Modeling Temporal-Spatial Correlations for Crime Prediction

Xiangyu Zhao
Data Science and Engineering Lab
Michigan State University
zhaoxi35@msu.edu

Jiliang Tang
Data Science and Engineering Lab
Michigan State University
tangjili@msu.edu

ABSTRACT

Crime prediction plays a crucial role in improving public security and reducing the financial loss of crimes. The vast majority of traditional algorithms performed the prediction by leveraging demographic data, which could fail to capture the dynamics of crimes in urban. In the era of big data, we have witnessed advanced ways to collect and integrate fine-grained urban, mobile, and public service data that contains various crime-related sources and rich temporal-spatial information. Such information provides better understandings about the dynamics of crimes and has potentials to advance crime prediction. In this paper, we exploit temporal-spatial correlations in urban data for crime prediction. In particular, we validate the existence of temporal-spatial correlations in crime and develop a principled approach to model these correlations into the coherent framework TCP for crime prediction. The experimental results on real-world data demonstrate the effectiveness of the proposed framework. Further experiments have been conducted to understand the importance of temporal-spatial correlations in crime prediction.

KEYWORDS

Crime Prediction; Crime Prevention; Temporal-Spatial correlation

1 INTRODUCTION

Crime prediction plays a tremendously impactful role in improving public security and reducing financial loss of crimes. Recent studies have shown that crime prediction is closely related to the sustainable development of urban and the quality of citizen's life [10]. Therefore, there is an increasing and urgent demand for accurate crime prediction. Efforts have been made on understanding crime prediction model based on demographic data, i.e., statistical socioeconomic characteristics of a population, such as education level [12], income level and wealth gap [21, 27], ethnic and religious difference [3].

However, it is still very challenging for researchers and police departments to predict high-accurate crime number with only demographic data in a big city due to the following facts. First, these demographic features are relatively stable over an extended period, which cannot capture the dynamics within a specific community. Second, the vast majority of communities in a city share similar demographic features so that capturing the differences between different communities becomes difficult [31]. Recently, with the tremendous development of new techniques to collect and integrate fine-grained data, a large amount of urban data has been recorded such as public safety data, meteorological data, point of interests (POIs) data, human mobility data and 311 public-service complaint data. Such data has been successfully utilized to advance a variety of urban computing tasks such as air quality prediction [38], noise indicator [39], urban region function discovery [34], social event recommendation [35], driving behavior analysis [33] and real estate ranking [15]. It could also be used to advance crime analysis.

Big urban data provides sources that contain helpful context information about crime. For instance, human mobility offers useful environmental factors such as the function of a region, population density, and residential stability, which can significantly affect criminal activities according to environmental criminology [4]; while meteorological data such as weather information has been proven to be related to crime [8, 28]. Meanwhile, criminal theories such as routine activity theory [7] and rational choice theory [9] suggest that crime distribution is highly determined by time and space. Thus, temporal-spatial factors play a crucial role in crime analysis [23]. Big urban data contains rich and fine-grained information about where and when the data is collected. Such information not only allows us to understand the dynamics of crime like how crime evolves; but also enables us to study spatial factors of crime such as geographical influence. These temporal and spatial understandings provide unprecedented and unique opportunities for us to conduct advanced research on crime analysis with urban data. As a consequence, it has great potential to help us build more accurate crime prediction.

In this paper, we exploit temporal-spatial correlations for crime prediction with urban data. In essence, we aim to investigate the following two challenging questions: (i) what temporal-spatial patterns can be observed about crimes with urban data; and (ii) how to model these patterns mathematically for crime prediction. For temporal-spatial patterns, we focus our investigation on (a) *intra-region temporal correlation* and (b) *inter-region spatial correlation*. Intra-region temporal correlation helps us to understand how crime evolves over time for a region in a city; while inter-region spatial correlation suggests the geographical influence among regions in the city. We propose a novel framework TCP, which captures temporal-spatial correlations for crime prediction. We summarize our major contributions as follows:

- We validate temporal-spatial correlations in crime including intra-region temporal correlation and inter-region spatial correlation;

CIKM'17 , November 6–10, 2017, Singapore, Singapore

ACM ISBN 978-1-4503-4918-5/17/11…$15.00
https://doi.org/10.1145/3132847.3133024

- We propose a novel crime prediction framework TCP, which captures intra-region temporal correlation and inter-region spatial correlation into a coherent model; and
- We conduct experiments on real-world data to verify the effectiveness of the proposed framework and the importance of temporal-spatial correlations in accurate crime prediction.

The rest of this paper is organized as follows. In Section 2, we formally define the problem of crime prediction. We describe the dataset and perform preliminary data analysis in Section 3. In Section 4, we provide approaches to model temporal-spatial correlations and introduce details about the proposed TCP framework with an optimization algorithm. Section 5 presents experimental results with discussions. Section 6 briefly reviews related work. Finally, Section 7 concludes with future work.

2 PROBLEM STATEMENT

In this section, we introduce some mathematical notations and formally define the problem we will study in this work. We use bold letters to denote matrices and vectors, e.g., $\mathbf{W}$ and $\mathbf{w}$; we use non-bold letters to represent scalars, e.g., N and n; and we employ Greek letters as parameters, e.g., α and β.

Let $\mathcal{R} = \{r_1, r_2, \ldots, r_N\}$ denote a set of regions in a city, where N is the number of regions. Suppose that there are totally K time slots (i.e., days, weeks, or months), i.e., $\mathbf{T} = \{t_1, t_2, \ldots, t_K\} \in \mathbb{R}^K$. Let $\mathbf{Y} \in R^{N \times K}$ denote the observed number of crime numbers where $\mathbf{Y}_n^k$ is the crime number observed at region r_n in time slot t_k.

Let $\mathbf{X}^k = [\mathbf{X}_1^k, \mathbf{X}_2^k, \ldots, \mathbf{X}_N^k]^T \in \mathbb{R}^{N \times M}$ denote the feature matrix of all regions in the time slot t_k, where M is the number of features. Note that more details about features will be discussed in the following section.

With the aforementioned notations and definitions, the problem of crime prediction can be formally stated as follows:

Definition 2.1 (Problem Statement). Given the feature matrices $\mathbf{X}^1, \mathbf{X}^2, \ldots, \mathbf{X}^K$ and the historical observed crime matrix $\mathbf{Y}$ of regions in $\mathcal{R}$, we aim to learn a crime predictor that can predict the number of crimes h time slots later (or in time t_{K+h}) for each region in $\mathcal{R}$ by leveraging $\mathbf{X}^1, \mathbf{X}^2, \ldots, \mathbf{X}^K$ and $\mathbf{Y}$.

Note that our goal is to perform crime prediction for a future time slot t_{K+h}. If we construct the feature matrix $\mathbf{X}^k$ using data in t_k, the feature matrix in a future time slot t_{K+h} is not available. Therefore, in our work, $\mathbf{X}^k$ is actually constructed based on data in t_{k-h} instead of t_k. We further assume that there is extra data where we can construct the feature matrix $\mathbf{X}^1$ in t_1. Since for different h values, the major differences for the proposed framework are how to construct the feature matrices and choose the target crime numbers, in the following subsections, we will choose $h = 1$ for illustrations.

3 DATA ANALYSIS

In this section, we will first introduce the dataset for this study, and then perform preliminary analysis about temporal-spatial correlations in crime data. Such understandings lay the groundwork to build a meaningful crime prediction framework.

3.1 Data

We collect crime data from July 1, 2012 to June 30, 2013 in New York City. We divide NYC into $N = 133$ disjointed regions, and each region is a $2km \times 2km$ grid. Note that we also can choose other ways define regions such as zip code. To construct the feature matrices, we also collect multiple sources that can be related to crime. Next we detail these sources.

- **Public Security Sources**: Intuitively, regions with many crime complaints could indicate more crimes in the near future. Therefore, we collect crime complaint data and it contains the complaint frequencies of multiple types of offenses such as assault, arson, harassment and criminal trespass. The NYC police claimed that stop-and-frisk contributes to a decline in the crime rate [32]. Thus, we also collect the Stop-and-Frisk data, which includes a NYC Police Department practice of temporarily detaining, questioning, and at times searching civilians on the street for weapons and other contraband.
- **Meteorological Source**: Informed by criminology [8, 28], meteorology and crime have been found to be correlated. Hence, we collect meteorological data, consisting of weather, temperature, wind strength, precipitation, snowfall, humidity, pressure, visibility, etc. In total, 30 features are collected from NYC meteorological station each day.
- **Point of interests (POIs)**: The density of POIs can characterize the neighborhood functions, which could be helpful for crime prediction[19]. We crawled point of interests from FourSquare. In total, 10 categories of POIs are obtained, i.e., food, shops, residence, nightlife, arts and entertainment, travel, outdoors and recreation, professional, college and education, and event.
- **Human Mobility**: Human mobility provides useful information, such as function of a region, population density, and residential stability, which are related to urban crime. We extract three features from this source. One is check-ins from the POI dataset. The other two are pick-up & drop-off points from the taxi trajectories dataset, which denote the number of people arriving at or departing from the target region.
- **311 Public-Service Complaint Source**: 311 is NYC's governmental non-emergency service number, allowing people in the city to complain about things that are not urgent by making a phone call. It includes air quality, animal, electric, fire, heat, homeless, parking, noise, traffic and water system, etc. 311 shows the citizens' dissatisfaction with government service, thus it is highly related with crime.

In this work, we focus on extracting features from the aforementioned sources. It is possible to also use other sources such as criminal networks, social networks[20, 36] and urban environment. We will leave it as one future work.

3.2 An Analysis on Temporal-Spatial Correlations

Previous studies suggest that temporal-spatial correlations exist in many kinds of urban datasets, such as air quality[38], noise[39], and water quality[25]. Crimes could share similar temporal-spatial correlations in the urban system. In this subsection, we investigate temporal-spatial correlations in crime data. Informed by criminologists, there could be a variety of temporal-spatial patterns [23].

However, in this work, we focus on: (i) intra-region temporal correlation and (ii) inter-region spatial correlation; while leaving the exploring of other temporal-spatial patterns such as weekly periodic patterns as future investigation. Next we will give more details.

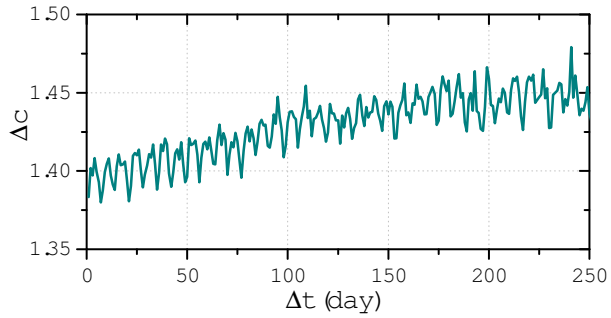

Figure 1: Intra-region Temporal Correlation

Intra-region Temporal Correlation: Within a region, even though the crimes of the region change over time, they should change smoothly. We assume that in time t, the crime number is c_t; while in time $t + \Delta t$, the crime number is $c_{t+\Delta t}$. To study intra-region temporal correlation, we show how $|c_t - c_{t+\Delta t}|$ changes with Δt on average over all regions. The result is shown in Figure 1 where x-axis denotes Δt and y-axis is the average crime differences of $|c_t - c_{t+\Delta t}|$. Note that we choose each time slot as one day in the figure; however, we have similar observations when choosing each time slot as a week and a month. From the figure, we note that the crime differences are highly correlated with Δt. Particularly, two consecutive time slots share similar crime numbers; while with the increase of Δt, the crime difference tends to increase.

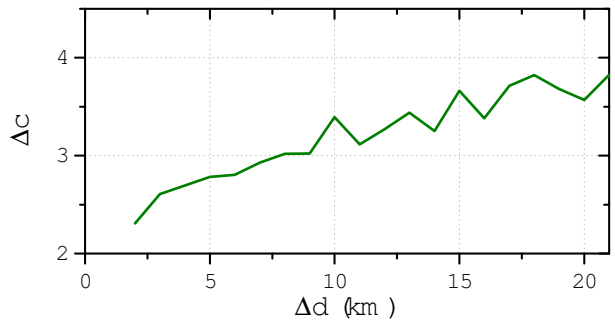

Figure 2: Inter-region Spatial Correlation

Inter-region Spatial Correlation: For inter-region spatial correlation across multiple regions, if two regions are spatially close to each other, it is likely that the two regions have similar crime numbers in the same time slot. Given a pair of regions in a certain time slot, we use Δd to denote the geographical distance between two regions and Δc to indicate the absolute crime difference. In Figure 2, we demonstrate how Δc changes with Δd averaged over all time slots, where x-axis and y-axis are Δd and Δc, respectively. From Figure 2, we observe that (i) if two regions are spatially close to each other, the two regions have similar crime numbers and (ii) with the increase of the geographical distance Δd, the crime difference Δc is likely to increase.

3.3 Discussion

We summarize the observations from our preliminary study as follows:

- For a region, we observe intra-region temporal correlation – (i) for two consecutive time slots, they are likely to share similar crime numbers; and (ii) with the increase of differences between two time slots, the crime difference has the propensity to increase.
- Over all regions, we note inter-region spatial correlation – (i) two geographically close regions have similar crime numbers; and (ii) with the increase of spatial distance between two regions, the crime difference tends to increase.

The above observations provide the groundwork for our proposed framework for crime prediction.

4 THE PROPOSED CRIME PREDICTION FRAMEWORK

In the last section, we validate the existence of temporal-spatial correlations in crime. In this section, we first introduce the basic model, then detail how to model temporal-spatial correlations into a coherent optimization framework and finally discussion how to optimize the framework and how to utilize the framework for crime prediction.

4.1 The Basic Model

Without considering temporal-spatial correlations, we can build an individual and basic model for each region r_n in each time slot t_k. We further assume that there is a vector $\mathbf{W}_n^k \in \mathbb{R}^{M\times 1}$ for the region r_n in the time slot t_k , which can map $\mathbf{X}_n^k$ to Y_n^k as: $\mathbf{X}_n^k\mathbf{W}_n^k$. All $\mathbf{W}_n^k$ can be learned via solving the following optimization problem:

$$\min_{\mathbf{W}_n^k} \sum_{k=1}^{K}\sum_{n=1}^{N} \left(L(\mathbf{X}_n^k\mathbf{W}_n^k, Y_n^k) + \theta\|\mathbf{W}_n^k\|_2^2\right). \quad (1)$$

where L is the loss function and we will choose square loss in this work. However, it is straightforward to extend it to other loss functions such as hinge loss and logistic loss. $\|\mathbf{W}_n^k\|_2^2$ is adopted to avoid overfitting, which is controlled by a non-negative parameter θ. The basic model completely overlooks the temporal-spatial correlations. In the following subsections, we will introduce model components to capture intra-region temporal and inter-region spatial correlations based on the basic model.

4.2 Modeling Intra-region Temporal Correlation

Our preliminary study in the last section suggested that for the same region in a city, with the increase of differences between two time slots, the crime difference tends to increase. This finding paves us a way to model intra-region temporal correlation.

For two time slots t_k and $t_k + \Delta t$ for a region r_n, we propose to minimize the following term to capture temporal correlation based

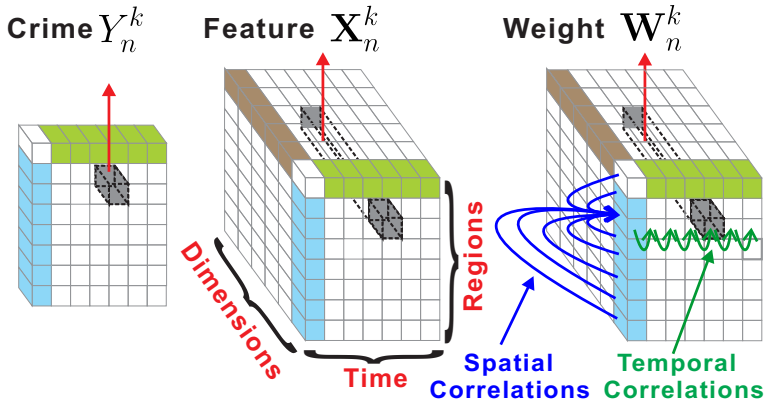

Figure 3: An illustration of the proposed framework with temporal-spatial correlations.

on our observation as:

$$f(\Delta t)\|\mathbf{W}_n^k - \mathbf{W}_n^{k+\Delta t}\|_1 \tag{2}$$

where $f(\cdot)$ is a decay function on Δt. Next we discuss the inner work of Eq. (2). When Δt is smaller, indicating that two time slots are closer, $f(\Delta t)$ is larger, which will push $\mathbf{W}_n^k$ as closer as $\mathbf{W}_n^{k+\Delta t}$. Note that the ℓ_1-norm makes it possible to encourage them exactly the same. Similar analysis can be applied to the case when Δt is larger. Eq. (3) contains the terms for all regions over all time slots.

similar Wn imply features influencing Yn in similar way

$$\lambda \sum_{n=1}^{N} \left(\sum_{k=1}^{K} \sum_{j>k}^{K} f(j-k)\|\mathbf{W}_n^k - \mathbf{W}_n^j\|_1 \right), \tag{3}$$

the parameter λ is introduced to control the contribution from the model component to capture intra-region temporal correlation.

Next we discuss a special definition of $f(\Delta t)$: $f(\Delta t) = 1$ if $\Delta t = 1$ and $f(\Delta t) = 0$, otherwise. Then Eq. (3) can be rewritten as:

$$\lambda \sum_{n=1}^{N} \left(\sum_{k=1}^{K-1} \|\mathbf{W}_n^k - \mathbf{W}_n^{k+1}\|_1 \right), \tag{4}$$

Instead of all pairs of time slots in Eq. (3), Eq. (4) only considers two consecutive time slots. One advantage of Eq. (4) is – it introduces $O(NK)$ terms rather than $O(NK^2)$ in Eq. (3). Since we can choose different time granularities such as hours, days, weeks or months, K can be very large. Therefore, Eq. (4) is more robust to different time granularities. Meanwhile, we empirically find that Eq. (4) and Eq. (3) work very similarly. Thus, we choose Eq. (4) to model intra-region temporal correlation in this work.

Eq. (4) can be rewritten as:

$$\lambda \sum_{n=1}^{N} \sum_{k=1}^{K-1} \|\mathbf{W}_n^k - \mathbf{W}_n^{k+1}\|_1 = \sum_{n=1}^{N} \|\mathbf{W}_n \mathbf{Q}\|_1, \tag{5}$$

where $\mathbf{Q} \in \mathbb{R}^{K\times(K-1)}$ is a sparse matrix. Specifically, $\mathbf{Q}(k,k) = \lambda, \mathbf{Q}(k+1,k) = -\lambda$ for $k = 1, \cdots, K-1$ and all the other terms 0. $\mathbf{W}_n = [\mathbf{W}_n^1, \mathbf{W}_n^2, \ldots, \mathbf{W}_n^K]$. We define $\|\mathbf{X}\|_1$ as $\sum_{i,j} |\mathbf{X}_{ij}|$ in this work.

4.3 Modeling Inter-region Spatial Correlation

Inter-region spatial correlation suggests that with the increase of geographical distance of two regions in a city, the crime difference between the regions in a certain time slot tends to increase. In this subsection, we will develop a model component to capture this observation.

Similar to intra-region temporal correlation, we propose to minimize the following terms to capture inter-region spatial correlation:

$$\sum_{i=1}^{N} \sum_{j=1}^{N} g(d_{ij})\|\mathbf{W}_i^k - \mathbf{W}_j^k\|_1, \tag{6}$$

where d_{ij} is the *Vincenty distance* between region r_i and region r_j. $g(d_{ij})$ is a non-increase function of d_{ij}. When d_{ij} is smaller, meaning r_i and r_j closer, $g(d_{ij})$ should be larger that pushes $\mathbf{W}_i^k$ and $\mathbf{W}_j^k$ closer. Similar analysis can be used when d_{ij} is larger. Above analysis supports that Eq. (6) can model our observations about inter-region spatial correlation. In this work, we find that a power law exponential function of g works well as:

$$g(d_{ij}) = d_{ij}^{-\mathcal{H}}, \tag{7}$$

where $\mathcal{H}$ is a regularization parameter controlling the degree of spatial correlation. This spatial penalty automatically encodes Tobler's first law of geography [30] and imposes a soft constraint that spatially close regions tend to have similar mapping vectors. We can rewrite Eq. (6) as:

$$\sum_{i=1}^{N} \sum_{j=1}^{N} g(d_{ij})\|\mathbf{W}_i^k - \mathbf{W}_j^k\|_1 = \|\mathbf{W}^k \mathbf{P}\|_1, \tag{8}$$

where $\mathbf{P} \in \mathbb{R}^{N\times N^2}$ is a sparse matrix. Specifically, $\forall i = 1, \cdots, N, j = 1, \cdots, N$ and $i \neq j$, we have $\mathbf{P}(i, (i-1)\cdot N + j) = g(d_{ij}), \mathbf{P}(j, (i-1)\cdot N + j) = -g(d_{ij})$ and all the other terms 0. We use $\mathbf{W}^k = [\mathbf{W}_1^k, \mathbf{W}_2^k, \ldots, \mathbf{W}_N^k] \in \mathbb{R}^{M\times N}$ to denote the projection matrix of all regions in the time slot t_k.

4.4 An Optimization Method

With model components to capture temporal and spatial correlations, the proposed framework TCP is to solve the following

optimization formulation:

$$\min_{\mathbf{W}} L = \sum_{k=1}^{K}\left(\sum_{n=1}^{N}(\mathbf{X}_n^k\mathbf{W}_n^k - Y_n^k)^2 + \frac{1}{2}\|\mathbf{W}^k\mathbf{P}\|_1\right) + \sum_{i=n}^{N}\|\mathbf{W}_n\mathbf{Q}\|_1. \tag{9}$$

An illustration of the proposed framework is demonstrated in Figure 3. The first term is the basic model, the second term captures spatial correlations and the third term models temporal correlations.

In this work, we utilize ADMM [2] to optimize the objective function in Eq. (9). We first introduce *auxiliary variable matrices* $\mathbf{E}^k = \mathbf{W}^k\mathbf{P}$ and $\mathbf{F}_n = \mathbf{W}_n\mathbf{Q}$. Then the optimization formulation becomes:

$$\begin{aligned}\min_{\mathbf{W}} L = &\sum_{k=1}^{K}\left(\sum_{n=1}^{N}(\mathbf{X}_n^k\mathbf{W}_n^k - Y_n^k)^2 + \frac{1}{2}\|\mathbf{E}^k\|_1\right) + \sum_{n=1}^{N}\|\mathbf{F}_n\|_1 \\ s.t.\quad &\mathbf{E}^k = \mathbf{W}^k\mathbf{P}, \quad \forall k = 1,\cdots,K \\ &\mathbf{F}_n = \mathbf{W}_n\mathbf{Q} \quad \forall n = 1,\cdots,N\end{aligned} \tag{10}$$

Then the scaled form of ADMM objective function of Eq (10) can be written as:

$$\begin{aligned}\min L_\rho(\mathbf{W},\mathbf{E},\mathbf{F},\mathbf{U},\mathbf{V}) = &\sum_{k=1}^{K}\sum_{n=1}^{N}(\mathbf{X}_n^k\mathbf{W}_n^k - Y_n^k)^2 \\ &+ \frac{1}{2}\sum_{k=1}^{K}\|\mathbf{E}^k\|_1 + \frac{\rho}{2}\sum_{k=1}^{K}\|\mathbf{W}^k\mathbf{P} - \mathbf{E}^k + \mathbf{U}^k\|_F^2 \\ &+ \sum_{n=1}^{N}\|\mathbf{F}_n\|_1 + \frac{\rho}{2}\sum_{n=1}^{N}\|\mathbf{W}_n\mathbf{Q} - \mathbf{F}_n + \mathbf{V}_n\|_F^2,\end{aligned} \tag{11}$$

where $\mathbf{U}^k \in \mathbb{R}^{M\times N^2}$ and $\mathbf{V}_n \in \mathbb{R}^{M\times(K-1)}$ are *scaled dual variable matrices*. ρ is a parameter to control the penalty for the violation of equality constraints $\mathbf{E}^k = \mathbf{W}^k\mathbf{P}$ and $\mathbf{F}_n = \mathbf{W}_n\mathbf{Q}$. $\|\cdot\|_F$ is the *Frobenius*-norm of a matrix. Following the standard ADMM process, the $t+1^{th}$ iteration of ADMM optimization of Eq (11) consists of the following five procedures:

$$\mathbf{W}_n^k(t+1) = \arg\min_{W_n^k} L_\rho(\mathbf{W},\mathbf{E},\mathbf{F},\mathbf{U},\mathbf{V}), \tag{12}$$

$$\mathbf{E}^k(t+1) = S_{\frac{1}{2}/\rho}\Big(\mathbf{W}^k(t+1)\mathbf{P} + \mathbf{U}^k(t)\Big), \tag{13}$$

$$\mathbf{F}_n(t+1) = S_{1/\rho}\Big(\mathbf{W}_n(t+1)\mathbf{Q} + \mathbf{V}_n(t)\Big), \tag{14}$$

$$\mathbf{U}^k(t+1) = \mathbf{U}^k(t) + \mathbf{W}^k(t+1)\mathbf{P} - \mathbf{E}^k(t+1), \tag{15}$$

$$\mathbf{V}_n(t+1) = \mathbf{V}_n(t) + \mathbf{W}_n(t+1)\mathbf{Q} - \mathbf{F}_n(t+1), \tag{16}$$

where the *soft thresholding operator* S is defined as

$$S_\alpha(x) = \begin{cases} x-\alpha & if\ x > \alpha \\ 0 & if\ \|x\| \le \alpha \\ x+\alpha & if\ x < -\alpha \end{cases} \tag{17}$$

In each iteration of ADMM, we leverage Stochastic Gradient Descent (SGD) to update weight matrices. Specifically, in each iteration of ADMM, we randomly select integer $k \in [1,K]$ and integer $n \in [1,N]$, and we approach n^{th} column of weight matrix $\mathbf{W}^k$ by deriving the gradient of $L_\rho(\mathbf{W},\mathbf{E},\mathbf{F},\mathbf{U},\mathbf{V})$ with respect to $\mathbf{W}_n^k$ as:

$$\begin{aligned}\frac{\partial L_\rho(\mathbf{W},\mathbf{E},\mathbf{F},\mathbf{U},\mathbf{V})}{\partial \mathbf{W}_n^k} = &2(\mathbf{X}_n^k\mathbf{W}_n^k(t) - Y_n^k)\cdot(\mathbf{X}_n^k)^T \\ &+ \rho(\mathbf{W}^k(t)\mathbf{P} - \mathbf{E}^k(t) + \mathbf{U}^k(t))\cdot(\mathbf{P}_n)^T \\ &+ \rho(\mathbf{W}_n(t)\mathbf{Q} - \mathbf{F}_n(t) + \mathbf{V}_n(t))\cdot(\mathbf{Q}^k)^T\end{aligned} \tag{18}$$

where $\mathbf{P}_n$ is the n^{th} row of $\mathbf{P}$ and $\mathbf{Q}^k$ is the k^{th} row of $\mathbf{Q}$. Thus in each iteration of ADMM, all $\mathbf{W}_{n'}^{k'}(k' \neq k \text{ or } n' \neq n)$ are fixed. We then use Gradient Descent method with the gradient calculated in Eq (18) to update the current $\mathbf{W}_n^k$ until converge. And then we proceed to update $\mathbf{E}^k, \mathbf{F}_n, \mathbf{U}^k$ and $\mathbf{V}_n$ with the selected k and n.

The detailed ADMM optimization algorithm is shown in Algorithm 1. In Algorithm 1, γ is the learning rate. Next, we briefly discuss the algorithm. In line 1, we initialize weight matrices $\mathbf{W}^k$, auxiliary variable matrices $\mathbf{E}^k$ and $\mathbf{F}_n$, and scaled dual variables matrices $\mathbf{U}^k$ and $\mathbf{V}_n$ randomly, for $k = 1,\cdots,K$ and $n = 1,\cdots,N$ respectively. In each iteration of ADMM, we select integer $k \in [1,K]$ and integer $n \in [1,N]$ randomly in line 3, and first update $\mathbf{W}_n^k$ leveraging gradient descent from line 4 to line 7, and then update $\mathbf{E}^k, \mathbf{F}_n, \mathbf{U}^k$ and $\mathbf{V}_n$ using aforementioned update rules from line 8 to line 11. After ADMM is convergent, Algorithm 1 will output the well trained weight matrices $\mathbf{W}^k$, for $k = 1,\cdots,K$ respectively.

Algorithm 1 The ADMM Optimization Procedures of the Proposed Framework.

Input: The feature matrices $\mathbf{X}^k, \forall k = 1,\cdots,K$, the target matrix $\mathbf{Y}$, the sparse matrices $\mathbf{P}$ and $\mathbf{Q}$, the parameter ρ of ADMM
Output: The weight matrices $\mathbf{W}^k, \forall k = 1,\cdots,K$

1: Initialize $\mathbf{W}^k, \mathbf{E}^k, \mathbf{U}^k$ randomly $\forall k = 1,\cdots,K$ and initialize $\mathbf{F}_n, \mathbf{V}_n$ randomly $\forall n = 1,\cdots,N$
2: **while** Not convergent **do**
3: Select integer $k \in [1,K]$ and integer $n \in [1,N]$ randomly
4: **while** Not convergent **do**
5: Calculate $\frac{\partial L_\rho(\mathbf{W},\mathbf{E},\mathbf{F},\mathbf{U},\mathbf{V})}{\partial \mathbf{W}_n^k}$ according Eq. (18)
6: Update $\mathbf{W}_n^k \leftarrow \mathbf{W}_n^k - \gamma\frac{\partial L_\rho(\mathbf{W},\mathbf{E},\mathbf{F},\mathbf{U},\mathbf{V})}{\partial \mathbf{W}_n^k}$
7: **end while**
8: Update $\mathbf{E}^k$ according to Eq. (13)
9: Update $\mathbf{F}_n$ according to Eq. (14)
10: Update $\mathbf{U}^k$ according to Eq. (15)
11: Update $\mathbf{V}_n$ according to Eq. (16)
12: **end while**

Now we analyze the time complexity of the Algorithm 1. In each iteration of ADMM, the most time consuming part is the SGD procedure. In each iteration of SGD, we calculate $\frac{\partial L_\rho(\mathbf{W},\mathbf{E},\mathbf{F},\mathbf{U},\mathbf{V})}{\partial \mathbf{W}_n^k}$ according to Eq. (18). First we consider the time complexity of $2(\mathbf{X}_n^k\mathbf{W}_n^k(t) - Y_n^k)\cdot(\mathbf{X}_n^k)^T$, in which $\mathbf{X}_n^k\mathbf{W}_n^k(t)$ can be computed in $O(M^2)$, then subtracting Y_n^k and multiplying $(\mathbf{X}_n^k)^T$ can be computed in $O(M)$, so the time complexity of $2(\mathbf{X}_n^k\mathbf{W}_n^k(t) - Y_n^k)\cdot(\mathbf{X}_n^k)^T$ is $O(M^2+M)$. For $\rho(\mathbf{W}^k(t)\mathbf{P} - \mathbf{E}^k(t) + \mathbf{U}^k(t))\cdot(\mathbf{P}_n)^T$, since the matrix representation of $\mathbf{P}$ is very sparse, i.e., each column of $\mathbf{P}$ has at most

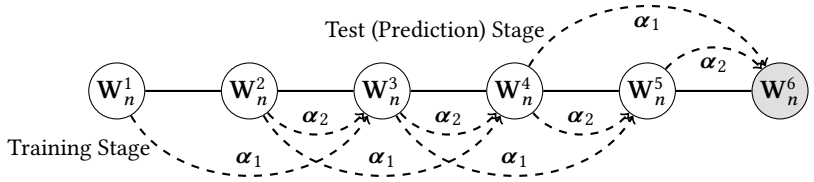

Figure 4: An Example of Learning Parameters for Crime Prediction.

two non-zero elements, thus $\mathbf{W}^k(t)\mathbf{P} - \mathbf{E}^k(t) + \mathbf{U}^k(t)$ can be computed in $O(M * N)$. Since $(\mathbf{P}_n)^T$ has $2 * (N-1)$ non-zero elements, then multiplying $(\mathbf{P}_n)^T$ can be computed in $O(M * N)$, so time complexity of $\rho(\mathbf{W}^k(t)\mathbf{P} - \mathbf{E}^k(t) + \mathbf{U}^k(t)) \cdot (\mathbf{P}_n)^T$ is $O(M * N)$. Similarly, for $\rho(\mathbf{W}_n(t)\mathbf{Q} - \mathbf{F}_n(t) + \mathbf{V}_n(t)) \cdot (\mathbf{Q}^k)^T$, since $\mathbf{Q}$ is very sparse, i.e., each row or column of $\mathbf{Q}$ has at most two non-zero elements, then time complexity of it is $O(N * K)$. Therefore, considering that there are N regions and K time slots, the time complexity of each ADMM iteration is $\#iter_{SGD} * O(NK(M^2 + M * N + N * K))$ where $\#iter_{SGD}$ is the number of iterations for the SGD procedure. Our optimization method is based on ADMM, hence, it is straightforward to be parallelized for large-scale datasets.

4.5 Crime Prediction

The proposed framework TCP can make use of temporal-spatial correlations to learn a $\mathbf{W}_n^k$ for each region r_n in the time slot t_k. In this subsection, we discuss how to utilize all $\{\mathbf{W}_n^k\}$ for crime prediction for a future time slot t_{K+1}.

As discussed in Section 2, the feature matrix $\mathbf{X}^k$ is constructed based on data in the time slot t_{k-1} instead of t_k. Therefore, we can get the feature matrix $\mathbf{X}^{K+1}$ using data from the time slot t_K. If we can get the mapping vector of $\mathbf{W}_n^{K+1}$ for the region r_n in the time slot t_{K+1}, then we can predict the crime number of r_n in t_{K+1} as $\mathbf{X}_n^{K+1}\mathbf{W}_n^{K+1}$. Based on the above process, we boil down the problem as how to utilize all $\{\mathbf{W}_n^k\}_{k=1}^K$ to estimate $\mathbf{W}_n^{K+1}$.

According to intra-region temporal correlation, the mapping vector $\mathbf{W}_n^k$ should be related to these in its previous time slots. We further assume that $\mathbf{W}_n^k$ can be estimated by its g previous time slots as:

$$\mathbf{W}_n^k = \alpha_1 \mathbf{W}_n^{k-g} + \alpha_2 \mathbf{W}_n^{k-g+1} + \cdots + \alpha_g \mathbf{W}_n^{k-1}, \quad (19)$$

the coefficients $\boldsymbol{\alpha} = \{\alpha_1, \alpha_2, \ldots, \alpha_g\}$ are introduced to control the contributions from $\{\mathbf{W}_n^{k-g}, \mathbf{W}_n^{k-g+1}, \mathbf{W}_n^{k-1}\}$, separately. We can empirically set values of $\{\alpha_1, \alpha_2, \cdots, \alpha_g\}$. But it is not practical especially when g is large. Therefore, it would be better to develop an algorithm that can automatically estimate these parameters $\boldsymbol{\alpha}$ from the training data. Actually $\boldsymbol{\alpha}$ can be estimated via solving the following optimization problem:

$$\min_{\boldsymbol{\alpha}} \sum_{k=g+1}^{K} (\mathbf{X}_n^k(\alpha_1 \mathbf{W}_n^{k-g} + \alpha_2 \mathbf{W}_n^{k-g+1} + \cdots + \alpha_g \mathbf{W}_n^{k-1}) - Y_n^k)^2 \quad (20)$$

Figure 4 depicts an example to demonstrate how we learn $\boldsymbol{\alpha}$ in a region r_n. Support that we use $K = 5$ time slots as training data and we want to predict crime number in the time slot 6. Via Algorithm 1, we can learn the mapping vectors $\{\mathbf{W}_n^1, \mathbf{W}_n^2, \mathbf{W}_n^3, \mathbf{W}_n^4, \mathbf{W}_n^5\}$. In this example, we use previous $g = 2$ time slots to predict $\mathbf{W}_n^6$ as $\mathbf{W}_n^6 = \alpha_1 \mathbf{W}_n^4 + \alpha_2 \mathbf{W}_n^5$. To learn $\{\alpha_1, \alpha_2\}$, we can construct $K - g = 3$ samples, i.e., $\mathbf{W}_n^3 = \alpha_1 \mathbf{W}_n^1 + \alpha_2 \mathbf{W}_n^2$, $\mathbf{W}_n^4 = \alpha_1 \mathbf{W}_n^2 + \alpha_2 \mathbf{W}_n^3$ and $\mathbf{W}_n^5 = \alpha_1 \mathbf{W}_n^3 + \alpha_2 \mathbf{W}_n^4$. Via solving Eq. (20), we can estimate $\{\alpha_1, \alpha_2\}$. After obtaining $\{\alpha_1, \alpha_2\}$, the crime number in the time slot 6 can be predicted as: $\mathbf{X}_n^6(\alpha_1 \mathbf{W}_n^4 + \alpha_2 \mathbf{W}_n^5)$.

For different regions, they may have different temporal patterns. Therefore, after learning the mapping vectors via Algorithm 1 together, we learn the parameters $\boldsymbol{\alpha}$ to estimate $\mathbf{W}_n^{K+1}$ for each region, respectively.

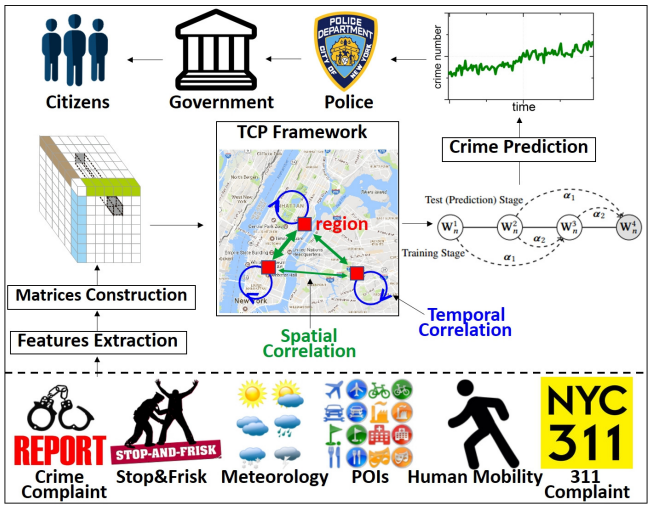

Figure 5: The Overall of the crime prediction system.

4.6 An Overview of the Crime Prediction System

As shown in Figure 5, our whole crime prediction system comprises of three major components: (i) feature extraction, (ii) the proposed framework TCP, and (iii) crime number prediction, as follows:

- **Step 1: Feature Extraction.** We first extract features from multiple sources, such as crime complaint dataset, stop-and-frisk dataset, meteorology, Point of Interests (POIs), human mobility and 311 complaint dataset. Then we combine these features into feature matrices $\{\mathbf{X}^k\}_{k=1}^K$.
- **Step 2: the proposed framework TCP.** Based on feature matrices and the historical crime numbers, the proposed framework

TCP learns the model parameters $\mathbf{W}_n^k$ for each region r_n in each time slot t_k.

- **Step 3: Crime Number Prediction.** Based on the well trained model $\mathbf{W}_n^k$ in Step 2, our system learns parameters to estimate $\mathbf{W}_n^{K+1}$ via solving Eq. (20) and performs crime prediction based on $\mathbf{W}_n^{K+1}$ for each region for t_{K+1}.

5 EXPERIMENTS

In this section, we conduct extensive experiments with real-world urban data to evaluate the effectiveness of the proposed framework. We mainly aim to answer the following two questions: (a) how the proposed framework performs compared to representative baselines; and (b) how the temporal and spatial patterns contribute to the performance. We first introduce experimental settings. Then we seek answers to the above two questions. Next we study the impact of important parameters on the performance of the proposed framework. Finally, to further demonstrate the potentials of temporal-spatial information, we investigate more temporal-spatial patterns in addition to intra-region temporal and inter-region spatial correlations.

5.1 Experimental Settings

We evaluate our method on the same dataset introduced in Section 3, which is collected from July 1, 2012 to June 30, 2013 (365 days) in New York City with $N = 133$ disjointed regions. In each region, we use data of previous K time slots to train the framework, and predict the crime number of h time slots later. Therefore, for each region, there are in total $K_S = T - K - h + 1$ testing samples where T is the total number of time slots.

The crime prediction performance is evaluated in terms of the average root-mean-square-error (RMSE) of all N regions as:

$$aRMSE = \frac{1}{N}\sum_{n=1}^{N}\sqrt{\frac{1}{K_S}\sum_{k=1}^{K_S}\left(\hat{\mathbf{Y}}_n^k - \mathbf{Y}_n^k\right)^2} \tag{21}$$

where $\hat{\mathbf{Y}}_n^k$ is the predicted crime number; while $\mathbf{Y}_n^k$ is the observed crime number. In this evaluation, we choose the time slots as days and $K = 5$ because crime numbers are related to recent past. We vary h as $\{1, 7\}$. For the parameters of the proposed framework such as λ, $\mathcal{H}$ and ρ, we select them via cross-validation. Correspondingly, we also do parameter-tuning for baselines for a fair comparison. We will discuss more details about parameter selection for the proposed framework in the following subsections.

5.2 Performance Comparison for Crime Prediction

To answer the first question, we compare the proposed framework with the following representative baseline methods:

- **CSI** [11]: Cubic Spline Interpolation trains piecewise third-order polynomials which pass through crime points of recent K days, and then predicts the crime number in the near future by the trained polynomials.
- **ARMA**: Auto-Regression-Moving-Average is well-known for predicting time series data. ARMA predicts the crime number of a region solely based on the historical crime records of the region, considering the recent K days for a moving average.
- **LASSO** [29]: Lasso tries to minimize the objective function $\frac{1}{2}\|\mathbf{Y}^k - \mathbf{X}^k\mathbf{W}^k\|_2^2 + \gamma\|\mathbf{W}^k\|_1$ and encodes the sparsity over all weights in $\mathbf{W}^k$.
- **LR** [26]: Linear Regression is applied for each region individually, which totally overlooks the temporal-spatial correlations.
- **stMTL** [37]: Spatio-Temporal Multi-Task Learning enhances static spatial smoothness regression framework by learning the temporal dynamics of features through an non-parametric term, i.e., $\sum_{m=1}^{M}(\sum_{k=2}^{K}|W_m^k - W_m^{k-1}|)^2$, where M is the dimension of $\mathbf{W}^k$, and W_m^k is the m^{th} element of $\mathbf{W}^k$.

Table 1: Overall performance Comparison in terms of aRMSE.

	1 day	7 days
CSI	13.223	33.562
ARMA	6.3135	12.2572
LASSO	2.8210	3.3956
LR	2.5498	2.8985
stMTL	2.2356	2.5365
TCP	**1.7205**	**1.7791**

Note that parameters in baselines and our model are determined via cross-validation. More details about parameter selection about our model will be discussed in the following subsections. The results are shown in Table 1. We make following observations:

- All the techniques perform relatively better for short-term prediction (1 day vs. 7 days), which suggests that prediction for the near time is easier than that of distant future. However, our TCP framework is much more robust in distant future prediction than baseline methods.
- ARMA performs better than CSI because ARMA introduces a moving average, which can capture the crime evolving trends once it has been established.
- CSI and ARMA achieve the worse performance than LASSO and LR, since these two methods are only base on the historic crime dataset, while LASSO and LR can incorporate multi-source heterogeneous information.
- stMTL and TCP outperform the other four methods, which demonstrates that the crimes among different regions are indeed correlated.
- TCP performs better than stMTL. stMTL only captures that features cross regions in the same time slot share the same weights; while TCP captures both temporal-spatial correlations. More details about the impact of temporal and spatial correlations in TCP will be discussed in the following subsection.

Via the performance comparison, we can draw an answer to the first question – by modeling temporal-spatial correlations, the proposed framework TCP outperforms representative baselines for crime prediction.

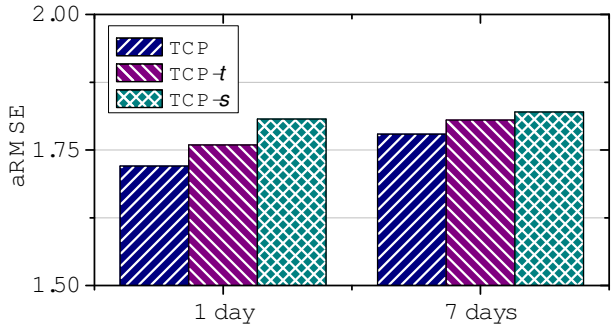

Figure 6: Impact of Temporal and Spatial Correlations.

5.3 Impact of Temporal and Spatial Correlations

To study the impact of temporal and spatial correlations on the proposed framework TCP, we systematically eliminate the corresponding model components by defining following variants of TCP:

- TCP−t: This variant is to evaluate the performance of temporal correlations, so we set parameters of spatial correlation as 0, i.e., $\forall i, j, g(d_{ij}) = 0$.
- TCP−s: In this variant, we evaluate the performance of spatial correlation, so we eliminate the impact from temporal correlation by setting $\lambda = 0$.

The experimental results are demonstrated in Figure 6. From this figure, it can be observed:

- For 1-day prediction, TCP−t outperforms TCP−s significantly, which indicates that $\mathbf{W}_n^{k+1}$ of near future is highly depended on $\mathbf{W}_n^k$ within a region.
- For 7-day prediction, the performance of TCP−t and TCP−s becomes close. This result supports that the temporal correlation becomes weak with the increase of time difference Δt.
- TCP outperforms both TCP−t and TCP−s. This observation further supports the importance of temporal-spatial correlations in crime prediction.

The above observations suggest that (a) both temporal and spatial patterns are useful; and (b) they contain complementary information.

5.4 Parametric Sensitivity Analysis

Our method has two key parameters, i.e., λ that controls temporal correlation and $\mathcal{H}$ that controls spatial correlation. To study the impact of these parameters, we investigate how the proposed framework TCP works with the changes of one parameter, while fixing other parameters.

Figure 7 (a) shows the parameter sensitivity of λ in crime number prediction task. The performance for 1-day prediction achieves the peak when $\lambda = 2$. In other words, $\mathbf{W}_n^{k+1}$ of near future is highly depended on $\mathbf{W}_n^k$ within a region. For 7-day prediction, performance achieves the peak when $\lambda = 1$, which indicates that the temporal correlation becomes weak with the Δt increasing.

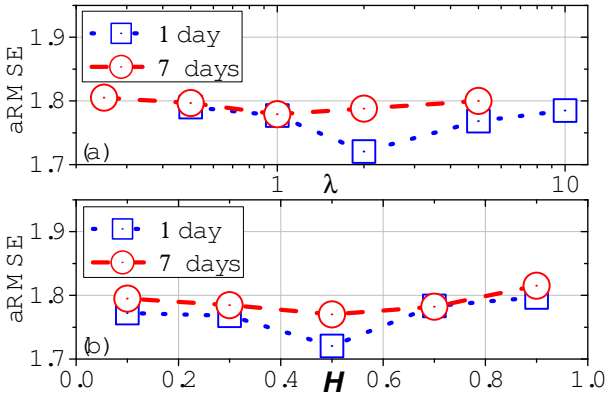

Figure 7: Parameter Sensitiveness. (a) λ for temporal correlation. (b) $\mathcal{H}$ for spatial correlation.

For spatial correlation, Figure 7 (b) shows how the performance changes with $\mathcal{H}$. When $\mathcal{H} \to 0, g(d_{ij}) \to 1$, i.e., all regions are all highly related to each other, or $\mathcal{H} \to +\infty, g(d_{ij}) \to 0$, i.e., all regions are independent to each other. TCP steadily achieves the best performance when $\mathcal{H} = 0.5$, which suggests the importance of spatial correlation in crime prediction.

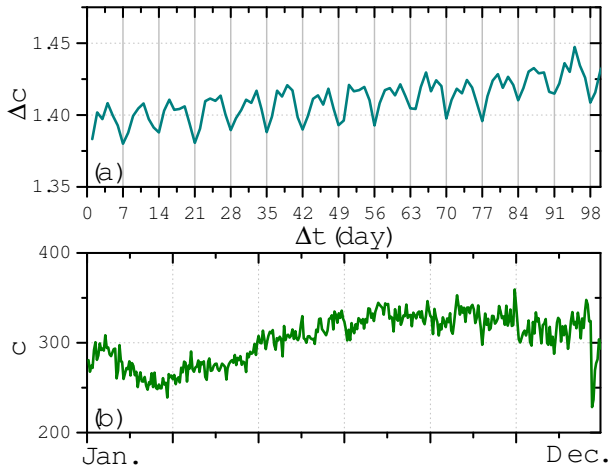

Figure 8: Temporal Pattern of Urban Crimes. (a) Weekly Periodicity. (b) Days of a Year.

5.5 Further Probing on Temporal-Spatial Patterns

Our experimental results show the promising of temporal-spatial correlations in crime analysis. In this subsection, we further demonstrate the potentials of temporal-spatial correlations in crime analysis by showing more interesting temporal-spatial patterns.

First, we only show the first 100 points of Figure 1 in Figure 8 (a). We note that even though the crime difference Δc tends to increase with the increase of Δt, we can see the weekly periodicity that Δc will decrease when $\Delta t = 7, 14, 21, \cdots$. This means that the crimes in

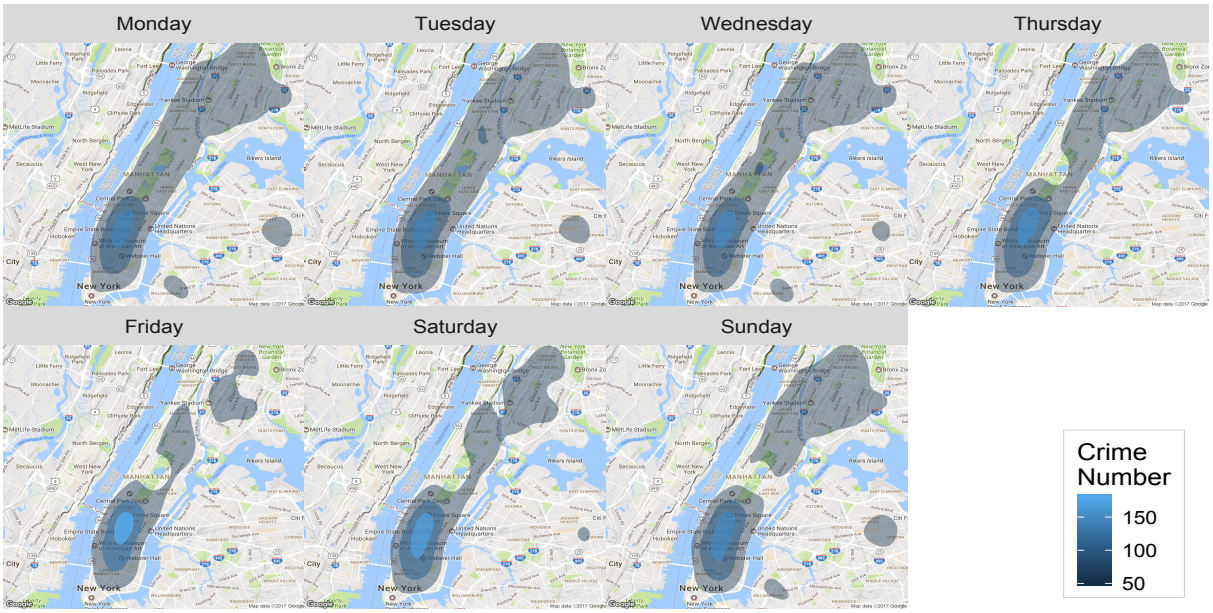

Figure 9: Crime spatial distribution map for NYC with respect to days of a week.

Monday may be similar on consecutive weeks. Furthermore, Figure 8 (b) shows the average daily crime number from 2006 to 2015. We note that the daily crime number gradually increases from March to September, while gradually decreases from October to February. An interesting observation is that crimes increase before Christmas, but decease dramatically during Christmas and New Year. These suggest that holidays could matter in crime analysis.

Also, motivated by the weekly periodicity of urban crimes, we also want to examine how the spatial distribution of urban crimes varies with respect to days of a week. Figure 9 shows the crime density of NYC from July 1, 2012 to June 30, 2013. From Figure 9, we have following observations: (a) typically, Monday to Thursday share similar spatial distributions of urban crimes, while Friday to Sunday have similar spatial distributions of urban crimes; and (b) Friday has the most uneven distributions, and it is also the most unsafe day in a week.

The aforementioned temporal-spatial patterns can be further captured to improve the crime prediction performance. We will leave it as one future investigation direction.

6 RELATED WORK

In this section, we briefly introduce the work related to our study. In general, the related work can be mainly grouped into the following categories.

The first category is about environmental criminology. In this paper, we analyze the crime prediction problem by incorporating temporal-spatial correlations. Criminal theories such as routine activity theory [7] and rational choice theory [9] suggest that crime distribution is highly determined by time and space. Specifically, according to routine activity theory [7], the union of three elements in time and space are required for a crime to occur: a likely offender, a suitable target and the absence of a capable guardian against crime. Crimes could be prevented or reduced by interacting with any aspect of the triangle. Rational choice perspective theory [6] focuses upon the offender's decision making processes with hypothesis that offending is purposive behavior which helps the offender in some way. Furthermore, crime pattern theory [14] is helpful in establishing how people interact with their spatial environment. Awareness theory [5] has suggested that crime has four dimensions: victim, offender, geo-temporal and legal and concentrating on the spatial element of crime is significant to understand the behavior of offenders.

The second category related to this paper is current crime prediction techniques. Typically, current techniques can be classified into three group. The first group of techniques are based on statistical methods. For example, researchers show that there is correlation between the characteristics of a population and the rate of violent crimes [18]. The author in [13] is able to discover a correlation between reported crime census statistics from the South African Police Service and crime events discussed in tweets. While authors in [17] conclude that there is a positive effect of symbolic racism on both preventive and punitive penalties. The second group of techniques are data mining methods. For instance, the author in [16] uses Latent Dirichlet Allocation for learning topics and related terms from tweets and eschews deep semantic analysis in favor of shallower analysis via topic modeling. Another researcher built a crime policing self-organizing map to extract information such as crime type and location from reports to provide for a more effective crime analysis and employs an unsupervised Sequential Minimization Optimization for clustering [1]. Finally, the third group of methods predict crime hotspot from geographical information system perspective. For example, some researchers studied the trend of using web-based crime mapping from the 100 highest GDP cities of the world [22]. The authors conclude that the main factors that

drive numerous crime mapping are e-governance and community policing. Another work proposes a novel Bayesian based prediction model to predict the accurate location of the next crime scene in a serial crime [24].

7 CONCLUSION

In this paper, we propose a novel framework TCP, which captures temporal-spatial correlations including intra-region temporal correlation and the inter-region spatial correlation for crime prediction. TCP utilizes heterogeneous urban sources, e.g., public security data, meteorological data, point of interests (POIs), human mobility data and 311 complaint data. We evaluate our approach with extensive experiments based on real-world urban data about New York City. The results show that (1) our framework can accurately predict crime numbers in the future; (2) temporal-spatial correlations can help crime prediction and (3) more temporal-spatial patterns could be used to advance crime prediction.

There are several interesting research directions. First, in addition to urban sources we used in this work, we would like to investigate more sources. Second, we would like to validate with more temporal-spatial patterns and investigate how to model them mathematically for crime prediction. Finally, the formulation proposed in the work is quite general to capture temporal-spatial correlations; hence we would like to investigate more applications of the proposed formulation.

ACKNOWLEDGEMENTS

This material is based upon work supported by, or in part by, the National Science Foundation (NSF) under grant number IIS-1714741 and IIS-1715940. We also would like to thank Dawei Yin, Xia Hu and Suhang Wang for their helpful discussions.

REFERENCES

[1] Meshrif Alruily. 2012. Using text mining to identify crime patterns from arabic crime news report corpus. (2012).
[2] Stephen Boyd, Neal Parikh, Eric Chu, Borja Peleato, and Jonathan Eckstein. 2011. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends® in Machine Learning* 3, 1 (2011), 1–122.
[3] John Braithwaite. 1989. *Crime, shame and reintegration.* Cambridge University Press.
[4] Paul J Brantingham and Patricia L Brantingham. 1981. *Environmental criminology.* Sage Publications Beverly Hills, CA.
[5] Patricia L Brantingham and Paul J Brantingham. 1981. Notes on the geometry of crime. *Environmental criminology* (1981).
[6] R. V. Clarke and M. Felson. 2008. Introduction: Criminology, routine activity, and rational choice. 5 (2008), 1–14.
[7] Lawrence E Cohen and Marcus Felson. 1979. Social change and crime rate trends: A routine activity approach. *American sociological review* (1979), 588–608.
[8] Ellen G Cohn. 1990. Weather and crime. *British journal of criminology* 30, 1 (1990), 51–64.
[9] Derek B Cornish and Ronald V Clarke. 2014. *The reasoning criminal: Rational choice perspectives on offending.* Transaction Publishers.
[10] Chris Couch and Annekatrin Dennemann. 2000. Urban regeneration and sustainable development in Britain: The example of the Liverpool Ropewalks Partnership. *Cities* 17, 2 (2000), 137–147.
[11] Carl De Boor, Carl De Boor, Etats-Unis Mathématicien, Carl De Boor, and Carl De Boor. 1978. *A practical guide to splines.* Vol. 27. Springer-Verlag New York.
[12] Isaac Ehrlich. 1975. On the relation between education and crime. In *Education, income, and human behavior.* NBER, 313–338.
[13] Coral Featherstone. 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.* IEEE, 1–8.
[14] Marcus Felson and Ronald V Clarke. 1998. Opportunity makes the thief. *Police research series, paper* 98 (1998).
[15] Yanjie Fu, Guannan Liu, Spiros Papadimitriou, Hui Xiong, Yong Ge, Hengshu Zhu, and Chen Zhu. 2015. Real estate ranking via mixed land-use latent models. In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.* ACM, 299–308.
[16] Matthew S Gerber. 2014. Predicting crime using Twitter and kernel density estimation. *Decision Support Systems* 61 (2014), 115–125.
[17] Eva GT Green, Christian Staerkle, and David O Sears. 2006. Symbolic racism and WhitesâĂŹ attitudes towards punitive and preventive crime policies. *Law and Human Behavior* 30, 4 (2006), 435–454.
[18] Paul J Gruenewald, Bridget Freisthler, Lillian Remer, Elizabeth A LaScala, and Andrew Treno. 2006. Ecological models of alcohol outlets and violent assaults: crime potentials and geospatial analysis. *Addiction* 101, 5 (2006), 666–677.
[19] Hao Guo, Xin Li, Ming He, Xiangyu Zhao, Guiquan Liu, and Guandong Xu. 2016. CoSoLoRec: Joint Factor Model with Content, Social, Location for Heterogeneous Point-of-Interest Recommendation. In *International Conference on Knowledge Science, Engineering and Management.* Springer, 613–627.
[20] Bin Huang, Xiang-Yu Zhao, Kai Qi, Ming Tang, and Younghae Do. 2013. Coloring the complex networks and its application for immunization strategy. *Acta Physica Sinica* 62 (2013), 218902.
[21] Bruce P Kennedy, Ichiro Kawachi, Deborah Prothrow-Stith, Kimberly Lochner, and Vanita Gupta. 1998. Social capital, income inequality, and firearm violent crime. *Social science & medicine* 47, 1 (1998), 7–17.
[22] Kelvin Leong and Stephen CF Chan. 2013. A content analysis of web-based crime mapping in the world's top 100 highest GDP cities. *Crime Prevention & Community Safety* 15, 1 (2013), 1–22.
[23] Kelvin Leong and Anna Sung. 2015. A review of spatio-temporal pattern analysis approaches on crime analysis. *International E-Journal of Criminal Sciences* 9 (2015), 1–33.
[24] Renjie Liao, Xueyao Wang, Lun Li, and Zengchang Qin. 2010. A novel serial crime prediction model based on bayesian learning theory. In *Machine Learning and Cybernetics (ICMLC), 2010 International Conference on*, Vol. 4. IEEE, 1757–1762.
[25] Ye Liu, Yu Zheng, Yuxuan Liang, Shuming Liu, and David S Rosenblum. 2016. Urban water quality prediction based on multi-task multi-view learning. In *Proceedings of the International Joint Conference on Artificial Intelligence.*
[26] John Neter, Michael H Kutner, Christopher J Nachtsheim, and William Wasserman. 1996. *Applied linear statistical models.* Vol. 4. Irwin Chicago.
[27] E Britt Patterson. 1991. Poverty, income inequality, and community crime rates. *Criminology* 29, 4 (1991), 755–776.
[28] Matthew Ranson. 2014. Crime, weather, and climate change. *Journal of environmental economics and management* 67, 3 (2014), 274–302.
[29] Robert Tibshirani. 1996. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)* (1996), 267–288.
[30] Waldo R Tobler. 1970. A computer movie simulating urban growth in the Detroit region. *Economic geography* 46 (1970), 234–240.
[31] Hongjian Wang, Daniel Kifer, Corina Graif, and Zhenhui Li. 2016. Crime rate inference with big data. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.* ACM, 635–644.
[32] David Weisburd, Alese Wooditch, Sarit Weisburd, and Sue-Ming Yang. 2016. Do Stop, Question, and Frisk Practices Deter Crime? *Criminology & public policy* 15, 1 (2016), 31–56.
[33] Tong Xu, Hengshu Zhu, Xiangyu Zhao, Qi Liu, Hao Zhong, Enhong Chen, and Hui Xiong. 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.* ACM, 1285–1294.
[34] Jing Yuan, Yu Zheng, and Xing Xie. 2012. Discovering regions of different functions in a city using human mobility and POIs. In *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining.* ACM, 186–194.
[35] Xiangyu Zhao, Tong Xu, Qi Liu, and Hao Guo. 2016. Exploring the Choice Under Conflict for Social Event Participation. In *International Conference on Database Systems for Advanced Applications.* Springer, 396–411.
[36] Xiang-Yu Zhao, Bin Huang, Ming Tang, Hai-Feng Zhang, and Duan-Bing Chen. 2015. Identifying effective multiple spreaders by coloring complex networks. *EPL (Europhysics Letters)* 108, 6 (2015), 68005.
[37] Jiangchuan Zheng and Lionel Ming-Shuan Ni. 2013. Time-dependent trajectory regression on road networks via multi-task learning. In *Proceedings of the 27th AAAI Conference on Artificial Intelligence, AAAI 2013, Bellevue, Washington, USA.* 1048.
[38] Yu Zheng, Furui Liu, and Hsun-Ping Hsieh. 2013. U-air: When urban air quality inference meets big data. In *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining.* ACM, 1436–1444.
[39] Yu Zheng, Tong Liu, Yilun Wang, Yanmin Zhu, Yanchi Liu, and Eric Chang. 2014. Diagnosing New York city's noises with ubiquitous data. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing.* ACM, 715–725.