

# UAPD: Predicting Urban Anomalies from Spatial-Temporal Data

Xian Wu<sup>1</sup>, Yuxiao Dong<sup>1,2</sup>, Chao Huang<sup>1</sup>, Jian Xu<sup>1</sup>, Dong Wang<sup>1</sup>,  
Nitesh V. Chawla<sup>1\*</sup>

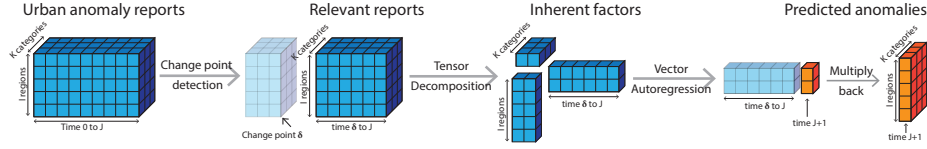
<sup>1</sup> Department of Computer Science and Engineering, and iCeNSA  
University of Notre Dame, USA,  
{xwu9,chuang7,jxu5,dwang5,nchawla\*}@nd.edu

<sup>2</sup> Microsoft Research Redmond, USA, yuxdong@microsoft.com

**Abstract.** Urban city environments face the challenge of disturbances, which can create inconveniences for its citizens. These require timely detection and resolution, and more importantly timely preparedness on the part of city officials. We term these disturbances as anomalies, and pose the problem statement: if it is possible to also predict these anomalous events (proactive), and not just detect (reactive). While significant effort has been made in detecting anomalies in existing urban data, the prediction of future urban anomalies is much less well studied and understood. In this work, we formalize the future anomaly prediction problem in urban environments, such that those can be addressed in a more efficient and effective manner. We develop the Urban Anomaly PreDiction (*UAPD*) framework, which addresses a number of challenges, including the dynamic, spatial varieties of different categories of anomalies. Given the urban anomaly data to date, *UAPD* first detects the change point of each type of anomalies in the temporal dimension and then uses a tensor decomposition model to decouple the interrelations between the spatial and categorical dimensions. Finally, *UAPD* applies an autoregression method to predict which categories of anomalies will happen at each region in the future. We conduct extensive experiments in two urban environments, namely New York City and Pittsburgh. Experimental results demonstrate that *UAPD* outperforms alternative baselines across various settings, including different region and time-frame scales, as well as diverse categories of anomalies.

## 1 Introduction

Timely resolution of urban environment and infrastructure problems is an essential component of a well-functioning city. Cities have developed various reporting systems, such as 311 services [1], for the citizens to report, track, and comment on the urban anomalies that they encounter in their daily lives. These systems enable the city government institutions to respond on a timely basis to disturbances or anomalies such as noise, blocked driveway and urban infrastructure malfunctions [14]. However, these systems and resulting actions rely on accurate and timely reporting by the citizens, and are by nature reactive. That leads us to



**Fig. 1.** The Urban Anomaly PreDiction (*UAPD*) Framework.

the question: what if we could *predict* which regions of a city will observe certain categories of anomalies in advance? We posit that this would enhance resource allocation and budget planning, and also result in more timely and efficient resolution of issues, thereby minimizing the disruption to the citizens’ lives [19]. However, the development of such an urban anomaly prediction system faces several challenges:

- 1) *Anomaly Dynamics*: The factors underlying urban anomalies may change over time. For example, anomalies in the winter may stem from winter related severe weather (e.g., snow), and it may be infeasible to train the predictive model by using historical data between spring and fall.
- 2) *Coupled Multi-Dimensional Correlations*: There may exist strong signals among the locations, time, and categories of occurred anomalies, that is, certain categories of anomalies occur at specific locations and/or specific time. Traditional time-series based models such as Autoregressive Moving-Average (ARMA) [28] and Gaussian Processing (GP) [8] largely rely on heuristic temporal features to make predictions. However, these approaches fail to capture the spatial-temporal factors that drive different categories of anomalies occur.

**Contributions.** To achieve a comprehensive solution that addresses these challenges and meets the goal of prediction, we develop a unified three-phase *Urban Anomaly PreDiction (UAPD)* framework to predict urban anomalies from spatial-temporal data—urban anomaly reports. As illustrated in Figure 1, at the first phase, we propose a probabilistic model, whose parameters are inferred via Markov chain Monte Carlo, to detect the change point of the historical anomaly records of a city, such that only most relevant reports are used for the prediction of future anomalies. At the second phase, we model the anomaly data starting from the detected change point of all regions in the city with a three-dimensional tensor, where each of three dimensions stands for the regions, time slots, and categories of occurred anomalies, respectively. We then decompose the tensor to incorporate the underlying relationships between each dimension into their corresponding inherent factors in the tensor. Subsequently, the prediction of anomalies in next time slot is furthered to a time-series prediction problem. At the third phase, we leverage Vector Autoregression to capture the inter-dependencies of inherent factors among multiple time series, generating the prediction results.

We evaluate the performance of *Urban Anomaly PreDiction (UAPD)* by using two real-world datasets collected from the 311 service in New York City and

Pittsburgh. The evaluation results with different region scales (i.e., precincts and regions divided by road segments) and different time-frames of training data show that our framework can predict different categories of anomalies more accurately than the state-of-the-art baselines.

In summary, our main contributions are as follows:

- We formalize the urban anomaly prediction problem from spatial-temporal data and develop a unified *UAPD* framework to predict future occurrences of different anomalies at different urban areas.
- In *UAPD*, we propose a change point detection solution to address the challenge of anomaly dynamics and leverage tensor decomposition to model the interrelations among multiple dimensions of urban anomaly data.
- We conduct extensive experiments with different scenarios in distinct urban environments to demonstrate the effectiveness of our presented framework.

## 2 Related Work

**Anomaly Detection.** Our work is related to urban sensing-based anomaly detection [6,5,23,7,18]. For example, Doan et al. presented a set of new clustering and anomaly detection techniques using a large-scale urban pedestrian dataset [6]. Chawla et al. used the Principal Component Analysis (PCA) scheme to discover events which may have caused anomalous behavior to appear in road traffic flow [5]. Pan et al. developed a novel system to detect anomalies according to drivers’ routing behavior [23]. Zheng et al. proposed a probability-based anomaly detection method to discover collective anomalies from multiple spatial-temporal datasets [32]. Le et al. proposed an online algorithm to detect anomaly occurrence probability given the data from various sensors in highly dynamic contexts [18]. However, all the above solutions identified urban anomalies *after* they happen. In contrast, this paper develops a principled approach to *predict* urban anomalies in different regions of a city. Although there exists one recent work on anomaly prediction [15] by considering the dependency of anomaly occurrence between regions, two significant limitations exist: (i) it ignored the fact that anomalies of different categories can be correlated; (ii) it assumed that some portion of ground truth data from the predicted time slot are known. However, such assumptions do not always hold in the practical scenario for a couple of reasons. Firstly, dependencies between anomalies with different categories are ubiquitous in urban sensing. Secondly, it is difficult to know the ground truth information from the predicted time slot beforehand. To overcome those limitations, this paper develops a new urban anomaly prediction framework to explicitly explore the inter-category dependencies and model the time-evolving inherent factors with respect to different categories, without requiring any ground truth information from the predicted time slot.

**Time Series Prediction.** Our work is related to the literature that focus on time series prediction [25,8,28,2]. In particular, Vallis et al. [25] proposed a novel statistical technique to automatically detect long-term anomalies in cloud data. A non-parametric approach Gaussian Processing (GP) has been developed

to solve the general time series prediction problem [8]. Additionally, Wiesel et al. [28] proposed a time varying Autoregression Moving Average (ARMA) model for co-variance estimation. Bao et al. [2] proposed a Particle Swarm Optimization (PSO)-based strategy to determine the number of sub-models in a self-adaptive mode with varying prediction horizons. Inspired by the work above, we develop a new scheme to explicitly consider the anomaly distribution dependency between regions and incorporate it into the time series prediction model.

**Bayesian Inference.** Bayesian inference has been widely used in the data analytics communities [30,21,20]. For example, Yang et al. studied the problem of learning social knowledge graphs by developing a multi-modal Bayesian embedding model [30]. Lian et al. proposed a sparse Bayesian content-aware collaborative filtering approach for implicit feedback recommendation [21]. Lee et al. proposed a Bayesian nonparametric model for medical risks prediction by exploring phenotype topics from electronic health records [20]. In this paper, we study the problem of urban anomaly prediction by proposing a Bayesian inference model to capture the evolving relationships of anomaly sequences.

### 3 Problem Definition

Given the historical sequences of anomalies within a city’s geographical regions, the objective is to predict whether certain categories of anomalies will happen at certain places in the future. We use a three-dimensional tensor  $\mathcal{X} \in \mathbb{R}^{I \times J \times K}$  to represent the anomaly sequences of all regions in a city.  $I, J, K$  denote the number of regions (e.g., geographical areas divided by the city’s road network), time slots, and anomaly categories (e.g., noise, blocked driveway, illegal parking, etc.) in the data, respectively. Each element  $\mathcal{X}_{i,j,k}$  represents the number of the  $k$ -category anomalies that happened at region  $i$  at time  $j$ .

**Urban Anomaly Prediction Problem:** Given the historical anomaly data  $\mathcal{X} \in \mathbb{R}^{I \times J \times K}$  of a city, the goal is to learn a predictive function  $f(\mathcal{X}) \rightarrow \mathcal{X}_{:,J+1,:}$  to infer future occurrences of each category  $k$  of anomalies at each region  $i$  at time  $J+1$ .

The urban anomaly prediction solution needs to incorporate additional factors as noted in Section 1. First, there may exist periodic and temporary patterns in urban anomalies. For example, the temporary road construction serves as the inherent factors for noise related anomalies. This pattern makes the model trained using data collected during the construction period ineffective for predictions over the following time-frame. Second, the urban anomaly data covers multiple aspects of coupled information—the location, time, and category of anomalies. It is unclear whether certain types of anomalies occur at correlated locations or time, and if so, how we can model the multidimensional data.

### 4 The *UAPD* Framework

Figure 1 illustrates the *UAPD* framework, with the following steps: 1) *UAPD* leverages Bayesian inference to detect the change point in the anomaly se-

**Table 1.** Symbols and Definitions

Symbol	Interpretation
$i, j, k$	the indices of regions, time slots, anomaly categories
$I, J, K$	the number of regions, time slots, anomaly categories
$M$	The anomaly sequence matrix
$M_j^k$	The element in the $k$ -th category, $j$ -th time slot of the abnormal region matrix
$\mathcal{X}$	The anomaly tensor
$R, T, C$	The inherent factor matrix with regions, time slots, anomaly categories, respectively
$L$	The number of inherent factors (i.e., the rank of tensor)
$\delta$	The change point
$\alpha, \beta$	Parameters in Poisson distributions before/after change point
$\mathbf{a}, \mathbf{b}$	Shape/Scale parameters in Gamma distributions
$\mathbf{p}$	Probability of each time slot being the change point

quence; 2) *UAPD* incorporates the spatial-temporal anomaly data into a three-dimensional tensor, which can be decomposed to learn the latent correlations between all dimensions; 3) *UAPD* applies the Vector Autoregression model to predict the occurrence of different categories of anomalies in each region of a city. We will explain these three steps in detail in the following subsections.

#### 4.1 Change Point Detection

Given an anomaly sequence, conventional prediction methods usually train a learning model by using all available data between time 0 and  $J$  to predict the future anomalies at time  $J+1$  [22]. However, the occurrence of urban anomalies can be largely influenced by temporary events and periodic patterns. For example, the road construction in a certain area can result in many noise reports during the construction period. As a result, the anomaly sequence of this area may be dramatically changed after the construction, i.e., a significant decrease in noise reports, leading to an ineffective prediction for post-construction anomalies by training models on the anomaly sequence that covers the construction period. To address this issue, we propose a Bayesian inference based method to detect the change point  $\delta$  in an anomaly sequence between time 0 and  $J$ . The sequence follows two different data distributions before and after the change point. Thus, *UAPD* learns the anomaly predictive function, over the time sequences, by leveraging the most relevant latent factors that caused the anomalies.

Specifically, we aim to detect the change point of the anomaly sequence matrix  $M$  with regions that have  $k$  categories of anomalies. To do so, we first model  $M_j^k$  that reports the  $k$ -th category of anomalies at time  $j$  as a Poisson distribution. Before and after the change point  $\delta$ ,  $M_j^k$  follows Poisson distributions with different parameter configurations,  $\alpha_k$  and  $\beta_k$ , respectively. Following the standard Bayesian inference, we set the conjugate prior of the Poisson distribution as the Gamma distribution with  $\mathbf{a}$  and  $\mathbf{b}$  as its shape and scale parameters, respectively. Finally, the change point  $\delta$  is set to obey a multinomial distribution with parameters  $\mathbf{p}=(p_0, \dots, p_j, \dots, p_J)$ , where  $p_j$  denotes the probability of time  $j$  to be the change point. The generative process behind our model can

---

**Algorithm 1:** Gibbs Sampling for Change Point

---

```

Initialize model parameter  $\{ \boldsymbol{\alpha}^{(1)}, \boldsymbol{\beta}^{(1)}, \delta^{(1)} \}$ ;
for  $itr = 1, \dots, iterations$  do
    for  $k = 1, \dots, K$  do
        | Sample the  $\alpha_k^{(itr+1)}$  according to  $\Pr(\alpha_k | \delta, \boldsymbol{\beta}^{itr}, M_{1:J})$  in Eq. (7);
    end
    for  $k = 1, \dots, K$  do
        | Sample the  $\beta_k^{(itr+1)}$  according to  $\Pr(\beta_k | \delta, \boldsymbol{\alpha}^{(itr+1)}, M_{1:J})$  in Eq. (8);
    end
    for  $\delta' = 1, \dots, J$  do
        | Calculate parameter  $p_{\delta'}^{(itr+1)}$  according to Eq. (9);
    end
    Sample the  $\delta$  according to  $\Pr(\delta | \boldsymbol{\alpha}^{(itr+1)}, \boldsymbol{\beta}^{(itr+1)}, M_{1:J})$ ;
end

```

---

be summarized as follows:

$$\begin{aligned}
 \delta &\sim \text{Multinomial}(\delta; \mathbf{p}) \\
 \alpha_k &\sim \text{Gamma}(\alpha_k; a_k^\alpha, b_k^\alpha) \\
 \beta_k &\sim \text{Gamma}(\beta_k; a_k^\beta, b_k^\beta) \\
 M_j^k &\sim \begin{cases} \text{Poisson}(M_j^k; \alpha_k), & 1 \leq j \leq \delta \\ \text{Poisson}(M_j^k; \beta_k), & \delta + 1 \leq j \leq J \end{cases}
 \end{aligned} \tag{1}$$

where  $M_j^k$  is the number of regions which have anomalies of  $k$ -th category at time  $j$ . The variable on the left side of the semi-colon is assigned a probability under the parameter on the right side. The posterior distribution is formally defined as follow:

$$\begin{aligned}
 \Pr(\boldsymbol{\alpha}, \boldsymbol{\beta}, \delta | M_{1:J}) &\propto \Pr(M_{1:\delta} | \boldsymbol{\alpha}) \Pr(M_{\delta+1:J} | \boldsymbol{\beta}) \Pr(\boldsymbol{\alpha}) \Pr(\boldsymbol{\beta}) \Pr(\delta) \\
 &= \prod_{k=1}^K \left( \prod_{j=1}^{\delta} \Pr(M_j^k | \alpha_k) \prod_{j=\delta+1}^J \Pr(M_j^k | \beta_k) \Pr(\alpha_k) \Pr(\beta_k) \right) \Pr(\delta)
 \end{aligned} \tag{2}$$

We could estimate the change point  $\delta$  and distribution parameters  $\boldsymbol{\alpha}, \boldsymbol{\beta}$  by maximizing the full joint distribution. However, to avoid fine parameter tuning and make accurate detection on change point, we apply the Markov Chain Monte Carlo (MCMC) method to infer parameter. In addition, since we have multiple random variables in this problem, we utilize the Gibbs sampling algorithm to execute the MCMC method. The basic idea in Gibbs sampling is that we sample the new value of each variable in each iteration accordingly by fixing other variables [24,17]. Algorithm 1 presents the outline of Gibbs sampling for change point detection. The conditional distribution of each variable and parameter derivations are given in the Appendix.

## 4.2 Tensor Decomposition

The inherent factors of anomalies among different regions, time slots and anomaly categories are revealed by tensor decomposition on a three dimensional tensor representing the anomaly records. Since the anomaly distributions of regions are often subject to changes in different time periods, the objective of tensor decomposition is to model the temporal effects of anomalies distribution in each region by learning the inherent factors as well as adopting these factors to different time periods.

We use the CANDECOMP/PARAFAC (CP) decomposition approach [3] to factorize the tensor into three different matrices  $R \in \mathbb{R}^{I \times L}$ ,  $T \in \mathbb{R}^{J \times L}$  and  $C \in \mathbb{R}^{K \times L}$ . Here  $L$  is the number of inherent factors and is indexed by  $l$ .  $R$ ,  $T$  and  $C$  are the inherent factor matrices with respect to  $I$  regions,  $J$  time slots and  $K$  anomaly categories, respectively. We can express the three-way tensor factorization of  $\mathcal{X}$  as:

$$\mathcal{X} \approx \sum_{l=1}^L R_{:,l} \circ T_{:,l} \circ C_{:,l} \quad (3)$$

where  $R_{:,l}$ ,  $T_{:,l}$  and  $C_{:,l}$  represent the  $l$ -th column of  $R$ ,  $T$  and  $C$ .  $\circ$  denotes the vector outer product. Each entry  $\mathcal{X}_{i,j,k}$  in tensor  $\mathcal{X}$  can be computed as the inner-product of three  $L$ -dimensional vectors as follows:

$$\mathcal{X}_{i,j,k} \approx \langle R_i, T_j, C_k \rangle \equiv \sum_{l=1}^L R_{i,l} T_{j,l} C_{k,l} \quad (4)$$

Although many techniques can be applied to CP decomposition, we utilize the *Alternative Least Square (ALS)* algorithm [4,12], which has been shown to outperform other algorithms in terms of solution quality [9].

## 4.3 Vector Autoregression

Using the tensor decomposition model discussed above, we can learn the inherent factors of anomalies from different time slots. Based on the three matrices generated by CP decomposition, we formulate the anomaly prediction problem as the time series prediction task on  $T \in \mathbb{R}^{J \times L}$ , since the inherent factors in other two dimensions remain constant. We use *Vector Autoregression (VAR)* [11] to capture the linear inter dependencies of inherent factors among multiple time series. In particular, VAR model formulates the evolution of a set of  $L$  inherent factors over the same sample period (i.e., time slot  $j = 1, 2, \dots, J$ ) as a linear function of their past values. We define the order  $S$  of VAR to represent the time series in the previous  $S$  time slots. Formally, it can be expressed as follows:

$$T_{j+1} = d + \sum_{s=1}^S B_s T_{j-s} + \epsilon_j \quad (5)$$

where  $d$  is a  $L \times 1$  vector of constants,  $B_s$  is a time-invariant  $L \times L$  matrix and  $\epsilon_j$  is a  $L \times 1$  vector of errors. Many techniques have been proposed for VAR order selection, such as Bayesian Information Criterion (BIC) [27], Final Prediction Error (FPE) [26] and Akaike information criteria (AIC) [10]. In this paper, we utilize the lowest AIC value to decide the order of VAR. Hence, the number of anomalies with the  $k$ -th category in region  $R_i$  in the next time slot can be derived as:

$$\mathcal{X}_{i,J+1,k} = \langle R_i, T_{J+1}, C_k \rangle \equiv \sum_{l=1}^L R_{i,l} T_{J+1,l} C_{k,l} \quad (6)$$

## 5 Evaluation

In this section, we conduct experiments on two real-world datasets collected from 311 Service in New York City (NYC) and Pittsburgh, respectively. In particular, we answer the following research questions:

- **Q1:** How does *UAPD* perform as compared to the state-of-the-art solutions in predicting different categories of urban anomalies?
- **Q2:** How does *UAPD* perform in anomaly prediction with respect to different geographical region scales?
- **Q3:** How does *UAPD* perform in anomaly prediction with respect to different time-frames (i.e., #time-slots  $J$ )?
- **Q4:** Can *UAPD* effectively capture the change point of an anomaly sequence? Is change point detection effective for urban anomaly prediction?
- **Q5:** Is *UAPD* stable with regard to the rank of tensor  $L$  (i.e., the number of inherent factors)?

### 5.1 Experimental Setup

**Datasets.** We evaluated *UAPD* on real-world urban anomaly reports datasets collected from New York City (NYC) OpenData portal <sup>3</sup> and Pittsburgh OpenData portal <sup>4</sup>, respectively. Those datasets are collected from 311 Service that is an urban anomaly report platform, which allows citizens to report complaints about urban anomalies such as blocked driveway by texting, phone call or mobile app. Each complaint record contains the timestamp, coordinates and category of anomaly. Different cities may have different anomaly categories due to different urban properties [29]. For the New York City datasets, we focus on 4 key anomaly categories (e.g., Blocked Driveway, Illegal Parking) which are selected in [32,15,31]. For the Pittsburgh datasets, we select the top frequently reported anomaly categories (e.g., Potholes, Refuse Violations) as the target predictive anomaly categories. The New York city data was collected from Jan 2014 to Dec 2014 and the Pittsburgh data was collected from Jan 2016 to Dec 2016.

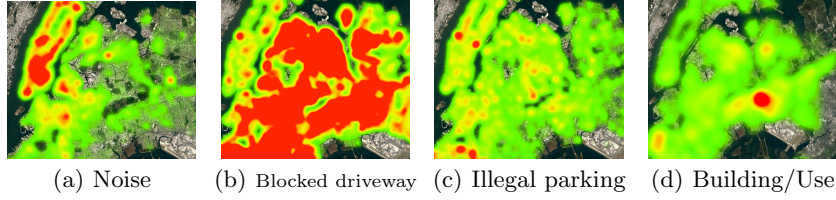
<sup>3</sup> <https://data.cityofnewyork.us/>

<sup>4</sup> <http://www.wprdc.org/>



**Table 2.** Data Statistics

City	Anomaly Category	Number of Instances
New York City	Noise	151,174
	Blocked Driveway	92,335
	Illegal Parking	69,100
	Building/Use	27,724
Pittsburgh	Potholes	3,361
	Snow/Ice removal	2,504
	Building Maintenance	2,352
	Weeds/Debris	2,023
	Refuse Violations	1,196
	Abandoned Vehicle	924
	Replace/Repair a Sign	620

**Fig. 2.** Geographical Distribution of Anomalies with Different Categories.

The statistics of these datasets are summarized in Table 2. Due to space limit, we only show the geographical distribution of different categories of anomalies in NYC in Figure 2. In those figures, firstly for the same anomaly category, we can observe that different regions have different number of anomalies. Secondly for the same region, we can observe that different anomaly categories have different distributions, which is the main motivation to use tensor decomposition model to identify the inherent factors by considering the correlations between different regions and different categories. Similar geographical distributions can be observed in Pittsburgh datasets.

**Baselines and Evaluation Metrics.** We compare our proposed scheme to the following state-of-the-art techniques:

- *CUAPS*: it predicts the anomaly by exploring the dependency of anomaly occurrence between different regions and use some portion of ground truth data from the predicted time slot [15].
- *Random Forest (RF)*: it incorporates the spatial-temporal features of anomaly sequence into the classification model.
- *Adaptive Boosting (AdaBoost)*: it predicts the anomaly occurrence of each region by maintaining a distribution of weights over the training examples.
- *Gaussian Processing (GP)*: it is a nonparametric approach that predicts the abnormal state of each region by exploring temporal features [8].

- *Autoregressive Moving-Average Model (ARMA)*: it is a time varying model that predicts the abnormal state of each region by exploring anomaly traces features [28].
- *Logistic Regression (LR)*: it is a regression model that estimates the abnormal state of each region by exploring temporal features [13].

## 5.2 Performance Validation

We use the following metrics to evaluate the estimation performance of the *UAPD* scheme: *F1-measure*, *Precision* and *Recall*.

**Effectiveness of *UAPD* (Q1, Q2, Q3).** In this subsection, we present the performance of *UAPD* on two real-world datasets with different geographical region scales and different time slots  $J$ . The source code of *UAPD* is publicly available <sup>5</sup>. In our evaluation, we consider two versions of our scheme: (i) *UAPD-CP*: a simplified version of *UAPD* framework which does not include the change point detection step; (ii) *UAPD*: the full version of the proposed scheme. We compare the *UAPD* scheme and its variant with the state-of-the-art algorithms.

**Results on New York City Dataset.** To investigate the effect of geographical scale, we partition the New York City into different geographical regions using the following methods:

- *High-Level Region*: New York City is divided into 76 geographical areas based on the political and administrative districts information <sup>6</sup>. We refer to each geographical region as a *high-level region*.
- *Fine-Grained Region*: we use major roads (road segments with levels from  $L_1$  to  $L_5$ ) to partition the entire city, which results in 862 geographical area [32]. We refer to each geographical area as a *fine-grained region*.

Based on the above region partition methods, each anomaly can be mapped into a particular region.

The evaluation results on NYC datasets with high-level region are shown in Table 3 (Jun) and Table 4 (Dec), respectively. The evaluation results on NYC datasets with fine-grained region are shown in Table 5 (Jun) and Table 6 (Dec), respectively. To investigate the effect of the number of time slots  $J$ , we study the performance of all compared schemes with different values of  $J$ . In Table 3 and Table 4, we set  $J$  to be 150 (from January to May). In Table 5 and Table 6, we set  $J$  to 330 (from January to November). The evaluation results are the average across the 10 consecutive days for prediction. From the results, we can observe that *UAPD* consistently outperforms the state-of-the-art baselines in most cases on different categories of anomalies. For example, *UAPD* outperforms the best baseline (AdaBoost) for high-level region-Building/Use case by 16.2% on F1-Score and the best baseline (RF) for fine-grained region-Blocked Driveway

<sup>5</sup> <https://bitbucket.org/xianwu9/uapd>

<sup>6</sup> <https://data.cityofnewyork.us/Public-Safety/Police-Precincts/78dh-3ptz/data>

**Table 3.** Prediction Results on Jun with High-Level Region in NYC

Anomaly	Noise			Blocked Driveway			Illegal Parking			Building/Use		
Algorithm	F1	Precision	Recall	F1	Precision	Recall	F1	Precision	Recall	F1	Precision	Recall
<b>UAPD</b>	0.966	0.944	0.989	0.878	0.838	0.923	0.867	0.776	0.980	0.679	0.602	0.779
UAPD-CP	0.952	0.940	0.964	0.837	0.809	0.867	0.867	0.776	0.982	0.628	0.549	0.733
CUAPS	0.930	0.941	0.920	0.824	0.826	0.825	0.808	0.794	0.822	0.576	0.620	0.539
RF	0.958	0.956	0.960	0.841	0.864	0.819	0.822	0.807	0.837	0.594	0.65	0.547
AdaBoost	0.950	0.947	0.953	0.841	0.827	0.856	0.812	0.783	0.843	0.584	0.572	0.596
GP	0.926	0.936	0.916	0.793	0.814	0.774	0.778	0.782	0.774	0.588	0.550	0.631
ARMA	0.926	0.962	0.892	0.818	0.946	0.720	0.688	0.866	0.571	0.408	0.804	0.274
LR	0.913	0.942	0.886	0.771	0.902	0.673	0.837	0.794	0.886	0.646	0.665	0.628

**Table 4.** Prediction Results on Dec with High-Level Region in NYC

Anomaly	Noise			Blocked Driveway			Illegal Parking			Building/Use		
Algorithm	F1	Precision	Recall	F1	Precision	Recall	F1	Precision	Recall	F1	Precision	Recall
<b>UAPD</b>	0.892	0.838	0.954	0.894	0.880	0.908	0.873	0.779	0.992	0.724	0.639	0.836
UAPD-CP	0.893	0.834	0.962	0.894	0.891	0.898	0.873	0.779	0.992	0.733	0.655	0.833
CUAPS	0.881	0.826	0.945	0.869	0.904	0.837	0.805	0.762	0.853	0.613	0.642	0.587
RF	0.875	0.866	0.885	0.855	0.868	0.842	0.810	0.819	0.8	0.607	0.673	0.552
AdaBoost	0.880	0.849	0.914	0.848	0.853	0.844	0.833	0.828	0.839	0.623	0.625	0.621
GP	0.795	0.819	0.772	0.834	0.855	0.814	0.802	0.805	0.798	0.583	0.567	0.599
ARMA	0.823	0.897	0.761	0.817	0.959	0.712	0.708	0.861	0.601	0.454	0.813	0.315
LR	0.773	0.844	0.714	0.805	0.891	0.734	0.834	0.786	0.887	0.605	0.710	0.527

case by 34.7% on F1-Score. In the occasional case that *UAPD* misses the best performance, it still generates very competitive results. The results are consistent with different geographical region scales and historical time slots.

**Table 5.** Prediction Results on June with Fine-grained Region in NYC

Anomaly	Noise			Blocked Driveway			Illegal Parking			Building/Use		
Algorithm	F1	Precision	Recall	F1	Precision	Recall	F1	Precision	Recall	F1	Precision	Recall
<b>UAPD</b>	0.630	0.534	0.768	0.555	0.491	0.639	0.421	0.404	0.441	0.262	0.268	0.257
UAPD-CP	0.635	0.593	0.684	0.511	0.550	0.478	0.373	0.468	0.310	0.122	0.367	0.073
CUAPS	0.486	0.599	0.409	0.415	0.478	0.366	0.220	0.187	0.268	0.170	0.118	0.309
RF	0.553	0.582	0.526	0.412	0.466	0.369	0.306	0.357	0.267	0.166	0.221	0.133
AdaBoost	0.528	0.522	0.535	0.397	0.370	0.429	0.301	0.264	0.350	0.183	0.155	0.222
GP	0.551	0.500	0.613	0.393	0.349	0.452	0.326	0.271	0.409	0.216	0.181	0.269
ARMA	0.404	0.787	0.272	0.199	0.631	0.118	0.104	0.517	0.058	0.028	0.412	0.014
LR	0.607	0.630	0.587	0.382	0.615	0.278	0.275	0.453	0.197	0.098	0.346	0.057

**Results on Pittsburgh Dataset.** We repeated the same experiments on Pittsburgh dataset. Considering the space limit, we only present the evaluation results for high-level regions in June. In particular, we partition the Pittsburgh city based on its council districts information<sup>7</sup>. The evaluation results on different categories of anomalies are shown in Table 7. The reported results are also the average across the 10 consecutive days for prediction. In Table 7, we

<sup>7</sup> <http://pittsburghpa.gov/council/maps>

**Table 6.** Prediction Results on Dec with Fine-grained Region in NYC

Anomaly	Noise			Blocked Driveway			Illegal Parking			Building/Use		
Algorithm	F1	Precision	Recall	F1	Precision	Recall	F1	Precision	Recall	F1	Precision	Recall
<b>UAPD</b>	0.521	0.466	0.589	0.558	0.570	0.546	0.403	0.455	0.362	0.171	0.406	0.109
UAPD-CP	0.373	0.366	0.380	0.530	0.537	0.522	0.373	0.389	0.359	0.210	0.247	0.183
CUAPS	0.423	0.437	0.410	0.452	0.508	0.407	0.242	0.270	0.218	0.111	0.096	0.133
RF	0.438	0.545	0.366	0.429	0.522	0.364	0.293	0.384	0.237	0.116	0.198	0.082
AdaBoost	0.380	0.338	0.433	0.393	0.367	0.422	0.290	0.257	0.332	0.160	0.124	0.227
GP	0.415	0.346	0.517	0.453	0.415	0.499	0.355	0.297	0.441	0.219	0.188	0.262
ARMA	0.358	0.715	0.239	0.253	0.663	0.156	0.151	0.641	0.086	0.023	0.286	0.012
LR	0.520	0.641	0.438	0.504	0.598	0.435	0.338	0.516	0.251	0.077	0.286	0.044

**Table 7.** Prediction Results on June with High-Level Regions in Pittsburgh

Anomaly	Potholes			Weeds/Debris			Building Maintenance		
Algorithm	F1	Precision	Recall	F1	Precision	Recall	F1	Precision	Recall
<b>UAPD</b>	0.848	0.736	1.000	0.909	0.844	0.984	0.624	0.531	0.756
UAPD-CP	0.839	0.769	0.923	0.887	0.851	0.926	0.698	0.577	0.882
CUAPS	0.772	0.736	0.811	0.724	0.730	0.718	0.523	0.466	0.595
RF	0.793	0.842	0.750	0.794	0.862	0.735	0.585	0.649	0.533
AdaBoost	0.760	0.754	0.766	0.596	0.739	0.500	0.558	0.585	0.533
GP	0.733	0.716	0.750	0.812	0.800	0.824	0.615	0.542	0.711
ARMA	0.560	0.778	0.437	0.631	0.813	0.514	0.451	0.615	0.616
LR	0.778	0.790	0.766	0.821	0.833	0.808	0.494	0.525	0.467
Anomaly	Refuse Violations			Abandoned Vehicle			Replace/Repair a Sign		
Algorithm	F1	Precision	Recall	F1	Precision	Recall	F1	Precision	Recall
<b>UAPD</b>	0.615	0.522	0.750	0.642	0.515	0.854	0.622	0.509	0.800
UAPD-CP	0.535	0.451	0.657	0.628	0.551	0.730	0.487	0.404	0.613
CUAPS	0.076	0.500	0.041	0.394	0.441	0.357	0.133	0.666	0.074
RF	0.490	0.522	0.462	0.560	0.618	0.512	0.364	0.526	0.278
AdaBoost	0.490	0.522	0.462	0.624	0.558	0.707	0.382	0.406	0.361
GP	0.393	0.342	0.461	0.494	0.478	0.512	0.459	0.447	0.472
ARMA	0.133	0.500	0.077	0.280	0.533	0.195	0.052	0.500	0.027
LR	0.324	0.545	0.231	0.500	0.543	0.463	0.217	0.500	0.139

can observe that our scheme *UAPD* still outperforms other baselines in most of the evaluation metrics. Additionally, we can observe that *UAPD* outperforms *UAPD-CP* (without detecting any change point on anomaly sequence) in most cases. Since there do not exist reports of Snow/Ice removal category of anomaly on June, we do not consider this category in this evaluation.

The performance improvements of *UAPD* are achieved by (i) carefully considering dynamic causes of anomalies; (ii) explicitly exploiting inherent factors of different categories of anomalies; (iii) carefully handling the evolving relationship among different anomaly categories and regions, with Bayesian inference and tensor decomposition, which are missing from the state-of-the-art solutions.

**Analysis of Change Point Detection (Q4).** Additionally, we conducted experiments to analyze the results of change point detection (i.e., the first compo-

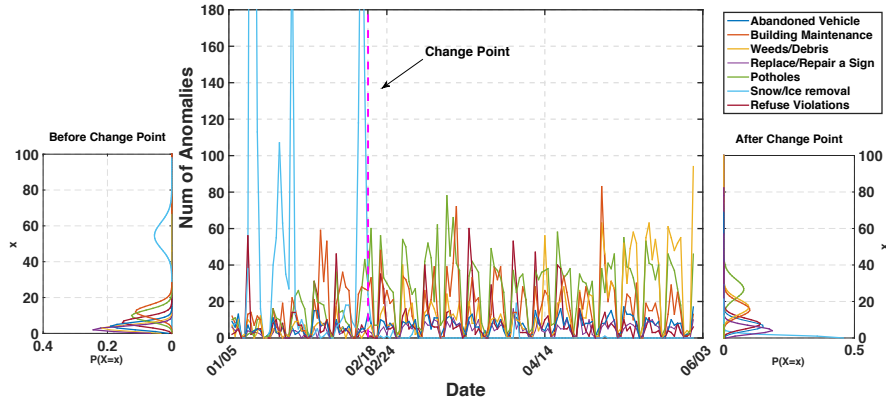


Fig. 3. Analysis of Change Point Detection.

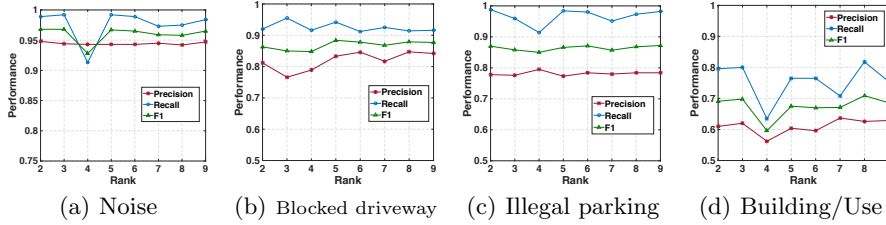


Fig. 4. Performance w.r.t Rank Parameter  $L$  on Jun. with high-level regions in NYC

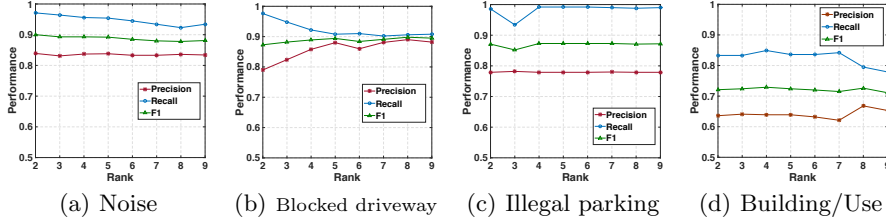


Fig. 5. Performance w.r.t Rank Parameter  $L$  on Dec. with high-level regions in NYC

ment of the proposed framework) and further visualize the result. The detection results on Pittsburgh datasets are shown in Figure 3. We kept the value of  $J$  the same as the above experiments. In Figure 3, we can observe that the detected starting point of anomaly sequences is Feb 18, 2016 instead of the actual beginning time (i.e., Jan 1, 2016). Moreover, the supplementary figures in both left and right side show that the anomaly distributions before the detected change point significantly differs from that after the change point, which demonstrates the effectiveness of our change point detection on anomaly sequences.

**Impact of Rank Parameter (Q5).** To investigate the effect of the only parameter (i.e., rank parameter  $L$ ) in our framework, we studied the performance of our proposed scheme by varying the value of  $L$ . Particularly, we vary the value of  $rank$  from 2 to 9. The evaluation results on NYC datasets are shown in Figure 5 and Figure 4. We can observe that the performance of *UAPD* is stable with the value of  $rank$  from 5 to 8 which satisfy the value selection rubric of rank parameter  $L$  in tensor decomposition [16]. In our experiments, we set the value of  $rank$  to 8. Considering the space limit, we only present the results on NYC datasets. In the results on Pittsburgh datasets, similar results can be observed.

## 6 Conclusion

In this paper, we develop a Urban Anomaly PreDiction(*UAPD*) framework to predict urban anomalies from spatial-temporal data. *UAPD* enables the accurate prediction of the occurrences of different anomalies at each region of a city in the future. *UAPD* explicitly detects the change point of the anomaly sequences and also explores the time-evolving inherent factors and their relationships with each dimension tensor (i.e., regions and anomaly categories). We evaluate our presented framework on two sets of urban anomaly reports collected from 311 Service in New York City and Pittsburgh, respectively. The results show that *UAPD* significantly outperforms state-of-the-art baselines, and provides a framework for being predictive about disturbances, enabling the city officials to be more proactive in their preparation and response.

## 7 Appendix

In this section, we give the specific form of conditional posterior functions used in algorithm 1. According to the posterior function in Equation (3), we can obtain the full conditional posterior of  $\alpha$  given  $\beta, \delta$  as follows:

$$\begin{aligned} \Pr(\alpha_k | \delta, \beta, M_{1:J}) &= \text{Gamma}(\alpha_k; {}^*a_\alpha^k, {}^*b_\beta^k) \\ {}^*a_\alpha^k &= a_\alpha^k + \sum_{j=1}^{\delta} M_j^k; \quad {}^*b_\alpha^k = \delta + b_\alpha^k \end{aligned} \quad (7)$$

The update of parameters in Gamma distribution of  $k$ -th category anomaly before the change point is given by  ${}^*a^k, {}^*b^k$ . Since we utilize the conjugate prior for the  $\alpha$ , the conditional distribution of  $\alpha$  follows the Gamma distribution. Similarly, the conditional posterior of  $\beta$  given  $\alpha, \delta$  has the same form as:

$$\begin{aligned} \Pr(\beta_k | \delta, \alpha, M_{1:J}) &= \text{Gamma}(\beta_k; {}^*a_\alpha^k, {}^*b_\beta^k) \\ {}^*a_\beta^k &= a_\alpha^k + \sum_{j=\delta+1}^J M_j^k; \quad {}^*b_\beta^k = J - \delta + b_\beta^k \end{aligned} \quad (8)$$

Finally, the full conditional posterior distribution of  $\delta$  is a Multinomial distribution:

$$\Pr(\delta|\alpha, \beta, M_{1:J}) = \text{Multinomial}(\delta; \mathbf{p})$$

$$p_{\delta'} = \exp\left(\sum_{k=1}^K \left(\sum_{j=1}^{\delta'} M_j^k \log \alpha_k + \sum_{j=\delta'}^J M_j^k \log \beta_k - \delta' \alpha_k - (J - \delta') \beta_k\right) + \log \sigma\right)$$
(9)

where  $\sigma$  is the normalization constant of  $p_{\delta'}$ .  $\mathbf{p} = (p_1, \dots, p_{\delta'}, \dots, p_J)$

## Acknowledgments

Research was in part sponsored by the Army Research Laboratory and was accomplished under Cooperative Agreement Number W911NF-09-2-0053 (Network Science CTA). The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Army Research Laboratory or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation here on.

## References

1. 3-1-1. <https://en.wikipedia.org/wiki/3-1-1> (Accessed on July 2017)
2. Bao, Y., Xiong, T., Hu, Z.: Pso-mismo modeling strategy for multistep-ahead time series prediction. *IEEE Transactions on Cybernetics* pp. 655–668 (2014)
3. Carroll, J.D., Chang, J.J.: Analysis of individual differences in multidimensional scaling via an n-way generalization of "eckart-young" decomposition. *Psychometrika* 35(3), 283–319 (1970)
4. Carroll, J.D., Pruzansky, S., Kruskal, J.B.: Candelinc: A general approach to multidimensional analysis of many-way arrays with linear constraints on parameters. *Psychometrika* 45(1), 3–24 (1980)
5. Chawla, S., Zheng, Y., Hu, J.: Inferring the root cause in road traffic anomalies. In: *ICDM*. pp. 141–150. *IEEE* (2012)
6. Doan, M.T., Rajasegarar, S., Salehi, M., Moshtaghi, M., Leckie, C.: Profiling pedestrian distribution and anomaly detection in a dynamic environment. In: *CIKM*. pp. 1827–1830. *ACM* (2015)
7. Dong, Y., Pinelli, F., Gkoufas, Y., Nabi, Z., Calabrese, F., Chawla, N.V.: Inferring unusual crowd events from mobile phone call detail records. In: *Machine Learning and Knowledge Discovery in Databases - European Conference, ECML PKDD 2015, Proceedings, Part II*. pp. 474–492 (2015)
8. Esling, P., Agon, C.: Time-series data mining. *ACM Computing Surveys (CSUR)* p. 12 (2012)
9. Faber, N.K.M., Bro, R., Hopke, P.K.: Recent developments in candecom/parafac algorithms: a critical review. *Chemometrics and Intelligent Laboratory Systems* 65(1), 119–137 (2003)

10. Gelman, A., Hwang, J., Vehtari, A.: Understanding predictive information criteria for bayesian models. *Statistics and Computing* pp. 997–1016 (2014)
11. Hamilton, J.D.: *Time series analysis*, vol. 2. Princeton university press Princeton (1994)
12. Harshman, R.A.: Foundations of the parafac procedure: Models and conditions for an” explanatory” multi-modal factor analysis (1970)
13. Hosmer Jr, D.W., Lemeshow, S., Sturdivant, R.X.: *Applied logistic regression*, vol. 398. John Wiley & Sons (2013)
14. Hsieh, H.P., Lin, S.D., Zheng, Y.: Inferring air quality for station location recommendation based on urban big data. In: *KDD*. pp. 437–446. ACM (2015)
15. Huang, C., Wu, X., Wang, D.: Crowdsourcing-based urban anomaly prediction system for smart cities. In: *CIKM*. ACM (2016)
16. Kolda, T.G., Bader, B.W.: Tensor decompositions and applications. *SIAM review* 51(3), 455–500 (2009)
17. Koller, D., Friedman, N.: *Probabilistic graphical models: principles and techniques*. MIT press (2009)
18. Le, V.D., Scholten, H., Havinga, P.: Flead: Online frequency likelihood estimation anomaly detection for mobile sensing. In: *UbiComp*. pp. 1159–1166. ACM (2013)
19. Lee, J.Y., Kang, U., Koutra, D., Faloutsos, C.: Fast anomaly detection despite the duplicates. In: *WWW*. pp. 195–196. ACM (2013)
20. Lee, W., Lee, Y., Kim, H., Moon, I.C.: Bayesian nonparametric collaborative topic poisson factorization for electronic health records-based phenotyping. In: *IJCAI* (2016)
21. Lian, D., Ge, Y., Yuan, N.J., Xie, X., Xiong, H.: Sparse bayesian content-aware collaborative filtering for implicit feedback. In: *IJCAI* (2016)
22. Lv, Y., Duan, Y., Kang, W., Li, Z., Wang, F.Y.: Traffic flow prediction with big data: a deep learning approach. *IEEE Transactions on Intelligent Transportation Systems* 16(2), 865–873 (2015)
23. Pan, B., Zheng, Y., Wilkie, D., Shahabi, C.: Crowd sensing of traffic anomalies based on human mobility and social media. In: *SIGSPATIAL*. pp. 344–353. ACM (2013)
24. Resnik, P., Hardisty, E.: Gibbs sampling for the uninitiated. Tech. rep., DTIC Document (2010)
25. Vallis, O., Hochenbaum, J., Kejariwal, A.: A novel technique for long-term anomaly detection in the cloud. In: *6th USENIX Workshop on Hot Topics in Cloud Computing (HotCloud 14)* (2014)
26. Wang, Y.: On efficiency properties of an r-square coefficient based on final prediction error. *Statistics & Probability Letters* 83(10), 2276–2281 (2013)
27. Watanabe, S.: A widely applicable bayesian information criterion. *Journal of Machine Learning Research* 14(Mar), 867–897 (2013)
28. Wiesel, A., Bibi, O., Globerson, A.: Time varying autoregressive moving average models for covariance estimation. *IEEE Transactions on Signal Processing* pp. 2791–2801 (2013)
29. Wikipedia: <https://en.wikipedia.org/wiki/3-1-1#availability> (2017)
30. Yang, Z., Tang, J., Cohen, W.: Multi-modal bayesian embeddings for learning social knowledge graphs. In: *IJCAI* (2016)
31. Zheng, Y., Liu, T., Wang, Y., Zhu, Y., Liu, Y., Chang, E.: Diagnosing new york city’s noises with ubiquitous data. In: *UbiComp*. pp. 715–725. ACM (2014)
32. Zheng, Y., Zhang, H., Yu, Y.: Detecting collective anomalies from multiple spatio-temporal datasets across different domains. In: *SIGSPATIAL*. ACM (2015)