Multi-Type Urban Crime Prediction

Xiangyu Zhao[1], **Wenqi Fan**[2*], **Hui Liu**[3], **Jiliang Tang**[3]

[1]City University of Hong Kong, [2]The Hong Kong Polytechnic University, [3]Michigan State University
xianzhao@cityu.edu.hk, wenqifan03@gmail.com, {liuhui7,tangjili}@msu.edu

Abstract

Crime prediction plays an impactful role in enhancing public security and sustainable development of urban. With recent advances in data collection and integration technologies, a large amount of urban data with rich crime-related information and fine-grained spatio-temporal logs have been recorded. Such helpful information can boost our understandings of the temporal evolution and spatial factors of urban crimes and can enhance accurate crime prediction. However, the vast majority of existing crime prediction algorithms either do not distinguish different types of crime or treat each crime type separately, which fails to capture the intrinsic correlations among different types of crime. In this paper, we perform crime prediction exploiting the cross-type and spatio-temporal correlations of urban crimes. In particular, we verify the existence of correlations among different types of crime from temporal and spatial perspectives, and propose a coherent framework to mathematically model these correlations for crime prediction. Extensive experiments on real-world datasets validate the effectiveness of our framework.

Introduction

It is well recognized that crime prediction is of great importance for enhancing the public security of urban so as to improve the life quality of citizens (Couch and Dennemann 2000). Efforts have been made on constructing crime prediction models to predict either the total crime amount (Zhao and Tang 2017b) or several specific types of crime such as *Burglary* (Wang and Liu 2017), *Felony Assault* (Barrett, Katsiyannis, and Zhang 2006), *Grand Larceny* (Fisher 1999), *Murder* (Revitch and Schlesinger 1978), *Rape* (Thornhill and Thornhill 1983), *Robbery* (Roesch and Winterdyk 1986), and *Vehicle Larceny* (Henry and Bryan 2000). In other words, most existing crime prediction methods either do not distinguish different types of crime or consider each crime type separately.

According to criminology and recent studies, different types of crime behave differently but are intrinsically correlated. For instance, social disorganization theory (Sampson and Groves 1989) and broken windows theory (Wilson and Kelling 1982) suggest that a series of minor crimes like vandalism or graffiti might cause the increase of more severe crimes like assaults and weapon violence; while relations between different types of crime in London are investigated, where *bicycle theft*, *burglary*, *robbery* and *theft from the person* are observed to be closely related in terms of spatial distribution. The above theories and observations indicate that different types of crime are intrinsically related to each other, and exploiting the correlations among crime types could boost accurate crime prediction.

Recently, driven by the advances in big urban data collection and integration techniques, a great quantity of urban data has been collected such as crime complaint data, stop-and-frisk data and 311 public-service complaint data (Zhao et al. 2016, 2017; Liu et al. 2017, 2020). Such data contains rich and useful context information about crime. For example, in the near future, more crimes tend to occur in the areas with many crime complaints; while public-service complaint data reveal citizens' dissatisfaction with government service, thus it is associated with crimes (Huang et al. 2018). In addition, big urban data contains fine-grained information about where and when the data is collected (Guo et al. 2016; Xu et al. 2016, 2018, 2019). Such spatio-temporal information not only enables us to study the geographical factors of crimes such as urban configuration, but also allows us to understand the dynamics and evolution of crimes over time (Leong and Sung 2015). According to environmental criminologies like awareness theory (Brantingham and Brantingham 1981) and crime pattern theory (Felson and Clarke 1998), the distribution of urban crimes is highly influenced by space and time. Thus, the spatio-temporal understandings from big urban data provide unprecedented opportunities for us to construct accurate crime predictions.

In this paper, we jointly explore cross-type and spatio-temporal correlations for crime prediction by leveraging big urban data. Specifically, we mainly seek answers for two challenging questions: (1) what correlations can be observed among different types of crime, and (2) how to mathematically model cross-type and spatio-temporal correlations for crime prediction. For cross-type correlations, we investigate temporal and spatial patterns of different types of crimes as well as their relationships; for spatio-temporal correlations, we focus our investigation on mathematically modeling (1) intra-region temporal correlation that suggests how crime evolves over time in a region, and (2) inter-region spatial cor-

*Wenqi Fan is corresponding author.

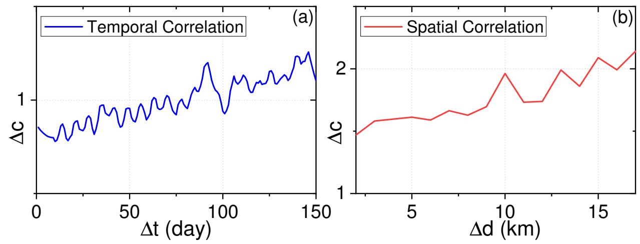

Figure 1: Temporal/spatial correlations of Grand Larceny.

relation that depicts the spatial relationship across regions in the city (Zhao and Tang 2018). We propose a novel framework **CCC**, which jointly captures **C**ross-type and spatio-temporal **C**orrelations for **C**rime prediction based on urban data. We conduct extensive experiments on real big urban datasets to validate the effectiveness of the CCC framework.

Problem Statement

This section introduces the mathematical notations and then formally defines the problem we study in this work.

Let $\mathbf{Y} \in \mathbb{R}^{N \times T \times K}$ denote the observed numbers of crime where $Y_n^t(k)$ is the number of k^{th} crime type observed at n^{th} region in t^{th} time slot. Here we suppose that there are totally (1) N regions in a city, i.e., $n = \{1, 2, \ldots, N\} \in \mathbb{R}^N$, (2) T time slots (e.g. days, weeks, or months) in the dataset, i.e., $t = \{1, 2, \ldots, T\} \in \mathbb{R}^T$, and (3) K types of crime (e.g. burglary, robbery and grand larceny), i.e., $k = \{1, 2, \ldots, K\} \in \mathbb{R}^K$. Suppose that $\mathbf{X} \in \mathbb{R}^{N \times T \times M}$ denotes the set of feature vectors, where $\mathbf{X}_n^t \in \mathbb{R}^{1 \times M}$ is the feature vector of n^{th} region in t^{th} time slot, and M is the number of features. Note that feature vector $\mathbf{X}_n^t$ is same for all types of crime of n^{th} region in t^{th} time slot. More details about features will be proposed in the experiment section.

With the above-mentioned notations and definitions, we formally state the problem of crime prediction as: *Given the observed historical crime amounts* $\mathbf{Y}$ *and feature vectors* $\mathbf{X}$*, we aim to predict the crime amount of time slot* $T + \tau$ *(or* τ *time slots later) for each type of crime based on* $\mathbf{Y}$ *and* $\mathbf{X}$.

It should be noted that our goal is to predict crime amount for future time slot $T + \tau$. However, if the feature vector $\mathbf{X}_n^t$ is constructed based on data in t^{th} time slot of n^{th} region, the future feature vector of $(T + \tau)^{th}$ time slot is not available. To this end, in this paper, we actually construct $\mathbf{X}_n^t$ using data in $(t - \tau)^{th}$ time slot rather than t^{th} time slot of n^{th} region. Without the loss of generality, in the following sections, we leverage $\tau = 1$ for illustrations, i.e., performing crime prediction for $(T + 1)^{th}$ time slot.

Preliminary Study

This section investigates spatio-temporal and cross-type correlations for different crime types in New York City, which contains 7 types, i.e., Burglary, Felony Assault, Grand Larceny, Murder, Rape, Robbery, and Vehicle Larceny.

Spatio-Temporal Correlations

Within a region, the amount of crime should change smoothly over time. We assume the crime amount is c_t and $c_{t+\Delta t}$ for time t and $t + \Delta t$. To study temporal correlation, we show how the crime amount differences $\Delta c = |c_t - c_{t+\Delta t}|$ changes with Δt on average of all regions. The result is illustrated in Figure 1(a) where x-axis is Δt (days) and y-axis is Δc. We can observe that the crime differences are highly related to Δt. To be specific, (i) two consecutive time slots share similar crime amounts; (ii) with the increase of Δt, the crime difference is likely to increase.

For regions in the city, if two regions are spatially close to each other, they are likely to have similar crime amounts at the same time slot. Given a pair of regions, we leverage Δd as their spatial distance and use Δc as their absolute crime difference. We show how Δc changes with Δd averaged over all time slots in Figure 1(b), where x-axis is Δd and y-axis is Δc. We note that (i) when two regions are spatially close, they have similar crime amounts, and (ii) with the increase of distance Δd, the Δc tends to increase.

The above observations suggest the existence of temporal and spatial correlations for each type of urban crime. Note that we omit other crime types with similar observations.

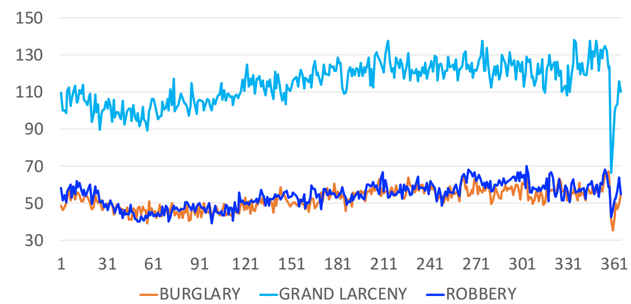

Figure 2: Average daily crime numbers from 2012 to 2015.

Cross-Type Correlations

To investigate temporal correlations among different crime types, we study how crime amounts of each type change with the days of a year. The average daily crime amounts from 2012 to 2015 are shown in Figure 2, where x-axis denotes the days of a year and y-axis is the crime amounts. We can observe obvious temporal correlations between Burglary, Grand Larceny and Robbery. Specifically, the daily crime amounts of each type increase from March to September and tend to decrease from January to February. Furthermore, the crime amounts increase in December before Christmas, but decrease dramatically during the Christmas and New Year. We omit other crime types having similar patterns at a different scale.

To study the spatial correlations among crime types, we show how crimes are spatially distributed in New York City of 2012 in Figure 3. We make the observations that (1) the majority of types concentrate in the Bronx and Manhattan; and (2) Burglary, Rape, Robbery, and Vehicle Larceny share some occurrence hotspots in the Brooklyn and Queens.

To study the correlations from both temporal and spatial perspectives, we first construct $K = 7$ matrices $\{\mathbf{R}^1, \mathbf{R}^2, \ldots, \mathbf{R}^K\}$, where each $\mathbf{R}^k \in \mathbb{R}^{N \times T}$. Each element $R_{n,t}^k \in \mathbf{R}^k$ is the crime amount for k^{th} type of crime

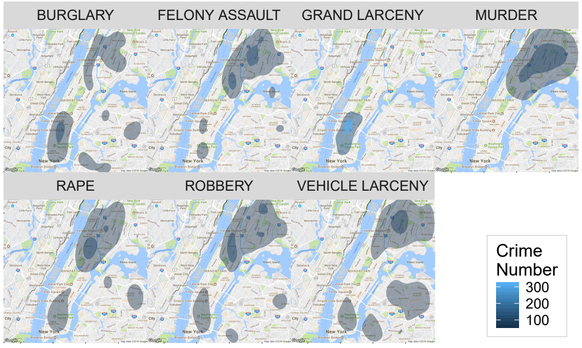

Figure 3: Spatial distribution of crimes in 2012.

in n^{th} region of t^{th} time slot. To study the spatio-temporal correlations between two types (e.g. the i^{th} and j^{th} type) of crime, we calculate the variant of *cosine similarity* between $\mathbf{R}^i$ and $\mathbf{R}^j$ as follows:

$$cosine(\mathbf{R}^i, \mathbf{R}^j) = \frac{<\mathbf{R}^i, \mathbf{R}^j>}{\|\mathbf{R}^i\|_F \|\mathbf{R}^j\|_F} \tag{1}$$

where $<\mathbf{R}^i, \mathbf{R}^j> = \sum_{n,t} R^i_{n,t} R^j_{n,t}$ and $\|\cdot\|_F$ is the *Frobenius norm*. The result is shown in Figure 4. We can observe that most types of crime are indeed correlated with each other. The least spatio-temporal correlation exists between Grand Larceny and Murder.

To sum up, we demonstrate the existence of temporal and spatial correlations among different types of crime. These observations provide the groundwork for leveraging the cross-type correlations for accurate crime prediction.

Framework

In this section, we first present the basic model without cross-type and spatio-temporal correlations, then propose the details of introducing cross-type correlations as well as spatio-temporal correlations into a coherent framework. Finally, we discuss the optimization process and leverage the framework to perform crime prediction.

The Basic Model

Without considering cross-type and spatio-temporal correlations, we build a basic and individual model of k^{th} crime type for n^{th} region in t^{th} time slot. Correspondingly, there is a weight vector $\mathbf{W}^t_n(k) \in \mathbb{R}^{M\times 1}$ for k^{th} crime type of n^{th} region in t^{th} time slot, which can map $\mathbf{X}^t_n$ to $Y^t_n(k)$ as: $\mathbf{X}^t_n \mathbf{W}^t_n(k) \rightarrow Y^t_n(k)$. All $\mathbf{W}^t_n(k)$ can be learned by solving the following regression problem:

$$\min_{\mathbf{W}^t_n(k)} \sum_{n=1}^{N}\sum_{t=1}^{T}\sum_{k=1}^{K} \left(\mathbf{X}^t_n \mathbf{W}^t_n(k) - Y^t_n(k)\right)^2 \tag{2}$$

where we use the square loss function for regression task in this work. Note that it is straightforward to leverage other loss functions such as logistic loss and hinge loss. This basic and individual model completely neglects the existence of correlations among different types of crime and spatio-temporal correlations within each type of crime.

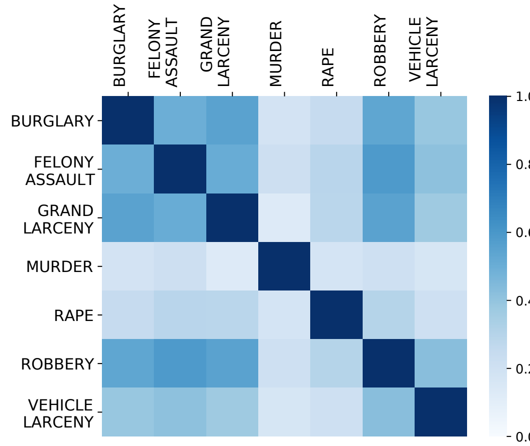

Figure 4: Cross-type correlation matrix heatmap.

Cross-Type Correlations

To exploit correlations of urban crimes, we first decompose the weight vector $\mathbf{W}^t_n(k)$ into the sum of two components $\mathbf{W}^t_n(k) = \mathbf{P}^t_n + \mathbf{Q}^t_n(k)$ (see Page 28 in (Zhou, Chen, and Ye 2012) as illustration), where we use $\mathbf{P}^t_n$ to capture the *common patterns* shared by all crime types, e.g., all types of crimes tend to occur more in the regions with more human check-ins; $\mathbf{Q}^t_n(k)$ captures the *specific patterns* for k^{th} crime type, e.g., Burglary occurs more in the residential areas. These common/specific patterns are not hand-craft, and will be automatically learned during the model optimization.

Since $\mathbf{Q}^t_n(k)$ captures the specific patterns for k^{th} crime type, we can model the cross-type correlations by capturing the relationships between crime $\mathbf{Q}^t_n(k), \forall k \in [1, K]$. We first combine all the type specific weight vectors into a weight matrix, i.e., $\mathbf{Q}^t_n = [\mathbf{Q}^t_n(1), \mathbf{Q}^t_n(2), \ldots, \mathbf{Q}^t_n(K)] \in \mathbb{R}^{M\times K}$. Then, adopting the task relationship regularization component in (Zhang and Yeung 2012), the relationships among crime types can be modeled as $tr\left(\mathbf{Q}^t_n {\mathbf{\Omega}^t_n}^{-1} \mathbf{Q}^{t\top}_n\right)$, where $\mathbf{\Omega}^t_n$ is the covariance matrix between crime types in n^{th} region of t^{th} time slot. Since $\mathbf{\Omega}^t_n$ is a covariance matrix, it should be positive semi-definite (i.e., $\mathbf{\Omega}^t_n \geq 0$). For all regions and time slots, we have:

$$\begin{aligned} &\sum_{n=1}^{N}\sum_{t=1}^{T} \alpha \cdot tr\left(\mathbf{Q}^t_n {\mathbf{\Omega}^t_n}^{-1} \mathbf{Q}^{t\top}_n\right) \\ &s.t. \quad \mathbf{\Omega}^t_n \geq 0 \qquad tr\left(\mathbf{\Omega}^t_n\right) = K \end{aligned} \tag{3}$$

where α is a non-negative hyper-parameter.

Intra-Region Temporal Correlation

Crime within a region is observed following intra-region temporal correlation – (1) for two consecutive time slots, they tend to share similar crime amounts; and (2) with the increase of distance between two time slots, the crime amounts difference is likely to increase. Inspired by this discovery, we propose a temporal regularization component to model the temporal correlations of crime amount within each region.

To be specific, considering the smooth evolution of crime amounts, the weight vectors should also change smoothly.

Therefore, we adopt a series of discrete weight vectors over time to represent the temporal dynamics of crime amounts, and we add a temporal regularization term to basic model:

$$\beta \sum_{n=1}^{N} \sum_{t=1}^{T-1} \left(\|\mathbf{P}_n^t - \mathbf{P}_n^{t+1}\|_1 + \sum_{k=1}^{K} \|\mathbf{Q}_n^t(k) - \mathbf{Q}_n^{t+1}(k)\|_1 \right) \quad (4)$$

where β is a non-negative parameter. The first term pushes $\mathbf{P}_n^t$ as closer as $\mathbf{P}_n^{t+1}$, i.e., the weight vector for common patterns shared by all crime types of n^{th} region change smoothly over time, while the second term captures the smooth evolution of weight vector for each specific crime type within a region. Note that we define $\|\mathbf{X}\|_1$ as $\sum_{i,j} |\mathbf{X}_{ij}|$ in this work, which makes it possible to encourage weight vectors of two consecutive time slots to be exactly same. We do not use ℓ_2-norm since it is likely to cause "wiggly" cost dynamics, which is not robust to noises and may hurt generalization (Zheng and Ni 2013). Eq. (4) is rewritten as:

$$\sum_{n=1}^{N} \left(\|\mathbf{P}_n \mathbf{A}\|_1 + \sum_{k=1}^{K} \|\mathbf{Q}_n(k)\mathbf{A}\|_1 \right) \quad (5)$$

where $\mathbf{P}_n = [\mathbf{P}_n^1, \mathbf{P}_n^2, \ldots, \mathbf{P}_n^T] \in \mathbb{R}^{M \times T}$ and $\mathbf{Q}_n(k) = [\mathbf{Q}_n^1(k), \mathbf{Q}_n^2(k), \ldots, \mathbf{Q}_n^T(k)] \in \mathbb{R}^{M \times T}$. $\mathbf{A} \in \mathbb{R}^{T \times (T-1)}$ is a sparse matrix. More specifically, $\mathbf{A}(t,t) = \beta$, $\mathbf{A}(t+1,t) = -\beta$ for $t = 1, \ldots, T-1$ and all the other terms 0.

Inter-Region Spatial Correlation

Aside from intra-region temporal correlation, the crime amounts across all regions follow inter-region spatial correlation – (1) two spatial close regions tend to have similar crime amounts; and (2) with the increase of geographical distance between two regions in a city, the crime difference between these two regions is likely to increase in a certain time slot. This observation inspires us to develop a spatial regularization component to capture the spatial correlation of crime amounts across regions in a city.

Specifically, we choose to minimize the following spatial component to capture inter-region spatial correlation:

$$\sum_{t=1}^{T} \sum_{i=1}^{N} \sum_{j=1}^{N} d(i,j)^{-\gamma} \left(\|\mathbf{P}_i^t - \mathbf{P}_j^t\|_1 + \sum_{k=1}^{K} \|\mathbf{Q}_i^t(k) - \mathbf{Q}_j^t(k)\|_1 \right) \quad (6)$$

where $d(i,j)$ is the spatial distance between i^{th} and j^{th} region. $d(i,j)^{-\gamma}$ is a power law exponential function, which is non-increase in terms of $d(i,j)$, where γ is the parameter controlling the degree of spatial correlations. Thus, when i^{th} and j^{th} regions are closer (i.e. $d(i,j)$ is smaller), $d(i,j)^{-\gamma}$ becomes larger that enforces weight vectors of two regions to be closer. Similar analysis can be used when the distance between i^{th} and j^{th} is larger.

Similar to intra-region temporal correlation, the first term pushes $\mathbf{P}_i^t$ and $\mathbf{P}_j^t$ to be closer, which means the weight vector for common patterns of all crime types in i^{th} and j^{th} region is similar if they are spatially close to each other. The second term captures the proximity across regions of each type of crime. This spatial component encodes Tobler's first law of geography (Tobler 1970) and performs a soft constraint that spatially close regions tend to have similar weight vectors. We can rewrite Eq. (6) as:

$$\sum_{t=1}^{T} \left(\|\mathbf{P}^t \mathbf{B}\|_1 + \sum_{k=1}^{K} \|\mathbf{Q}^t(k)\mathbf{B}\|_1 \right) \quad (7)$$

where $\mathbf{P}^t = [\mathbf{P}_1^t, \mathbf{P}_2^t, \ldots, \mathbf{P}_N^t] \in \mathbb{R}^{M \times N}$ and $\mathbf{Q}^t(k) = [\mathbf{Q}_1^t(k), \mathbf{Q}_2^t(k), \ldots, \mathbf{Q}_N^t(k)] \in \mathbb{R}^{M \times N}$. $\mathbf{B} \in \mathbb{R}^{N \times N^2}$ is a sparse matrix. To be specific, we have $\mathbf{B}(i, (i-1)\cdot N + j) = d(i,j)^{-\gamma}$ and $\mathbf{B}(j, (i-1)\cdot N + j) = -d(i,j)^{-\gamma}$ for $i, j \in [1, N]$ and $i \neq j$, while all the other terms 0.

Optimization

With above components for cross-type and spatio-temporal correlations, the objective loss function of the proposed framework is to solve the following optimization task:

$$\begin{aligned}
\min_{\mathbf{P},\mathbf{Q},\mathbf{\Omega}} L = & \sum_{n=1}^{N} \sum_{t=1}^{T} \sum_{k=1}^{K} \left(\frac{\mathbf{X}_n^t \left(\mathbf{P}_n^t + \mathbf{Q}_n^t(k) \right) - Y_n^t(k)}{\mathcal{Y}_n(k)} \right)^2 \\
& + \sum_{n=1}^{N} \sum_{t=1}^{T} \alpha \cdot tr\left(\mathbf{Q}_n^t {\mathbf{\Omega}_n^t}^{-1} \mathbf{Q}_n^{t\top} \right) \\
& + \sum_{n=1}^{N} \left(\|\mathbf{P}_n \mathbf{A}\|_1 + \sum_{k=1}^{K} \|\mathbf{Q}_n(k)\mathbf{A}\|_1 \right) \\
& + \sum_{t=1}^{T} \left(\|\mathbf{P}^t \mathbf{B}\|_1 + \sum_{k=1}^{K} \|\mathbf{Q}^t(k)\mathbf{B}\|_1 \right) \\
s.t. \quad & \mathbf{\Omega}_n^t \geq 0 \qquad \forall n \in [1,N] \quad \forall t \in [1,T] \\
& tr\left(\mathbf{\Omega}_n^t\right) = K \qquad \forall n \in [1,N] \quad \forall t \in [1,T]
\end{aligned} \quad (8)$$

Since the crime amounts can vary a lot in types and regions, we respectively normalize $\mathbf{X}_n^t \left(\mathbf{P}_n^t + \mathbf{Q}_n^t(k) \right) - Y_n^t(k)$ by the historical maximum of each crime type in each region, i.e., a constant $\mathcal{Y}_n(k)$, then the types or regions with much higher numbers cannot dominate the overall loss function.

In this work, we leverage ADMM technique (Boyd et al. 2011) to optimize the objective loss function Eq. (8). We first suppose $\mathbf{C}_n = \mathbf{P}_n \mathbf{A} \in \mathbb{R}^{M \times T-1}$, $\mathbf{D}_n(k) = \mathbf{Q}_n(k)\mathbf{A} \in \mathbb{R}^{M \times T-1}$, $\mathbf{E}^t = \mathbf{P}^t \mathbf{B} \in \mathbb{R}^{M \times N^2}$ and $\mathbf{F}^t(k) = \mathbf{Q}^t(k)\mathbf{B} \in \mathbb{R}^{M \times N^2}$, where $\mathbf{C}_n$, $\mathbf{D}_n(k)$, $\mathbf{E}^t$ and $\mathbf{F}^t(k)$ are *auxiliary variable matrices* in ADMM. Then the objective becomes:

$$\begin{aligned}
\min_{\mathbf{P},\mathbf{Q},\mathbf{\Omega}} L = & \sum_{n=1}^{N} \sum_{t=1}^{T} \sum_{k=1}^{K} \frac{1}{\mathcal{Y}_n(k)^2} \left(\mathbf{X}_n^t \left(\mathbf{P}_n^t + \mathbf{Q}_n^t(k) \right) - Y_n^t(k) \right)^2 \\
& + \sum_{n=1}^{N} \sum_{t=1}^{T} \alpha \cdot tr\left(\mathbf{Q}_n^t {\mathbf{\Omega}_n^t}^{-1} \mathbf{Q}_n^{t\top} \right) \\
& + \sum_{n=1}^{N} \left(\|\mathbf{C}_n\|_1 + \sum_{k=1}^{K} \|\mathbf{D}_n(k)\|_1 \right) \\
& + \sum_{t=1}^{T} \left(\|\mathbf{E}^t\|_1 + \sum_{k=1}^{K} \|\mathbf{F}^t(k)\|_1 \right) \\
s.t. \quad & \mathbf{\Omega}_n^t \geq 0 \qquad tr\left(\mathbf{\Omega}_n^t\right) = K \\
& \mathbf{C}_n = \mathbf{P}_n \mathbf{A} \qquad \mathbf{D}_n(k) = \mathbf{Q}_n(k)\mathbf{A} \\
& \mathbf{E}^t = \mathbf{P}^t \mathbf{B} \qquad \mathbf{F}^t(k) = \mathbf{Q}^t(k)\mathbf{B} \\
& \forall n \in [1,N] \quad \forall t \in [1,T] \quad \forall k \in [1,K]
\end{aligned} \quad (9)$$

Then the scaled form of ADMM optimization formulation of Eq. (9) can be written as:

$$
\begin{aligned}
&\min L_\rho(\mathbf{P}, \mathbf{Q}, \mathbf{\Omega}, \mathbf{C}, \mathbf{D}, \mathbf{E}, \mathbf{F}, \mathbf{S}, \mathbf{U}, \mathbf{V}, \mathbf{Z}) \\
&= \sum_{n=1}^{N}\sum_{t=1}^{T}\sum_{k=1}^{K} \frac{1}{\mathcal{Y}_n(k)^2}\left(\mathbf{X}_n^t\left(\mathbf{P}_n^t + \mathbf{Q}_n^t(k)\right) - Y_n^t(k)\right)^2 \\
&+ \sum_{n=1}^{N}\sum_{t=1}^{T} \alpha \cdot tr\left(\mathbf{Q}_n^t {\mathbf{\Omega}_n^t}^{-1} \mathbf{Q}_n^{t\top}\right) \\
&+ \sum_{n=1}^{N}\left(\|\mathbf{C}_n\|_1 + \frac{\rho}{2}\|\mathbf{P}_n\mathbf{A} - \mathbf{C}_n + \mathbf{S}_n\|_F^2\right) \\
&+ \sum_{n=1}^{N}\sum_{k=1}^{K}\left(\|\mathbf{D}_n(k)\|_1 + \frac{\rho}{2}\|\mathbf{Q}_n(k)\mathbf{A} - \mathbf{D}_n(k) + \mathbf{U}_n(k)\|_F^2\right) \\
&+ \sum_{t=1}^{T}\left(\|\mathbf{E}^t\|_1 + \frac{\rho}{2}\|\mathbf{P}^t\mathbf{B} - \mathbf{E}^t + \mathbf{V}^t\|_F^2\right) \\
&+ \sum_{t=1}^{T}\sum_{k=1}^{K}\left(\|\mathbf{F}^t(k)\|_1 + \frac{\rho}{2}\|\mathbf{Q}^t(k)\mathbf{B} - \mathbf{F}^t(k) + \mathbf{Z}^t(k)\|_F^2\right) \\
s.t. \quad & \mathbf{\Omega}_n^t \geq 0 \qquad tr\left(\mathbf{\Omega}_n^t\right) = K \\
& \forall n \in [1, N] \qquad \forall t \in [1, T]
\end{aligned} \tag{10}
$$

where $\|\cdot\|_F$ is the *Frobenius*-norm of a matrix. We introduce *scaled dual variable matrices* $\mathbf{S}_n \in \mathbb{R}^{M\times(T-1)}$, $\mathbf{U}_n(k) \in \mathbb{R}^{M\times(T-1)}$, $\mathbf{V}^t \in \mathbb{R}^{M\times N^2}$ and $\mathbf{Z}^t(k) \in \mathbb{R}^{M\times N^2}$ of ADMM. The penalty for the violation of equality constraints $\mathbf{C}_n = \mathbf{P}_n\mathbf{A}$, $\mathbf{D}_n(k) = \mathbf{Q}_n(k)\mathbf{A}$, $\mathbf{E}^t = \mathbf{P}^t\mathbf{B}$, $\mathbf{F}^t(k) = \mathbf{Q}^t(k)\mathbf{B}$ is controlled by a non-negative parameter ρ. According to ADMM technique, each optimization iteration of Eq (10) consists of the following steps:

$$\mathbf{P}_n^t \leftarrow \mathbf{P}_n^t - \eta\frac{\partial L_\rho}{\partial \mathbf{P}_n^t} \tag{11}$$

$$\mathbf{Q}_n^t(k) \leftarrow \mathbf{Q}_n^t(k) - \eta\frac{\partial L_\rho}{\partial \mathbf{Q}_n^t(k)} \tag{12}$$

$$\mathbf{\Omega}_n^t \leftarrow \frac{K\left(\mathbf{Q}_n^{t\top}\mathbf{Q}_n^t\right)^{1/2}}{tr\big(\left(\mathbf{Q}_n^{t\top}\mathbf{Q}_n^t\right)^{1/2}\big)} \tag{13}$$

$$\mathbf{C}_n \leftarrow S_{1/\rho}\big(\mathbf{P}_n\mathbf{A} + \mathbf{S}_n\big) \tag{14}$$

$$\mathbf{S}_n \leftarrow \mathbf{S}_n + \mathbf{P}_n\mathbf{A} - \mathbf{C}_n, \tag{15}$$

$$\mathbf{D}_n(k) \leftarrow S_{1/\rho}\big(\mathbf{Q}_n(k)\mathbf{A} + \mathbf{U}_n(k)\big) \tag{16}$$

$$\mathbf{U}_n(k) \leftarrow \mathbf{U}_n(k) + \mathbf{Q}_n(k)\mathbf{A} - \mathbf{D}_n(k) \tag{17}$$

$$\mathbf{E}^t \leftarrow S_{1/\rho}\big(\mathbf{P}^t\mathbf{B} + \mathbf{V}^t\big) \tag{18}$$

$$\mathbf{V}^t \leftarrow \mathbf{V}^t + \mathbf{P}^t\mathbf{B} - \mathbf{E}^t \tag{19}$$

$$\mathbf{F}^t(k) \leftarrow S_{1/\rho}\big(\mathbf{Q}^t(k)\mathbf{B} + \mathbf{Z}^t(k)\big) \tag{20}$$

$$\mathbf{Z}^t(k) \leftarrow \mathbf{Z}^t(k) + \mathbf{Q}^t(k)\mathbf{B} - \mathbf{F}^t(k) \tag{21}$$

where η is the learning rate of gradient descent. The *soft thresholding operator* $S_{1/\rho}(x)$ is defined as follows:

$$S_{1/\rho}(x) = \begin{cases} x - 1/\rho & if \quad x \ > \ 1/\rho \\ 0 & if \ \|x\| \leq \ 1/\rho \\ x + 1/\rho & if \quad x \ < -1/\rho \end{cases} \tag{22}$$

Crime Prediction Task

When optimization converges, it will output the well-trained weight vectors $\mathbf{P}_n^t$ and $\mathbf{Q}_n^t(k)$, $\forall n \in [1, N], t \in [1, T], k \in [1, K]$ respectively. In this subsection, we introduce how to perform crime prediction for a future time slot (i.e. $(T+1)^{th}$ time slot) based on well-trained $\mathbf{P}_n^t$ and $\mathbf{Q}_n^t(k)$.

As mentioned in Problem Statement, we actually construct feature vector $\mathbf{X}_n^t$ using data in $t-1^{th}$ time slot rather than t^{th} time slot of n^{th} region. Thus for the $T+1^{th}$ time slot, we can construct $\mathbf{X}_n^{T+1}$ based on data in T^{th} time slot. Therefore, in order to predict crime amount $Y_n^{T+1}(k) = \mathbf{X}_n^{T+1}\left(\mathbf{P}_n^{T+1} + \mathbf{Q}_n^{T+1}(k)\right)$ for k^{th} type of crime in n^{th} region of $T+1^{th}$ time slot, we need the mapping vectors $\mathbf{P}_n^{T+1}$ and $\mathbf{Q}_n^{T+1}(k)$. To sum up, the problem becomes to estimate $\mathbf{P}_n^{T+1}$ and $\mathbf{Q}_n^{T+1}(k)$ based on $\{\mathbf{P}_n^t\}_{t=1}^T$ and $\{\mathbf{Q}_n^t(k)\}_{t=1}^T$.

The mapping vectors $\mathbf{P}_n^t$ and $\mathbf{Q}_n^t(k)$ should be related to these of previous time slots according to intra-region temporal correlation. Therefore, we assume that $\mathbf{W}_n^t(k) = \mathbf{P}_n^t + \mathbf{Q}_n^t(k)$ is the weighted sum of its previous $\mathcal{G}$ time slots:

$$\mathbf{W}_n^t(k) = \frac{\sum_{\Delta t=1}^{\mathcal{G}} f(\Delta t)\left(\mathbf{P}_n^{t-\Delta t} + \mathbf{Q}_n^{t-\Delta t}(k)\right)}{\sum_{\Delta t=1}^{\mathcal{G}} f(\Delta t)} \tag{23}$$

where $f(\Delta t)$ should be a non-increase function of Δt, i.e., $f(\Delta t)$ should be larger when Δt is smaller, since $\mathbf{W}_n^t$ should be closer related to its just previous few time slots. In this work, we use a power law exponential function of $f(\Delta t) = \sigma^{-\Delta t}$, where $\sigma \in [1, +\infty)$ is introduced to control the contributions from $\{\mathbf{W}_n^{t-1}, \mathbf{W}_n^{t-2}, \ldots, \mathbf{W}_n^{t-\mathcal{G}}\}$. Note that when $\sigma = 1$, $\{\mathbf{W}_n^{t-1}, \mathbf{W}_n^{t-2}, \ldots, \mathbf{W}_n^{t-\mathcal{G}}\}$ contributes equally to $\mathbf{W}_n^t$. We propose to automatically estimate optimal σ from the training data via solving the following optimization problem:

$$\min_{\sigma} \sum_{t=\mathcal{G}+1}^{T}\left(\mathbf{X}_n^t \frac{\sum_{\Delta t=1}^{\mathcal{G}} \sigma^{-\Delta t}\left(\mathbf{P}_n^{t-\Delta t} + \mathbf{Q}_n^{t-\Delta t}(k)\right)}{\sum_{\Delta t=1}^{\mathcal{G}} \sigma^{-\Delta t}} - Y_n^t(k)\right)^2 \tag{24}$$

Experiments

In this section, we conduct extensive experiments to evaluate the effectiveness of the proposed framework. We seek to answer two questions: (1) how the proposed framework performs compared to the state-of-the-art baselines; and (2) how the cross-type correlations and spatio-temporal correlations benefit crime prediction.

Data

The data of K=7 types of crime is collected from 07/01/2012 to 06/30/2013 (T=365 days) in New York City. We respectively segment the city into disjointed $2km\times 2km$ grids (regions), and select N=100 regions with the most reported crimes. For the feature matrices, we collect multiple data resources that are related to crime: historical crime, stop-and-frisk, weather, Point of Interests (region function), human mobility and 311 public-service complaint data.

Experimental Setting

For each type of crime in each region, we leverage previous $\mathcal{T} = 7$ time slots' data to train the parameters since crime amounts are typically associated to recent previous time

Metrics	Prediction Time τ	Crime Prediction Models										
		ARIMA	VAR	LSTM	DMove	DST	STRN	CCRF	NCCRF	TCP	DCrime	CCC
RMSE	1-day	0.735	0.726	0.631	0.610	0.526	0.493	0.401	0.337	0.274	0.255	**0.223**
	7-day	1.2926	1.281	1.132	1.094	0.900	0.894	0.677	0.582	0.396	0.360	**0.330**
MAE	1-day	0.149	0.142	0.121	0.116	0.096	0.081	0.063	0.049	0.038	0.029	**0.021**
	7-day	0.236	0.221	0.198	0.187	0.152	0.139	0.099	0.085	0.055	0.042	**0.033**

Table 1: Overall performance comparison.

slots, and predict the crime amount of τ time slots later (we vary $\tau = \{1, 7\}$). Thus, in each region, each type of crime has $T_S = T - \mathcal{T} - \tau + 1$ test samples in total, where $T = 365$ is the total number of time slots. Then the total training sample amount is $T_S \times N \times K \times \mathcal{T}$, and the total test sample amount is $T_S \times N \times K$. Specifically, the amount of training/test sample is 1,754,200/250,600 and 1,724,800/246,400 for 1-day/7-day prediction. The performance of crime prediction is evaluated via the root-mean-square-error (RMSE) and mean absolute error (MAE):

$$RMSE = \sqrt{\frac{1}{NKT_S} \sum_{n=1}^{N} \sum_{k=1}^{K} \sum_{t_s=1}^{T_S} \left(\widehat{\mathbf{Y}}_n^{t_s}(k) - \mathbf{Y}_n^{t_s}(k) \right)^2} \quad (25)$$

$$MAE = \frac{1}{NKT_S} \sum_{n=1}^{N} \sum_{k=1}^{K} \sum_{t_s=1}^{T_S} \left| \widehat{\mathbf{Y}}_n^{t_s}(k) - \mathbf{Y}_n^{t_s}(k) \right| \quad (26)$$

where $\widehat{\mathbf{Y}}_n^{t_s}(k)$ is the predicted crime amount and $\mathbf{Y}_n^{t_s}(k)$ is the observed number. It is worth to note that: (i) We select parameters of the proposed framework such as α, β, γ, ρ and σ by cross-validation; (ii) For a fair comparison, we conduct parameter-tuning for baselines. More details about parameter selection will be discussed in following subsections.

Baseline

We compare our framework with the following baseline methods: **ARIMA** (Chen, Yuan, and Shu 2008), **VAR** (Corman, Joyce, and Lovitch 1987), **LSTM** (Cortez et al. 2018), DeepMove (**DMove**) (Feng et al. 2018), DeepST (**DST**) (Zhang et al. 2016), ST-ResNet (**STRN**) (Zhang, Zheng, and Qi 2017), **CCRF** (Yi et al. 2018), NN-CCRF (**NCCRF**) (Yi et al. 2019), **TCP** (Zhao and Tang 2017b) and DeepCrime (**DCrime**) (Huang et al. 2018).

Performance Comparison for Crime Prediction

To seek answer to the first question, we compare the proposed framework with the state-of-the-art baseline methods. The overall results are shown in Table 1. We have following observations: (1) DeepST and ST-ResNet outperform the previous four methods, which demonstrates that the crimes among different regions are indeed spatially correlated. The first four baselines only consider the temporal dependencies, while overlooking the spatial correlations. (2) CCRF, NCCRF, TCP and DCrime achieve better performance than previous six methods, since these three methods incorporate multiple sources that are related to crime, while previous six methods are solely based on the historical crime records

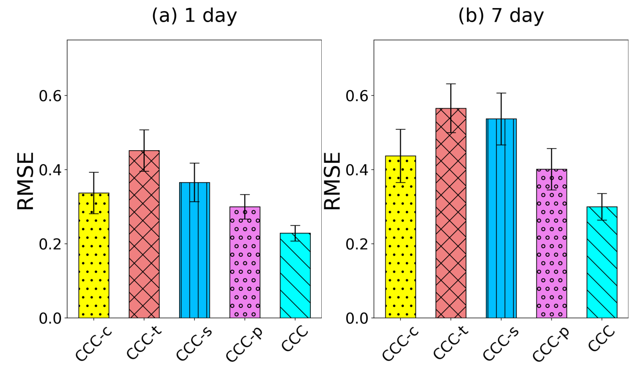

Figure 5: Ablation study in NYC.

(DeepST and ST-ResNet also can incorporate weather data). (3) CCC performs better than CCRF, NCCRF, TCP and DCrime, because CCC jointly captures cross-type correlations among multiple types of crime and spatio-temporal correlations for each type of crime, while CCRF, NCCRF, TCP and DCrime overlook the cross-type correlations. (4) All methods perform relatively better in short-term (1 day) crime prediction, which indicates that prediction of distant future is harder than that of near future. However, the proposed framework performs more robustly in long-term (7 days) prediction than baseline techniques. In summary, the proposed CCC framework can outperform the state-of-the-art baselines with significant margin ($p < 0.05$ in two-sided t-test over the best baseline) for crime prediction.

Ablation Study

In this subsection, we study the contribution of each important component of the proposed framework. We systematically eliminate each component and define following variants of CCC: (1) **CCC**$-c$: This variant evaluates the contribution of cross-type correlation, so we eliminate the impact from cross-type correlation by setting $\alpha = 0$. (2) **CCC**$-t$: This variant evaluates the performance of intra-region temporal correlations, so we set parameters of temporal correlation as 0, i.e., $\beta = 0$. (3) **CCC**$-s$: This variant evaluates the contribution of inter-region spatial correlation, so we eliminate the impact from it by setting all $d(i,j)^{-\gamma}$ as 0. (4) **CCC**$-p$: This variant evaluates the performance of weight $\mathbf{P}$ that captures the common features for all types of crime, so we remove all $\mathbf{P}_n^t$ for $n \in [1, N], k \in [1, K]$.

The results on NYC datasets are shown in Figure 5. From this figure, we can observe: (1) CCC achieves better performance than CCC$-c$ in both 1-day and 7-day prediction, which verifies that different types of crime are intrinsically correlated and introducing cross-type correlations can boost the performance of crime prediction; (2) CCC$-s$ outperforms CCC$-t$ in both 1-day and 7-day prediction, which shows that intra-region temporal correlation contributes more to crime prediction. Their performance becomes close in 7-day prediction. This indicates that temporal correlation becomes weak in long-term prediction; (3) CCC performs better than CCC$-p$. This result supports that introducing weight vectors $\mathbf{P}$ to capture the common features for all crime types is helpful for crime prediction. To sum up, CCC outperforms all variants, which proves that all components are useful in crime prediction and they contain complementary information.

Parametric Sensitivity Analysis

In this section, we evaluate three key parameters of the proposed framework, i.e., (1) α that controls cross-type correlation, (2) β that controls temporal correlation, and (3) γ that controls spatial correlation. To investigate the sensitivity of the proposed framework CCC with respect to these parameters, we study how CCC performs with changing the value of one parameter, while keeping other parameters fixed.

Figure 6 (a) illustrates the parameter sensitivity of α. CCC achieves the best performance when $\alpha = 2$ for 1-day prediction, while $\alpha = 3$ for 7-day prediction. This result indicates that cross-type correlation plays a more important role in long-term prediction. For temporal correlation, Figure 6 (b) shows how the performance changes with β. The performance achieves the peak when $\beta = 1.25$ for 1-day prediction and $\beta = 1$ for 7-day prediction, which suggests that weight vectors $\mathbf{P}_n^t$ and $\mathbf{Q}_n^t(k)$ are closely related to these of just the last few time slots; while in distant future prediction, the temporal correlation becomes weak. In Figure 6, when $\gamma \to 0, d(i,j)^{-\gamma} \to 1$, i.e., all regions are equally related to each other, or $\gamma \to +\infty, d(i,j)^{-\gamma} \to 0$, i.e., all regions are independent of each other. CCC approaches the best performance when $\gamma = 0.5$ for both 1-day and 7-day prediction, which demonstrates the importance of spatial correlation.

Related Work

The first category of works related to our study is traditional statistical, data mining and seismic analysis methods for crime predictions. The first group is statistical methods (Gruenewald et al. 2006; Green, Staerkle, and Sears 2006; Featherstone 2013). For example, researchers show that there is correlation between the characteristics of a population and the rate of violent crimes (Gruenewald et al. 2006). The second group is data mining methods (Mu et al. 2011; Yu et al. 2014; Zhao and Tang 2017a). For instance, a four-order tensor for crime forecasting is presented in (Mu et al. 2011). The tensor encodes the longitude, latitude, time, and other related crimes. The third group is seismic analysis (Lewis et al. 2012; Mohler et al. 2011). For instance, temporal patterns of dynamics of violence are analyzed using a point process model (Lewis et al. 2012).

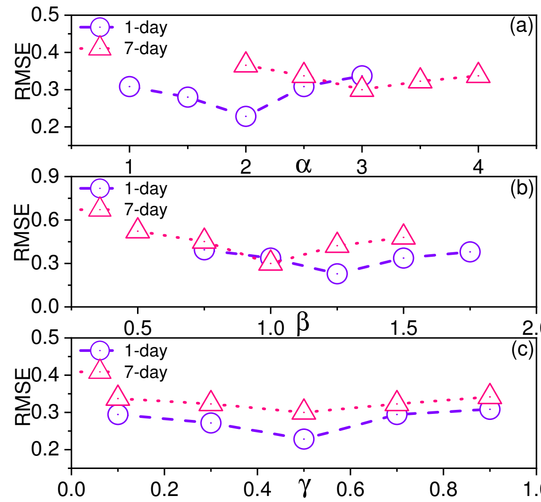

Figure 6: Parameter sensitiveness study in NYC.

The second category related to our study are deep learning methods (Cortez et al. 2018; Feng et al. 2018; Yi et al. 2018, 2019; Huang et al. 2018). For instance, DeepMove integrates RNNs with attention mechanisms to interpret the relations between the current and past values in predicting future crimes (Feng et al. 2018). DeepCrime is a hierarchical recurrent model that is capable of capturing the dynamic crime patterns and their relationships with other ubiquitous data (Huang et al. 2018). However, these methods either do not distinguish different crime types or consider each crime type separately, while ignoring cross-type correlations. Other deep learning methods related to our models are (Huang et al. 2013; Zhao et al. 2015, 2021; Zhu et al. 2016; Liu, Zhao, and Cong 2018; Liu et al. 2021a,b; Wang et al. 2020; Wang, Cao, and Yu 2020; Wang et al. 2019a,b; Wang, Cao, and Yu 2020; Zou et al. 2021).

Conclusion

In this paper, we propose a novel framework CCC, which jointly captures cross-type and spatio-temporal correlations for crime prediction. CCC leverages heterogeneous big urban data, e.g., crime complaint, stop-and-frisk, weather, point of interests (POIs), human mobility and 311 public-service complaint data. We evaluate our framework with extensive experiments based on real-world urban data from New York City. The results show that (1) different types of crime are intrinsically correlated with each other, (2) the proposed framework can accurately predict crime amounts in the near future and (3) cross-type and spatio-temporal correlations can boost crime prediction.

There are several interesting research directions. First, in addition to the cross-type and spatio-temporal correlations we studied in this work, we would like to investigate more crime patterns (e.g. periodicity and tendency). Second, we would like to introduce and develop more advanced techniques for crime analysis. Third, besides crime prediction task, we would like to design more sophisticated models to tackle more practical policing tasks in the real world.

Acknowledgments

This work is supported by the Start-up Grant (No.9610565) for New Faculty of the City University of Hong Kong, CCF-Tencent Open Fund, National Science Foundation (NSF) under grant numbers IIS1714741, CNS1815636, IIS1845081, IIS1907704, DRL2025244, IIS1928278, IIS1955285, IOS2107215, IOS2035472 and Army Research Office (ARO) under grant number W911NF-21-1-0198, a general research fund (GRF) from the Hong Kong Research Grants Council (RGC) (project no. 15200021), the National Natural Science Foundation of China (NSFC project no. 62102335), and an internal research grant from the Hong Kong Polytechnic University (project no. P0036200).

References

Barrett, D. E.; Katsiyannis, A.; and Zhang, D. 2006. Predictors of offense severity, prosecution, incarceration and repeat violations for adolescent male and female offenders. *Journal of Child and Family Studies*, 15(6): 708–718.

Boyd, S.; Parikh, N.; Chu, E.; Peleato, B.; and Eckstein, J. 2011. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends® in Machine Learning*, 3(1): 1–122.

Brantingham, P. L.; and Brantingham, P. J. 1981. Notes on the geometry of crime. *Environmental criminology*.

Chen, P.; Yuan, H.; and Shu, X. 2008. Forecasting crime using the arima model. In *Fuzzy Systems and Knowledge Discovery, 2008. FSKD'08. Fifth International Conference on*, volume 5, 627–630. IEEE.

Corman, H.; Joyce, T.; and Lovitch, N. 1987. Crime, deterrence and the business cycle in New York City: A VAR approach. *The Review of Economics and Statistics*, 695–700.

Cortez, B.; Carrera, B.; Kim, Y.-J.; and Jung, J.-Y. 2018. An architecture for emergency event prediction using LSTM recurrent neural networks. *Expert Systems with Applications*, 97: 315–324.

Couch, C.; and Dennemann, A. 2000. Urban regeneration and sustainable development in Britain: The example of the Liverpool Ropewalks Partnership. *Cities*, 17(2): 137–147.

Featherstone, C. 2013. Identifying vehicle descriptions in microblogging text with the aim of reducing or predicting crime. In *Adaptive Science and Technology (ICAST), 2013 International Conference on*, 1–8. IEEE.

Felson, M.; and Clarke, R. V. 1998. Opportunity makes the thief. *Police research series, paper*, 98.

Feng, J.; Li, Y.; Zhang, C.; Sun, F.; Meng, F.; Guo, A.; and Jin, D. 2018. Deepmove: Predicting human mobility with attentional recurrent networks. In *Proceedings of the 2018 world wide web conference*, 1459–1468.

Fisher, M. 1999. Grand Larceny. *Amaranthus*, 1999(1): 5.

Green, E. G.; Staerkle, C.; and Sears, D. O. 2006. Symbolic racism and Whites' attitudes towards punitive and preventive crime policies. *Law and Human Behavior*, 30(4): 435–454.

Gruenewald, P. J.; Freisthler, B.; Remer, L.; LaScala, E. A.; and Treno, A. 2006. Ecological models of alcohol outlets and violent assaults: crime potentials and geospatial analysis. *Addiction*, 101(5): 666–677.

Guo, H.; Li, X.; He, M.; Zhao, X.; Liu, G.; and Xu, G. 2016. CoSoLoRec: Joint Factor Model with Content, Social, Location for Heterogeneous Point-of-Interest Recommendation. In *International Conference on Knowledge Science, Engineering and Management*, 613–627. Springer.

Henry, L.; and Bryan, B. 2000. Visualising the spatio-temporal patterns of motor vehicle theft in Adelaide, South Australia. In *Conference on crime mapping: Adding value to crime prevention and control*.

Huang, B.; Zhao, X.-Y.; Qi, K.; Tang, M.; et al. 2013. Coloring the complex networks and its application for immunization strategy. *Acta Physica Sinica*, 62(21): 218902–218902.

Huang, C.; Zhang, J.; Zheng, Y.; and Chawla, N. V. 2018. DeepCrime: attentive hierarchical recurrent networks for crime prediction. In *Proceedings of the 27th ACM International Conference on Information and Knowledge Management*, 1423–1432.

Leong, K.; and Sung, A. 2015. A review of spatio-temporal pattern analysis approaches on crime analysis. *International E-Journal of Criminal Sciences*, 9: 1–33.

Lewis, E.; Mohler, G.; Brantingham, P. J.; and Bertozzi, A. L. 2012. Self-exciting point process models of civilian deaths in Iraq. *Security Journal*, 25(3): 244–264.

Liu, H.; Han, J.; Fu, Y.; Zhou, J.; Lu, X.; and Xiong, H. 2021a. Multi-Modal Transportation Recommendation with Unified Route Representation Learning. *Proceedings of the VLDB Endowment*, 14(3): 342–350.

Liu, H.; Wu, Q.; Zhuang, F.; Lu, X.; Dou, D.; and Xiong, H. 2021b. Community-Aware Multi-Task Transportation Demand Prediction. In *Proceedings of the Thirty-Fifth AAAI Conference on Artificial Intelligence*, 320–327.

Liu, Y.; Pham, T.-A. N.; Cong, G.; and Yuan, Q. 2017. An experimental evaluation of point-of-interest recommendation in location-based social networks. *Proceedings of the VLDB Endowment*, 10(10): 1010–1021.

Liu, Y.; Zhao, K.; and Cong, G. 2018. Efficient similar region search with deep metric learning. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 1850–1859.

Liu, Y.; Zhao, K.; Cong, G.; and Bao, Z. 2020. Online anomalous trajectory detection with deep generative sequence modeling. In *2020 IEEE 36th International Conference on Data Engineering (ICDE)*, 949–960. IEEE.

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

Mu, Y.; Ding, W.; Morabito, M.; and Tao, D. 2011. Empirical discriminative tensor analysis for crime forecasting. *Knowledge Science, Engineering and Management*, 293–304.

Revitch, E.; and Schlesinger, L. 1978. Murder: evaluation, classification, and prediction. *Violence: perspectives on murder and aggression. San Francisco, CA: Jossey-Bass.*

Roesch, R.; and Winterdyk, J. 1986. The implementation of a robbery information/prevention program for convenience stores. *Canadian J. Criminology*, 28: 279.

Sampson, R. J.; and Groves, W. B. 1989. Community structure and crime: Testing social-disorganization theory. *American journal of sociology*, 94(4): 774–802.

Thornhill, R.; and Thornhill, N. W. 1983. Human rape: An evolutionary analysis. *Ethology and Sociobiology*, 4(3): 137–173.

Tobler, W. R. 1970. A computer movie simulating urban growth in the Detroit region. *Economic geography*, 46: 234–240.

Wang, J.; Wu, J.; Wang, Z.; Gao, F.; and Xiong, Z. 2019a. Understanding urban dynamics via context-aware tensor factorization with neighboring regularization. *IEEE Transactions on Knowledge and Data Engineering*.

Wang, J.; Wu, N.; Lu, X.; Zhao, X.; and Feng, K. 2019b. Deep Trajectory Recovery with Fine-Grained Calibration using Kalman Filter. *IEEE Transactions on Knowledge and Data Engineering*.

Wang, S.; Cao, J.; and Yu, P. 2020. Deep learning for spatio-temporal data mining: A survey. *IEEE transactions on knowledge and data engineering*.

Wang, S.; Miao, H.; Chen, H.; and Huang, Z. 2020. Multi-task adversarial spatial-temporal networks for crowd flow prediction. In *Proceedings of the 29th ACM international conference on information & knowledge management*, 1555–1564.

Wang, Z.; and Liu, X. 2017. Analysis of Burglary Hot Spots and Near-Repeat Victimization in a Large Chinese City. *ISPRS International Journal of Geo-Information*, 6(5): 148.

Wilson, J. Q.; and Kelling, G. L. 1982. Broken windows. *Atlantic monthly*, 249(3): 29–38.

Xu, T.; Zhu, H.; Xiong, H.; Zhong, H.; and Chen, E. 2019. Exploring the social learning of taxi drivers in latent vehicle-to-vehicle networks. *IEEE Transactions on Mobile Computing*, 19(8): 1804–1817.

Xu, T.; Zhu, H.; Zhao, X.; Liu, Q.; Zhong, H.; Chen, E.; and Xiong, H. 2016. Taxi driving behavior analysis in latent vehicle-to-vehicle networks: A social influence perspective. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 1285–1294. ACM.

Xu, T.; Zhu, H.; Zhong, H.; Liu, G.; Xiong, H.; and Chen, E. 2018. Exploiting the dynamic mutual influence for predicting social event participation. *IEEE Transactions on Knowledge and Data Engineering*, 31(6): 1122–1135.

Yi, F.; Yu, Z.; Zhuang, F.; and Guo, B. 2019. Neural Network based Continuous Conditional Random Field for Fine-grained Crime Prediction. In *IJCAI*, 4157–4163.

Yi, F.; Yu, Z.; Zhuang, F.; Zhang, X.; and Xiong, H. 2018. An integrated model for crime prediction using temporal and spatial factors. In *2018 IEEE International Conference on Data Mining (ICDM)*, 1386–1391. IEEE.

Yu, C.-H.; Ding, W.; Chen, P.; and Morabito, M. 2014. Crime forecasting using spatio-temporal pattern with ensemble learning. In *Pacific-Asia Conference on Knowledge Discovery and Data Mining*, 174–185. Springer.

Zhang, J.; Zheng, Y.; and Qi, D. 2017. Deep Spatio-Temporal Residual Networks for Citywide Crowd Flows Prediction. In *AAAI*, 1655–1661.

Zhang, J.; Zheng, Y.; Qi, D.; Li, R.; and Yi, X. 2016. DNN-based prediction model for spatio-temporal data. In *SIGSPATIAL*, 92. ACM.

Zhang, Y.; and Yeung, D.-Y. 2012. A convex formulation for learning task relationships in multi-task learning. *arXiv preprint arXiv:1203.3536*.

Zhao, X.; and Tang, J. 2017a. Exploring Transfer Learning for Crime Prediction. In *2017 IEEE International Conference on Data Mining Workshops (ICDMW)*, 1158–1159. IEEE.

Zhao, X.; and Tang, J. 2017b. Modeling Temporal-Spatial Correlations for Crime Prediction. In *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*, 497–506. ACM.

Zhao, X.; and Tang, J. 2018. Crime in Urban Areas:: A Data Mining Perspective. *ACM SIGKDD Explorations Newsletter*, 20(1): 1–12.

Zhao, X.; Xia, L.; Zou, L.; Liu, H.; Yin, D.; and Tang, J. 2021. UserSim: User Simulation via Supervised GenerativeAdversarial Network. In *Proceedings of the Web Conference 2021*, 3582–3589.

Zhao, X.; Xu, T.; Fu, Y.; Chen, E.; and Guo, H. 2017. Incorporating Spatio-Temporal Smoothness for Air Quality Inference. In *2017 IEEE International Conference on Data Mining (ICDM)*, 1177–1182. IEEE.

Zhao, X.; Xu, T.; Liu, Q.; and Guo, H. 2016. Exploring the Choice Under Conflict for Social Event Participation. In *International Conference on Database Systems for Advanced Applications*, 396–411. Springer.

Zhao, X.-Y.; Huang, B.; Tang, M.; Zhang, H.-F.; and Chen, D.-B. 2015. Identifying effective multiple spreaders by coloring complex networks. *EPL (Europhysics Letters)*, 108(6): 68005.

Zheng, J.; and Ni, L. M. 2013. Time-dependent trajectory regression on road networks via multi-task learning. In *AAAI*, 1048–1055.

Zhou, J.; Chen, J.; and Ye, J. 2012. Multi-task learning: Theory, algorithms, and applications. *SDM tutorials*.

Zhu, H.; Xiong, H.; Tang, F.; Liu, Q.; Ge, Y.; Chen, E.; and Fu, Y. 2016. Days on market: Measuring liquidity in real estate markets. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 393–402.

Zou, L.; Xia, L.; Hou, L.; Zhao, X.; and Yin, D. 2021. Data-Efficient Reinforcement Learning for Malaria Control. *arXiv preprint arXiv:2105.01620*.