we then perform boosting iterations where (i) the pseudo-derivative λ_{it} is computed using the current score guess, (ii) a regression tree is fit to the derivative λ_{it} as a function of the features z_{it}, and (iii) the score s_{it} is updated by a gradient ascent step. In practice we find that using *stochastic* gradient ascent (Friedman 2002) performs better where a random subset of λ_i

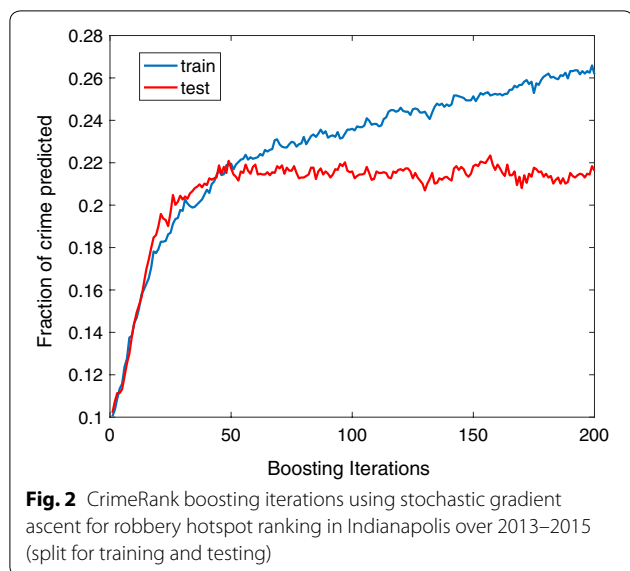Mohler *et al. Crime Sci* (2020) 9:3

Page 5 of 12

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

are used to estimate the regression tree Γ at each iteration. In Fig. 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 "Results and discussion" section.

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 Fig. 3 we provide an illustration of the problem that arises with fixed grids used in spatial hotspot ranking. Here four
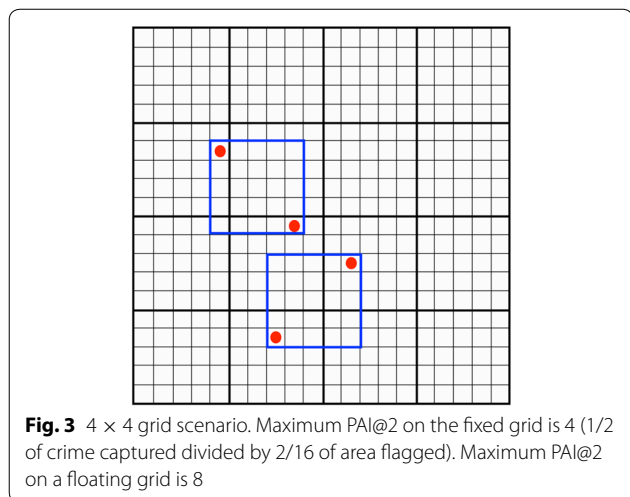

Fig. 3 4×4 grid scenario. Maximum PAI@2 on the fixed grid is 4 (1/2 of crime captured divided by 2/16 of area flagged). Maximum PAI@2 on a floating grid is 8

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 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 "Optimization of PAI@k" section. In particular, we train CrimeRank on a fixed regular grid obtaining the fitted CrimeRank model (i.e., the collection of regression trees). The CrimeRank model is then used to estimate the risk score, during the evaluation period, for a larger collection of grid cells and a greedy sort algorithm is used to find the set of k non-overlapping cells with the largest scores.

The CrimeRank model is fit one time, on a given grid from the training data, and then used to estimate the score, for all times in the evaluation period, at additional grid cells. The additional collection of grid cells can be generated, e.g., by translating and rotating the original grid used for model fitting. Because the model features must be calculated for the new grid cells, it is important to use the same size cells. In "Indianapolis crime hotspot ranking and Improvised Explosive Device (IED) attacks in Baghdad, Iraq" sections 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. Figure 3 illustrates the setting of $g = 5$; the thick lines shows the original 16 grid used for training the model and the collection of 200 additional grid cells are the square regions obtained by centering on each small square. In practice we find that $g = 10$ works well in balancing accuracy and storage/computational costs. In "2017 NIJ Crime Forecasting challenge" section, we also incorporated rotated grid cells to expand the number of potential hotspots.

Once all of the grid cells are scored, we utilize a greedy sort algorithm (Algorithm 2) to identify the top k non-overlapping hotspots. 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 $1, \ldots, j - 1$.

We note that there is a connection between the offgrid methodology we have proposed here and spatial scan statistics used to detect anomalies (for example disease outbreaks) in spatial-temporal event data (Kulldorff 2001; Assunção and Correa 2009; Neill 2009). The goal of the scan statistic approaches is to detect emerging

spatio-temporal clusters that have anomalous event rates by *scanning* over many possible spatial regions and time periods. For example, in Kulldorff (2001) circles Z of varying radius and center location are defined and then a likelihood ratio test using the statistic $L(Z)/L_0$ (where L is a Poisson likelihood) is used to flag clusters. Our goal is different, namely identifying the regions with the largest expected event rate in the future rather than identifying the regions that have the most unusual event rates in the recent past. For this purpose we are using features within each region to predict future risk and then directly optimizing a ranking loss function. We note that the scan statistic methods developed to search for irregularly shaped clusters (Duczmal et al. 2008, 2006; Speakman et al. 2016; Neill 2012; Tango and Takahashi 2005) could be used to generalize the rectangular regions we considered here and speed the search process. We will return to this idea in the discussion in "Conclusion" section.

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 $\text{Area}(V) < k \cdot \text{Area}(U_1)$
      $j \leftarrow \arg\max\{s_j : U_j \cap V = \emptyset\}$
      $V \leftarrow V \cup U_j$
    end while

Results and discussion
Baseline models
We compare CrimeRank to several existing methods including random forest (Mohler and Porter 2017; Alves et al. 2018), generalized linear model (GLM) (Kennedy et al. 2011; Wang et al. 2012), gradient boosting machine (GBM) applied to the ranking metric NDCG (Ridgeway 2007), a Hawkes point process (Mohler et al. 2015, 2011), kernel density estimation (Chainey et al. 2008), and a CNN-LSTM (Stec and Klabjan 2018; Groß et al. 2017). CrimeRank, random forest, GLM, and GBM use the same features (weekly event counts in the grid over the last 52 weeks). The self-exciting Hawkes model and kernel density estimation use the raw events as input. For the CNN-LSTM we use a 52 week time series of event counts in the 5×5 grid cell patch surrounding and including the target cell as input. We use 2 convolution layers with 3×3 filters followed by a LSTM and dense layer.

Indianapolis crime hotspot ranking
In our first example we test the CrimeRank methodology using crime and vehicle crash incident data from the city of Indianapolis, Indiana. Crime incidents for years 2012–2015, specifically robbery and residential burglary, were

provided electronically by the Indianapolis Metropolitan Police Department (IMPD). Vehicle crash data for years 2012–2013 were provided electronically from the Indiana State Police using the Automated Reporting Information Exchange System (ARIES). One of two characteristics must occur for collisions to be included in ARIES; if the incident resulted in personal injury or death, or property damage to an apparent extent greater than one thousand dollars. Both crime and crash data included date and time stamp as well as state-plane coordinates from a composite address locator that were converted to WGS84 coordinates. Robbery (Haberman and Ratcliffe 2012; Youstin et al. 2011; Ratcliffe and Rengert 2008), residential burglary (Nobles et al. 2016; Piza and Jeremy 2017; Bernasco 2008), and vehicle crashes (Carter and Piza 2018; Drawve et al. 2017; Kuo et al. 2013) have demonstrated spatiotemporal patterns in criminological research that are likely to inform strategic police operations to mitigate risk and deter offending. Thus, these three incident types are the focus of the present demonstration.

In the data set there are 35,225 burglary incidents, 13,135 robbery incidents, and 42,328 traffic accidents and we model and evaluate each event type separately. We consider weekly time periods and, following (Mohler et al. 2015), use grid cells of size $150\,\text{m} \times 150\,\text{m}$. We use the time period 1/1/2013 to 6/31/2014 for training and evaluate the methods on each week during the time period 7/1/2014 to 12/31/2015 (for traffic accidents we use 1/1/2013 to 6/31/2013 for training and 7/1/2013 to 12/31/2013 for testing). 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 hotspot policing deployments (Mohler et al. 2015).

In Table 1 we list the PAI results for CrimeRank and the baseline methods applied to crime and traffic crash incident data in Indianapolis. For all three incident types CrimeRank outperforms the other methodologies. CrimeRank captures 36% more events for burglary and 28% more events for robbery than the next best method. The improvement for traffic crashes is lower, but CrimeRank still has a PAI of over 60 compared to the other methods with a maximum PAI of 55. 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 Fig. 4). Traffic crashes are clustered at intersections and burglary is more spatially disaggregated and thus the PAI values are lower compared to those for robbery.

Mohler *et al. Crime Sci*   (2020) 9:3

Page 7 of 12

Table 1 PAI results of CrimeRank vs. baseline models for ranking hotspots of burglary, robbery, traffic crashes, and IED attacks
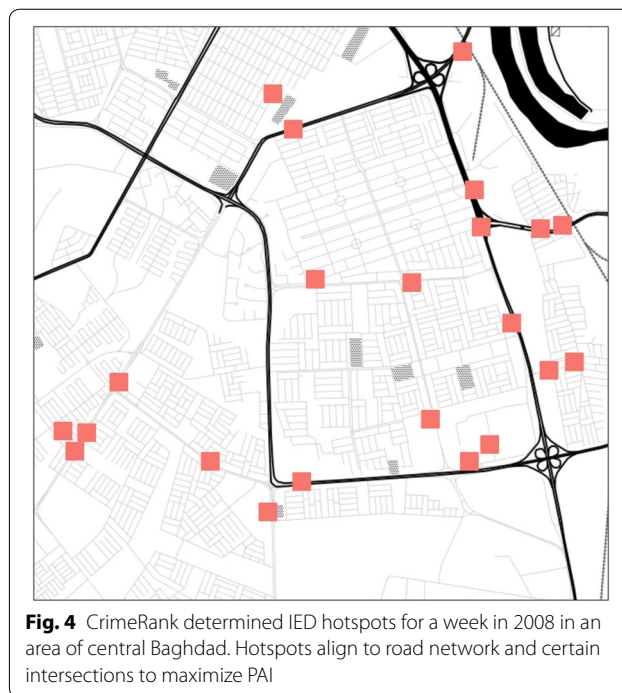
	CrimeRank	Rand. forest	GLM	Hawkes	CNN-LSTM	KDE	GBM-NDCG
Indpls. burglary	41.56	28.61	25.52	30.41	29.14	3.47	16.31
Indpls. robbery	88.71	53.61	69.68	68.32	32.68	10.46	46.48
Traffic crash	60.56	54.37	54.63	55.13	54.84	43.10	40.38
Baghdad IED	95.13	57.11	66.77	65.05	64.74	62.56	47.84

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. The attack data are based on Significant Activity (SIGACT) reports by Coalition forces in Iraq. Unclassified data from the MNU-I SIGACTS III database were provided to the Empirical Studies of Conflict (ESOC) project (Berman et al. 2011). The data set includes a wide range of activity but our analysis here is limited to IEDs. The SIGACT data have two weaknesses that are relevant here. First, they capture violence against civilians and between non-state actors only when U.S. forces are present and so likely undercount sectarian violence (Leonard 2009; Fischer 2008). Given that our emphasis is on IEDs, missing sectarian violence should not bias our results. Second, these data almost certainly suffer from measurement error in that units vary in their thresholds for reporting specific events as significant activity. Fortunately, there is no evidence that such error is nonrandom with regard to the IED locations. Missing data is inherent in all of the applications we consider in this paper; crimes and traffic crashes also may go unreported and adjusting forecasting models to compensate is beyond the scope of the paper.
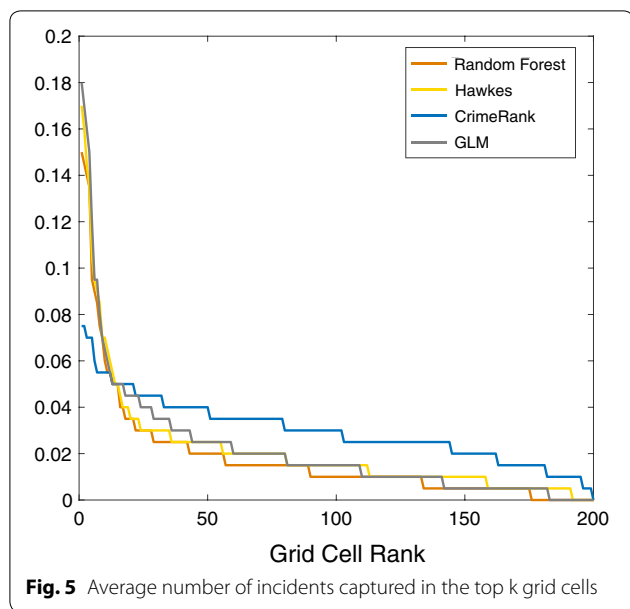
We again make weekly predictions and use grid cells of size $150\,\text{m} \times 150\,\text{m}$. For CrimeRank we use a max leaf size of 500 for the regression trees and subsample 1/4 of the training data when constructing each tree. We compare CrimeRank to the same baseline methods as in "Indianapolis crime hotspot ranking" section using identical 52 week time series features. 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 approximately 0.4% of the central area of Baghdad (chosen for the study to be a similar size to Indianapolis).

In Table 1 we list the PAI results for CrimeRank and the baseline methods applied to the IED incident data. Similar to robbery, CrimeRank outperforms the other methodologies by over 42%. In Fig. 4 we provide an example

Fig. 4 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

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

In Fig. 5 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 methods that use maximum likelihood estimation. 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 methods, explaining the overall improvement in PAI.

Mohler *et al. Crime Sci* (2020) 9:3

Page 8 of 12

Fig. 5 Average number of incidents captured in the top k grid cells

Table 2 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

2017 NIJ Crime Forecasting challenge

The 2017 NIJ Crime Forecasting challenge tasked participants with forecasting the spatial locations containing the highest volume of crime-related calls 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. 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 $62,500\,\text{ft}^2$ and $360,000\,\text{ft}^2$ with the smallest dimension being at least 125 ft. The cells flagged as hotspots were required to have aggregate area between $0.25\,\text{mi}^2$ and $0.75\,\text{mi}^2$, but there was no requirement that the hotspot cells be connected.

For the competition, we developed a Rotational Grid PAI maximization strategy (RGPM) (Mohler and Porter 2017) under the team name PASDA that was designed for jointly learning an optimal grid and scoring function for the purpose of maximizing PAI in crime forecasts under the rules of the NIJ competition. We used a regular grid of equally sized rectangles with the minimum allowable area $(62,500\,\text{ft}^2)$. The grid was parametrized with three parameters: cell height h, a grid translation parameter γ and a rotation angle θ. The overall procedure is captured in Algorithm 3, where the model $\mathcal{M}$ mapping features to the target variable was either a point process based GLM or a random forest (depending on crime category). A simplex method was used to maximize PAI with respect to the rotational grid parameters.

In Table 2 we include overall competition results illustrating the accuracy of our 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 2 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.

Next we compare CrimeRank and the baseline models from the previous section to the top performing methods of the NIJ competition. The methods again use 52 week count features (or the raw events for the Hawkes process and KDE). For training we use the time period 3/1/2013 to 5/31/2016 and then we evaluate the CrimeRank method using the competition validation data set.

Mohler *et al. Crime Sci* (2020) 9:3

Page 9 of 12

Algorithm 3 Rotational grid PAI maximization

1: **Function PAI**$(h,\theta,\gamma,\vec{x}_i,t_i,\omega,A_{min})$
    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: **Function OptimizeGrid**$(\vec{x}_i,t_i,\omega,A_{min})$
    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.

For comparison we also add a rotational version of CrimeRank. We consider (250 ft $\times$ 250 ft) squares as well as (125 ft $\times$ 500 ft) rectangles with four orientations (0, $\pi/4$, $\pi/2$ and $3\pi/4$). To reduce the memory requirements of using the offgrid search, we generate the additional grid cells by creating rectangles centered at a sub-sample of the event locations in the training period (10000 events).

We use a max leaf size of 100 for street crime and 50 for all calls for service for the regression trees and sub-sample 1/4 of the training data when constructing each tree. Examples of the Rotational CrimeRank hotspot cells are shown in Fig. 6. The code to reproduce our CrimeRank results is available at Github (Crimerank 2018).

We restrict our attention to the categories street crime and all calls for service over the 3 month forecasting window. We use the 3 months forecasting window so that variance does not play a large role in method ranking (in the NIJ competition short-term windows such as 1 week had very few events). In Table 3 we list CrimeRank PAI values (NIJ validation data set) compared to the baseline models. In the case of street crime, CrimeRank and its

rotational version achieve a PAI of 91 and 100 respectively compared to the 1st place solution PASDA (PAI 87) and the 2nd place solution TAMERZONE (PAI 84). For all calls for service, CrimeRank achieves a PAI of 64 compared to the 1st place solution CODILIME (PAI 60.5). We note in Fig. 6, where examples of Rotational CrimeRank hotspots are shown, that rectangles at diagonal angles are heavily favored in certain areas of Portland where major streets run diagonally. This effect was not possible within the rules of the NIJ competition, but meets the spirit of the rules in terms of cell shape, size, and non-overlapping requirements. Given the high societal cost of crime (McCollister et al. 2010), we believe a PAI improvement of 4 to 13 (over competition winning methods) is a significant result.

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 percentage of events captured in hotspots by up to 35% compared to commonly used methods for crime, traffic and IED event data. This 35% improvement could have important policy implications, as hotspot policing has been shown to yield greater crime rate reductions when the PAI of the hotspots is higher (Mohler et al. 2015). Beyond hotspot policing, CrimeRank may be used in conjunction with other proactive efforts such as community policing (Weisburd et al. 2020) and direct alerts for citizens (Groff and Taniguchi 2019).

In this work we restricted our attention to searching for rectangularly shaped hotspots. While we do develop the offgrid approach that considers shifting, rotating, and scaling the rectangles, hotspots with more general shapes may better capture location specific geometries and lead to higher PAI scores. Furthermore, it may be advantageous to consider network versions of CrimeRank that more naturally align with event locations that are restricted to streets. Future research in these areas may lead to further improvements in accuracy. One other research question that needs to be addressed in the future is how off-grid, rotated and non-standard polygon representations of crime hotspots may impact end-user trust in event forecasts. There also are data structure advantages and disadvantages of the method relative to
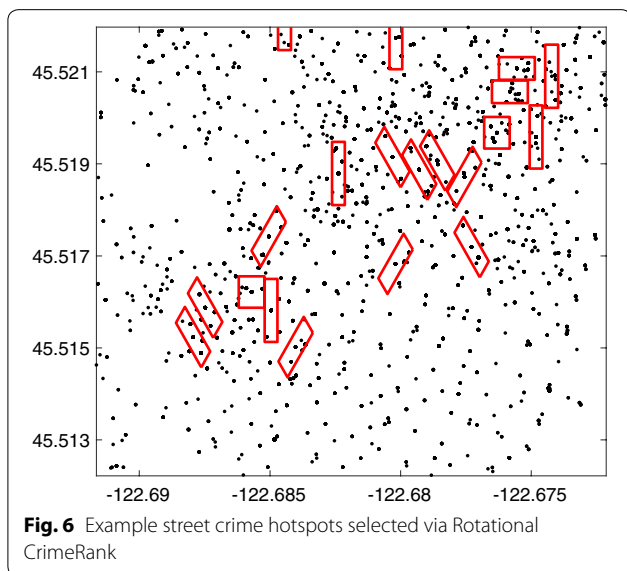
Fig. 6 Example street crime hotspots selected via Rotational CrimeRank

Mohler *et al. Crime Sci*   (2020) 9:3

Page 10 of 12

Table 3 PAI results of CrimeRank vs. baseline models for NIJ Competition Data

	CrimeRank Rotational	CrimeRank	Random forest	GLM	Hawkes	CNN-LSTM	KDE	GBM-NDCG	RGPM
All calls	64.5	61.5	19.4	57.5	57.5	58.4	36.3	57.8	58.8
Street	100.6	91.6	68.7	84.2	84.0	75.0	58.3	80.9	87.0

spatial rasters. We also only considered forecasts over 1-week and 3-month intervals in this paper and in the future it would be useful to consider hourly forecasting that can capture daily and hourly trends in crime.

While hotspot policing has been shown to yield crime rate reductions, there is the possibility of unwanted side effects of hotspot policing such as traffic stops that unfairly target minority populations, stop and frisk, and other police activities that have negative societal consequences. There has been some recent work on improving fairness of spatial crime forecasting algorithms (Wheeler 2019; Mohler et al. 2018) where a fairness penalty is added to the optimization algorithm. Future research may focus on incorporating fairness into learning to rank models of crime, similar to methods that incorporate fairness into learning to rank for information retrieval (Zehlike and Castillo 2018).

The methods introduced here will complement recent work on the incorporation of social sensing data into crime predictions (Wang et al. 2012, 2016; Bogomolov et al. 2014; Khosla et al. 2014). For example, real-time human movement data collected via smart phones or fixed city sensors has been shown to improve crime hotspot prediction accuracy. Implementing real-time, offgrid learning to rank and spatial scan methods at scale presents several computational and algorithmic challenges. The current model takes several minutes to hours to train on a laptop for each dataset. While this is not an issue for commercial predictive analytics software that runs in dynamic cloud servers, the runtime may be too long for desktop solutions used by crime analysts. Making these methods faster will be another focus of future research.

Abbreviations
PAI: Predictive accuracy index; IED: Improvised Explosive Device; KDE: Kernel density estimation; LGCP: log-Gaussian Cox Process; NDCG: Normalized Discounted Cumulative Gain; NIJ: National Institute of Justice; GBM: Gradient boosted machine; GLM: Generalized linear model; CNN: Convolution neural network; LSTM: Long short-term memory; IMPD: Indianapolis Metropolitan Police Department.

Acknowledgements
None.

Authors' contributions
All authors contributed equally. All authors read and approved the final manuscript.

Author details
[1] Indiana University Purdue University Indianapolis, Indianapolis, USA. [2] University of Virginia, Charlottesville, USA. [3] University of Maryland, College Park, USA.

References
Alves, L. G. A., Ribeiro, H. V., & Rodrigues, F. A. (2018). Crime prediction through urban metrics and statistical learning. *Physica A: Statistical Mechanics and its Applications*, *505*, 435–443.

Assunção, R., & Correa, T. (2009). Surveillance to detect emerging space-time clusters. *Computational Statistics & Data Analysis*, *53*(8), 2817–2830.

Berk, R., Sherman, L., Barnes, G., Kurtz, E., & Ahlman, L. (2009). 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)*, *172*(1), 191–211.

Berman, E., Shapiro, J. N., & Felter, J. H. (2011). Can hearts and minds be bought? the economics of counterinsurgency in Iraq. *Journal of Political Economy*, *119*(4), 766–819.

Bernasco, W. (2008). Them again? same-offender involvement in repeat and near repeat burglaries. *European Journal of Criminology*, *5*(4), 411–431.

Bogomolov, A., Lepri, B., Staiano, J., Oliver, N., Pianesi, F., & Pentland, A. (2014). Once upon a crime: towards crime prediction from demographics and mobile data. In *Proceedings of the 16th international conference on multi-modal interaction* (pp. 427–434). New York: ACM.

Braga, A. A., Turchan, B. S., Papachristos, A. V., & Hureau, D. M. (2019). Hot spots policing and crime reduction: An update of an ongoing systematic review and meta-analysis. *Journal of Experimental Criminology*, *15*(3), 289–311.

Burges, C. J. C. (2010). From ranknet to lambdarank to lambdamart: An overview. *Learning*, *11*(23–581), 81.

Caplan, J. M., Kennedy, L. W., & Miller, J. (2011). Risk terrain modeling: Brokering criminological theory and gis methods for crime forecasting. *Justice Quarterly*, *28*(2), 360–381.

Carter, J. G., & Piza, E. (2018). Spatiotemporal convergence of crime and vehicle crash hot spots: Additional consideration for policing places. *Crime & Delinquency*, *64*(14), 1795–1819. https://doi.org/10.1177/0011128717714793.

Chainey, S., Tompson, L., & Uhlig, S. (2008). The utility of hotspot mapping for predicting spatial patterns of crime. *Security Journal*, *21*(1), 4–28.

Cohen, J., Gorr, W. L., & Olligschlaeger, A. M. (2007). Leading indicators and spatial interactions: A crime-forecasting model for proactive police deployment. *Geographical Analysis*, *39*(1), 105–127.

Crimerank. (2018). https://github.com/gomohler/crimerank.

Curiel, R.P. (2019). Is crime concentrated or are we simply using the wrong metrics? arXiv preprint arXiv:1902.03105.

Mohler *et al. Crime Sci*   (2020) 9:3

Page 11 of 12

Diggle, P. J., Moraga, P., Rowlingson, B., Taylor, B. M., et al. (2013). Spatial and spatio-temporal log-gaussian cox processes: Extending the geostatistical paradigm. *Statistical Science*, *28*(4), 542–563.

Drawve, G., Belongie, M., & Steinman, H. (2017). The role of crime analyst and researcher partnerships: A training exercise in green bay, wisconsin. *Policing: A Journal of Policy and Practice*.

Duczmal, L., Cançado, A. L. F., & Takahashi, R. H. C. (2008). Delineation of irregularly shaped disease clusters through multiobjective optimization. *Journal of Computational and Graphical Statistics*, *17*(1), 243–262.

Duczmal, L., Kulldorff, M., & Huang, L. (2006). Evaluation of spatial scan statistics for irregularly shaped clusters. *Journal of Computational and Graphical Statistics*, *15*(2), 428–442.

Fischer, H. (2008). *Iraqi civilian casualties estimates*. Washington DC: Library of congress Washington DC congressional research service.

Flaxman, S., Chirico, M., Pereira, P., & Loeffler, C. (2018). Scalable high-resolution forecasting of sparse spatiotemporal events with kernel methods: A winning solution to the nij "real-time crime forecasting challenge". arXiv preprint arXiv:1801.02858.

Friedman, J. H. (2002). Stochastic gradient boosting. *Computational Statistics & Data Analysis*, *38*(4), 367–378.

Gao, P., Guo, D., Liao, K., Webb, J. J., & Cutter, S. L. (2013). Early detection of terrorism outbreaks using prospective space-time scan statistics. *The Professional Geographer*, *65*(4), 676–691.

Gorr, W. L. (2009). Forecast accuracy measures for exception reporting using receiver operating characteristic curves. *International Journal of Forecasting*, *25*(1), 48–61.

Gorr, W. L., & Lee, Y. J. (2015). Early warning system for temporary crime hot spots. *Journal of Quantitative Criminology*, *31*(1), 25–47.

Groff, E., & Taniguchi, T. (2019). Using citizen notification to interrupt near-repeat residential burglary patterns: the micro-level near-repeat experiment. *Journal of Experimental Criminology*, *15*(2), 115–149.

Groß, W., Lange, S., Bödecker, J., & Blum, M. (2017). Predicting time series with space-time convolutional and recurrent neural networks. In *Proceeding of European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning* (pp. 71–76).

Haberman, C. P., & Ratcliffe, J. H. (2012). The predictive policing challenges of near repeat armed street robberies. *Policing: A Journal of Policy and Practice*, *6*(2), 151–166.

Johnson, S. D., Bernasco, W., Bowers, K. J., Elffers, H., Ratcliffe, J., Rengert, G., et al. (2007). Space-time patterns of risk: A cross national assessment of residential burglary victimization. *Journal of Quantitative Criminology*, *23*(3), 201–219.

Kennedy, L. W., Caplan, J. M., & Piza, E. (2011). 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.

Khosla, A., An An, B., L., J.J., & Torralba, A. (2014). Looking beyond the visible scene. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 3710–3717).

Kulldorff, M. (2001). Prospective time periodic geographical disease surveillance using a scan statistic. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *164*(1), 61–72.

Kuo, P.-F., Lord, D., & Walden, T. D. (2013). Using geographical information systems to organize police patrol routes effectively by grouping hotspots of crash and crime data. *Journal of Transport Geography*, *30*, 138–148.

Leonard, B. (2009). *Measuring stability and security in Iraq*. Darby: DIANE Publishing.

Lewis, E., & Mohler, G. (2011). A nonparametric EM algorithm for multiscale Hawkes processes. preprint.

Liu, H., & Brown, D. E. (2003). Criminal incident prediction using a point-pattern-based density model. *International Journal of Forecasting*, *19*(4), 603–622.

Liu, T.-Y., et al. (2009). Learning to rank for information retrieval. *Foundations and Trends® in Information Retrieval*, *3*(3), 225–331.

McCollister, K. E., French, M. T., & Fang, H. (2010). The cost of crime to society: New crime-specific estimates for policy and program evaluation. *Drug & Alcohol Dependence*, *108*(1), 98–109.

Mohler, G., & Porter, M.D. (2017). Rotational grid, PAI-maximizing crime forecasts. *NIJ Report*.

Mohler, G., Raje, R., Carter, J., Valasik, M., & Brantingham, J. (2018). A penalized likelihood method for balancing accuracy and fairness in predictive policing. In *2018 IEEE International Conference on Systems, Man, and Cybernetics (SMC)* (pp. 2454–2459). New York: IEEE.

Mohler, G., et al. (2013). Modeling and estimation of multi-source clustering in crime and security data. *The Annals of Applied Statistics*, *7*(3), 1525–1539.

Mohler, G., Brantingham, P. J., Carter, J., & Short, M. B. (2019). Reducing bias in estimates for the law of crime concentration. *Journal of Quantitative Criminology*, *35*, 747–765.

Mohler, G. O., Short, M. B., Brantingham, P. J., Schoenberg, F. P., & Tita, G. E. (2011). Self-exciting point process modeling of crime. *Journal of the American Statistical Association*, *106*(493), 100–108.

Mohler, G. O., Short, M. B., Malinowski, S., Johnson, M., Tita, G. E., Bertozzi, A. L., et al. (2015). Randomized controlled field trials of predictive policing. *Journal of the American Statistical Association*, *110*(512), 1399–1411.

National Insititue of Justice. Nij real-time crime forecasting challenge, 2017.

Neill, D. B. (2009). Expectation-based scan statistics for monitoring spatial time series data. *International Journal of Forecasting*, *25*(3), 498–517.

Neill, D. B. (2012). Fast subset scan for spatial pattern detection. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *74*(2), 337–360.

Nobles, M. R., Ward, J. T., & Tillyer, R. (2016). The impact of neighborhood context on spatiotemporal patterns of Burglary. *Journal of Research in Crime and Delinquency*, *53*(5), 711–740.

Perry, W. L. (2013). *Predictive policing: The role of crime forecasting in law enforcement operations*. Santa Monica: Rand Corporation.

Piza, E.L., & Carter, J.G. (2017). Predicting initiator and near repeat events in spatiotemporal crime patterns: An analysis of residential burglary and motor vehicle theft. *Justice Quarterly* (pp. 1–29).

Piza, E. L., & Carter, J. G. (2018). Predicting initiator and near repeat events in spatiotemporal crime patterns: An analysis of residential burglary and motor vehicle theft. *Justice Quarterly*, *35*(5), 842–870.

Porter, M. D., & Reich, B. J. (2012). Evaluating temporally weighted kernel density methods for predicting the next event location in a series. *Annals of GIS*, *18*(3), 225–240.

Ratcliffe, J. H., & Rengert, G. F. (2008). Near-repeat patterns in philadelphia shootings. *Security Journal*, *21*(1–2), 58–76.

Ridgeway, G. (2007). Generalized boosted models: A guide to the gbm package. *Update*, *1*(1), 2007.

Shaw, B., Shea, J., Sinha, S., & Hogue, A. (2013). Learning to rank for spatiotemporal search. In *Proceedings of the sixth ACM international conference on Web search and data mining* (pp. 717–726). ACM.

Shirota, S., Gelfand, A. E., et al. (2017). Space and circular time log gaussian cox processes with application to crime event data. *The Annals of Applied Statistics*, *11*(2), 481–503.

Short, M. B., D'Orsogna, M. R., Brantingham, P. J., & Tita, G. E. (2009). Measuring and modeling repeat and near-repeat burglary effects. *Journal of Quantitative Criminology*, *25*(3), 325–339.

Speakman, S., Somanchi, S., McFowland, E, I. I. I., & Neill, D. B. (2016). Penalized fast subset scanning. *Journal of Computational and Graphical Statistics*, *25*(2), 382–404.

Stec, A., & Klabjan, D. (2018). Forecasting crime with deep learning. arXiv preprint arXiv:1806.01486.

Taddy, M. A. (2010). Autoregressive mixture models for dynamic spatial poisson processes: Application to tracking intensity of violent crime. *Journal of the American Statistical Association*, *105*(492), 1403–1417.

Tango, T., & Takahashi, K. (2005). A flexibly shaped spatial scan statistic for detecting clusters. *International journal of health geographics*, *4*(1), 11.

Wang, B., Yin, P., Bertozzi, A.L., Brantingham, P.J., Osher, S.J., & Xin, J. (2017). Deep learning for real-time crime forecasting and its ternarization. arXiv preprint arXiv:1711.08833.

Wang, H., Kifer, Daniel, G., Corina, & Li, Z. (2016). Crime rate inference with big data. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 635–644). New York: ACM.

Wang, X., & Brown, D. E. (2012). The spatio-temporal modeling for criminal incidents. *Security Informatics*, *1*(1), 1–17.

Wang, X., Gerber, M.S., & Brown, D.E. (2012). Automatic crime prediction using events extracted from twitter posts. In *International Conference on Social Computing, Behavioral-Cultural Modeling, and Prediction* (pp. 231–238). Berlin: Springer.

Mohler *et al. Crime Sci*   *(2020) 9:3*

Page 12 of 12

Weisburd, D., Gill, C., Wooditch, A., Barritt, W., & Murphy, J. (2020). Building collective action at crime hot spots: Findings from a randomized field experiment. *Journal of Experimental Criminology*, 1–31.

Weisburd, D. (2015). The law of crime concentration and the criminology of place. *Criminology*, *53*(2), 133–157.

Wheeler, A.P. (2019). Allocating police resources while limiting racial inequality. *Justice Quarterly*, 1–27.

Youstin, T. J., Nobles, M. R., Ward, J. T., & Cook, C. L. (2011). Assessing the generalizability of the near repeat phenomenon. *Criminal Justice and Behavior*, *38*(10), 1042–1063.

Zammit-Mangion, A., Dewar, M., Kadirkamanathan, V., & Sanguinetti, G. (2012). Point process modelling of the afghan war diary. *Proceedings of the National Academy of Sciences*, *109*(31), 12414–12419.

Zehlike, M., & Castillo, C. (2018). Reducing disparate exposure in ranking: A learning to rank approach. arXiv preprint arXiv:1805.08716.

Publisher's Note