CityOutlook: Early Crowd Dynamics Forecast towards Irregular Events Detection with Synthetically Unbiased Regression

Soto Anno
Tokyo Institute of Technology
Tokyo, Japan
anno@miubiq.cs.titech.ac.jp

Kota Tsubouchi
Yahoo Japan Corporation
Tokyo, Japan
ktsubouc@yahoo-corp.jp

Masamichi Shimosaka
Tokyo Institute of Technology
Tokyo, Japan
simosaka@miubiq.cs.titech.ac.jp

ABSTRACT

Early crowd dynamics forecasting, such as one week in advance, plays an important role in risk-aware decision-making in urban regions such as congestion mitigation or crowd control for public safety. Although previous approaches have addressed crowd dynamics prediction, they have failed to deal with the scarcity of anomalous events, which results in a large model bias and could not quantify the number of visitors in anomalous crowd gathering. To provide an elaborate early forecast, we focus on the successive properties of importance weighting (IW) to penalize the anomalous data in terms of model bias; however, leveraging the concept of IW is challenging because dividing dataset into normal and abnormal sets is difficult. Motivated by these challenges, we propose *CityOutlook*, a novel forecasting model based on unbiased regression with importance-based reweighting. To make IW applicable to our approach, we design an anomaly-aware data annotation scheme by utilizing the heterogeneous property of mobility data to determine the data anomaly. We evaluate CityOutlook using the datasets of large-scale mobility and transit search logs. The experimental results show that CityOutlook outperforms the state-of-the-art models on crowd anomaly forecast, providing the same level accuracy in forecasting normal dynamics.

CCS CONCEPTS

• **Information systems** → *Information systems applications*; • **Human-centered computing** → *Empirical studies in ubiquitous and mobile computing*;

KEYWORDS

crowd dynamics, urban computing, mobility logs, transit search logs, imbalanced learning

ACM Reference Format:
Soto Anno, Kota Tsubouchi, and Masamichi Shimosaka. 2021. CityOutlook: Early Crowd Dynamics Forecast towards Irregular Events Detection with Synthetically Unbiased Regression. In *29th International Conference on Advances in Geographic Information Systems (SIGSPATIAL '21), November 2–5, 2021, Beijing, China.* ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3474717.3483945

SIGSPATIAL '21, November 2–5, 2021, Beijing, China

ACM ISBN 978-1-4503-8664-7/21/11.
https://doi.org/10.1145/3474717.3483945

1 INTRODUCTION

This study explores early forecast of crowd dynamics, i.e., one week in advance, to detect regular and irregular events, and enable appropriate countermeasures for anomalous people movements. Crowd dynamics, i.e., the crowd density changes over the time, considerably increases during unusual events, which has a tremendous threat to public safety (e.g., the New Year's celebration at Shanghai in 2014). Forecasting crowd dynamics in the early stage is of great value to congestion mitigation or crowd control in anomalous people gatherings [11]; however, the task of forecasting crowd dynamics becomes much more difficult when it comes to both the normal dynamics (i.e., daily patterns of density changes) and abnormal dynamics (i.e., changes under irregular events).

Owing to the recent affluence of data (e.g., GPS-based mobility logs), analyzing and forecasting crowd dynamics in a city has been intensively studied [9]. In terms of crowd anomaly forecast, simulating the crowd flows using regressive models in an online learning manner is one of the prominent methods [5]; however, these approaches cannot provide long-term predictions (e.g., one week ahead) because the crowd flow starts to change only just before the anomalies. In contrast, given the fact that people's behavioral schedules reflect future human mobility patterns, empowering the early forecast with people's schedule patterns using additional data (e.g., searching histories of train transit) has also been explored [2, 6].

However, as yet, there are no methods to forecast in advance the number of people visiting unusually, because the existing methods suffer from the rarity of anomalous events, and consequently, the problem of data imbalance. Since events that cause abnormal congestion (e.g., fireworks displays, New Year's celebrations) are very infrequent, most of the data becomes normal and the number of anomalous records is limited. This results in a significant estimator bias when we consider a regression model of anomalous crowd dynamics; that is, while it well represents normal patterns of dynamics, the model cannot fit to anomalous data. A similar problem has discussed in the literature of cost-sensitive learning [4]. However, determining the criteria that indicate the relevance of anomalous data in learning the regression model is still challenging.

Although density ratio-based importance, which is commonly used in covariate shift adaptation [7], reasonably elucidates the relevance of anomalies, employing it for the crowd dynamics forecast is non-trivial. Importance weighting (IW) involves reweighting the loss function to be minimized (e.g., ordinary least squared loss) and provides an unbiased estimate for the distribution of anomalous data. In applying IW to the crowd dynamics forecast, we encounter the difficulty: To build the importance, the set of schedule patterns (as inputs of the regression model) should be divided into normal and abnormal sets. However, unlike the problem settings of covariate shift adaptation and outlier detection, dividing the input set is challenging because whether the schedule patterns are regular or irregular complicatedly depends on the contextual information (e.g., weekday or weekend, schedules for one week ahead or 10 days ahead).

Motivated by these challenges, we propose *CityOutlook*, a novel regression approach with importance-based reweighting of anomalous data. Specifically, the model regresses the irregularity score of crowd dynamics from people's schedule patterns and contextual information such as time, weather, and weekday or weekend. We leverage the concept of IW to penalize the loss to reduce the estimation bias for anomalous crowd dynamics. To tackle the difficulty in estimating the importance, we design an anomaly-aware data annotation scheme by utilizing the heterogeneous property of mobility data. This schema can generate the quantified relevance of anoamlity without depending on the contextual information. Consequently, the model can learn the anomalous patterns effectively.

The contributions of this work are summarized as follows: 1) We explore the problem of crowd dynamics forecast from the viewpoint of a regression problem with imbalanced data in an attempt to provide the effective criteria for the relevance of anomalies. 2) We propose a novel regression framework with importance estimation-based reweighting, called *CityOutlook*, to robustly model both normal and abnormal crowd dynamics. 3) We evaluate the proposed method on a large-scale real dataset. The experimental results and several case studies on real events demonstrate that the proposed model outperforms the baselines on abnormal dynamics, while providing the same level of accuracy in forecasting normal dynamics.

2 PROBLEM SETTING AND BASELINE

2.1 Problem Setting

Let t be a time segment on a day, and each day be divided into T time segments (i.e., $t = 1, 2, ..., T$). In addition, l denotes the point of interest (POI), which is a certain urban region on which we are focusing. We consider the number of mobility records in a certain area in a certain time segment as the ground-truth crowd density. The crowd density observed at the POI l on the date d and time segment t is denoted by $y^{(l)}_{d,t}$. To capture the crowd dynamics in the future, we define the scheduled crowd density. Especially, we use the transit search logs to obtain the scheduled crowd density patterns. The future activity patterns are described by the number of transit search queries as $s^{(l)}_{d,t|d'}$ with the scheduled date d and time t, searching date d', and destination POI l. In general, d' denotes the prior to d (i.e., $s^{(l)}_{d,t|d-i}$ denotes the number of logs searched i days before the scheduled date d). Finally, we define the scheduled crowd density set $\mathcal{S}^{(l)}_{d,t}$ for date d and time t at a destination POI l as $\mathcal{S}^{(l)}_{d,t} = \{s^{(l)}_{d,t|d-i} \mid i = p_d,\ p_d + 1,\ \dots,\ p_d + p_w\}$, where p_d is the earliest day before the scheduled date, and p_w denotes the range of days considered.

Under anomalous crowds, the irregularity score should be large to reflect the high degree of congestion. In contrast, it should be close to zero if no abnormal crowds occur. Based on this notion, we define the crowd dynamics irregularity score $v^{(l)}_{d,t}$ to represent the deviation of the ground-truth crowd density $y^{(l)}_{d,t}$ from the normal dynamics $\bar{y}^{(l)}_{d,t}$ as $v^{(l)}_{d,t} = (y^{(l)}_{d,t} - \bar{y}^{(l)}_{d,t})/\bar{y}^{(l)}_{d,t}$, which is forecasted by the scheduled crowd density and the contextual information such as time, day, and the weather.

2.2 Baseline: Supervised-CityProphet

Drawing on the previous work [2], we design a predictive model of the irregularity score $\hat{v}$. In this model, the crowd anomaly is forecasted by associating the mobility logs and schedule patterns by transit search logs. We build a regression function $\hat{v}^{(l)}_{d,t} = f(\xi^{(l)}_{d,t}; \boldsymbol{\theta})$, where $\boldsymbol{\theta}$ is the learning parameter. $\xi^{(l)}_{d,t} \in \mathbb{R}^{p_w}$ is the schedule deviation score calculated by the scheduled crowd density set $\mathcal{S}^{(l)}_{d,t}$, and the normal scheduled crowd density $\bar{s}^{(l)}_{d,t}$. This is expressed as $\xi^{(l)}_{d,t} = \{\xi^{(l)}_{d,t-j|d-i} \mid \xi^{(l)}_{d,t-j|d-i} = (s^{(l)}_{d,t-j|d-i} - \bar{s}^{(l)}_{d,t})/\bar{s}^{(l)}_{d,t}\}$, where $s^{(l)}_{d,t|d-i} \in \mathcal{S}^{(l)}_{d,t-j}$, and $j = -1, 0, 1$. Based on the defined terms, we formulate the irregularity prediction model $f(\xi^{(l)}_{d,t}; \boldsymbol{\theta})$ using an autoregressive model [1] as $\hat{v}^{(l)}_{d,t} = f(\xi^{(l)}_{d,t}; \boldsymbol{\theta}) = [1, {\xi^{(l)}_{d,t}}^{\top}]\boldsymbol{\theta}$, where $\boldsymbol{\theta} \in \mathbb{R}^{3p_w+1}$. For estimating the normal crowd dynamics $\bar{y}^{(l)}_{d,t}$ and $\bar{s}^{(l)}_{d,t}$, we internally use bilinear Poisson regression [8], which predicts ordinary crowd dynamics from external contextual factors such as holiday-or-not, weekday-or-weekend, or the weather. A detailed formulation of this model is described in Section ??. To simplify the notation for readability, we omit l, d, and t from the description.

The learning parameters are inferred by minimizing the ordinary least squared (OLS) loss $\mathcal{L}$, as $\min_{\boldsymbol{\theta}\in\Theta}[\frac{1}{N}\sum_n \mathcal{L}(v, f(\xi; \boldsymbol{\theta}))] = \min_{\boldsymbol{\theta}\in\Theta}[\frac{1}{N}\sum_n (v_n - f(\xi_n; \boldsymbol{\theta}))^2]$, where Θ is the parameter space, and N and n denote the number of data and its index, respectively. Note that minimizing OLS loss over the imbalanced dataset increases the estimator bias for the anomalous patterns because the model the model fits a large amount of normal data and tends to ignore rare patterns.

3 PROPOSED METHOD: CITYOUTLOOK

3.1 Bias Reduction by Importance Weighting

To address the problem of large estimator bias, we quantify the relevance of anomalous patterns for reweighting the data. To define the relevance for anomalous crowd dynamics data, we consider the closeness of the normal and abnormal data distributions. A promising measure of the closeness between distributions is importance based on the density ratio. We leverage the basic concept of the density ratio estimation-based IW technique. The importance $w(\xi)$ is defined by considering the density ratio between normal and abnormal data as $w(\xi) = \frac{p(s=1|\xi)}{p(s=0|\xi)}$, where s is the flag of anomalies which takes 1 when ξ is abnormal and 0 otherwise. The importance weighted least squared loss can be minimized as $\min_{\boldsymbol{\theta}\in\Theta}\left[\frac{1}{N}\sum_n \frac{p(s_n=1|\xi)}{p(s_n=0|\xi)}\mathcal{L}(v, f(\xi; \boldsymbol{\theta}))\right]$. Theoretically, the density ratio-based importance-weighted least squared loss provides consistent estimates over the abnormal distribution $p(\xi, s = 1)$, which means the estimator bias is reduced for the anomalous data and the model can learn anomalous patterns effectively. In practice, we use relative importance [10] to prevent learning instability caused by importance explosion and allow the model to learn both normal and abnormal patterns. This is defined as $\tilde{w}(\xi)\frac{p(s=1,\xi)}{\beta p(s=1,\xi)+(1-\beta)p(s=0,\xi)}$, where $\beta \in [0,1]$ is a hyper-parameter.

3.2 Heterogeneous Anomaly-Aware Annotation Scheme

However, setting up the importance is very complicated because the abnormality of the input ξ highly depends on the contextual factors, which results in the fact that there are no explicit anomality labels, and consequently makes it much more challenging to separate the input dataset into normal and abnormal. To address this issue, we propose the heterogeneous anomaly-aware annotation scheme which is tailored for penalizing the data by importance-based relevance and anomalous crowd dynamics learning. This scheme refers to the crowd dynamics irregularity score v which is defined based on the number of mobility logs, and explicitly define the anomality labels for the input ξ. We spuriously separate the input dataset by using the upper bound of the normality $\bar{v}_{\text{thre}}$. The normal input dataset $\mathcal{D}_{no}$ and anomalous input dataset $\mathcal{D}_{ano}$ are defined as $\mathcal{D}_{no} = \{\xi \mid (v,\xi), v < \bar{v}_{\text{thre}}\}$ and $\mathcal{D}_{ano} = \{\xi \mid (v,\xi), \bar{v}_{\text{thre}} \leq v\}$. We estimate the density $p(s=0,\xi)$ and $p(s=1,\xi)$ respectively in a non-parametric manner by using kernel density estimation [3] with a Gaussian kernel as follows:

$$p(s=0,\xi) = \frac{1}{|\mathcal{D}_{no}|}\sum_{\tilde{\xi}_i \in \mathcal{D}_{no}} \frac{1}{(2\pi h^2)^{D/2}} \exp\left\{-\frac{||\xi-\xi_i||^2}{2h^2}\right\}, \quad (1)$$

$$p(s=1,\xi) = \frac{1}{|\mathcal{D}_{ano}|}\sum_{\xi_j \in \mathcal{D}_{ano}} \frac{1}{(2\pi h^2)^{D/2}} \exp\left\{-\frac{||\xi-\xi_j||^2}{2h^2}\right\}, \quad (2)$$

where h denotes the Gaussian kernel width, and $D = 3p_w + 2$. From Bayes' theorem, we can obtain the density ratio without directly observing $p(\xi)$. Based on the estimated density, we calculate the importance as $\tilde{w}(\tilde{\xi}) = \frac{p(s=1,\tilde{\xi})}{\beta p(s=1,\tilde{\xi})+(1-\beta)p(s=0,\tilde{\xi})}$, We use the defined importance $\tilde{w}$ to penalize least squared loss.

3.3 Parameter Learning and Dynamics Forecast

In the learning process of the proposed model, we minimize the importance-weighted least squared loss. The learned parameter $\hat{\boldsymbol{\theta}}$ can be obtained by solving the following optimization problem: $\hat{\boldsymbol{\theta}} = \underset{\boldsymbol{\theta}\in\Theta}{\arg\min}\ \frac{1}{N}\left[\sum_{n=1}^{N} \tilde{w}(\xi_n)\mathcal{L}(v_n, f(\xi_n;\boldsymbol{\theta}))\right] + \gamma||\boldsymbol{\theta}||_2^2$, where $\mathcal{L}$ is the least squared loss, and $\gamma||\boldsymbol{\theta}||_2^2$ is the L2 regularization term with hyper-parameter γ. In the forecasting process, the crowd density $\hat{y}_{d,t}^{(l)}$ is rebate from the inferred irregularity score $\hat{v}_{d,t}^{(l)}$ as $\hat{y}_{d,t}^{(l)} = (1+\hat{v}_{d,t}^{(l)})\bar{y}_{d,t}^{(l)}$, where $\bar{y}_{d,t}^{(l)}$ is the normal dynamics discussed in Section 2.

4 EXPERIMENTS

4.1 Dataset

The mobility logs were collected via a disaster alert mobile application[1] from Yahoo! JAPAN by masking user IDs with dummies. Each record was completely anonymized, and characterized by timestamp, latitude, and longitude. We aggregated the mobility logs in the POIs at each time segment, and counted their number as crowd dynamics. For the scheduled crowd dynamics, we also utilized transit search history data, which were searched by passengers of train, bus, or taxi. These logs are gathered by the transit search engine[2], also released by Yahoo! JAPAN. Each record contains an anonymized user ID, searching timestamp, scheduled timestamp, and destination. Similar to the mobility logs, we added the number of search records per stations and time segment; therefore, we did not use any personal information for model learning. We utilized the data collected over six months (from October 1, 2019, to March 31, 2020). We selected 58 specific square areas as POIs and their corresponding stations. POIs contain not only the Greater Tokyo Area (where many people are observed daily), but also stadiums, shrines, and venues of the fireworks displays. The size of each POI was set to $600 \times 600m^2$.

4.2 Experimental Setups and Baselines

We consider one day as a 24-h period, and the number of time segments T is set to 24. Following previous research [8], the start of a day was 3:00 AM, which had the least active population, and the end was 3:00 AM the next day (i.e. 27:00 in 24-h notation).

As mentioned in the Section 2, we used bilinear Poisson regression [8] to predict normal crowd dynamics. In this model, the crowd density $y_{d,t}$ is assumed to follow a Poisson distribution as $y_{d,t} \sim \text{Pois}(\cdot|\lambda_{\mathbf{c}_d,t})$, where $\lambda_{\mathbf{c}_d,t}$ denotes the parameter of the Poisson distribution. We model $\lambda_{\mathbf{c}_d,t}$ with parameter matrix $\boldsymbol{W} \in \mathbb{R}^{C\times T}$ by using the context $\mathbf{c}_d$ of date d and time segment t as $\ln \lambda_{\mathbf{c}_d,t} = \boldsymbol{\varphi}(\mathbf{c}_d)^\top \boldsymbol{W}\boldsymbol{\phi}(t)$, where $\boldsymbol{\varphi}(\mathbf{c}) \in \mathbb{R}^C$ is the external factor vector expressing the contexts $\mathbf{c}$ by the one-hot encoding method, and $\boldsymbol{\phi}(t) \in \mathbb{R}^T$ is a time factor denoted by $\boldsymbol{\phi}(t) = \{t_s|t_s = \mathcal{N}(s|\tau,\sigma^2), s=1,\ldots,T\}$. $\mathcal{N}(\cdot)$ is a Gaussian distribution with mean τ and variance σ^2. For the context denoted by $\mathbf{c}_d$, we used holiday-or-not, weekday-or-weekend, and weather information in one-hot encoding. Weather information is a four-dimensional vector: sunny, cloudy, rainy, and the others. We used the tensor product to compose these features into one input vector. For the regularization term, we set $\gamma = 0.01$. For the hyper-parameter settings, we set $\beta = 0.1$, $\bar{v}_{\text{thre}} = 6.0$, $h = 5.0$.

We evaluated the performance for two types of predictions: **irregularity score forecast** and **crowd density forecast**. To evaluate the performance fairly, we conducted five-fold cross validation. On the each round of the cross validation, we divided the dataset into 88 days for training and 22 days for testing.

To evaluate the performance of our model, we adopted a mean absolute error (MAE) conditioned by the anomaly score-based threshold $\bar{v}$ in an attempt to measure the prediction performance on

[1] http://emg.yahoo.co.jp/
[2] https://transit.yahoo.co.jp/

Table 1: Performance comparison for forecasting one week in advance on 58 POIs across different thresholds.

Model	NS-MAE						AS-MAE					
	$\bar{\nu} = 10.0$		$\bar{\nu} = 15.0$		$\bar{\nu} = 20.0$		$\bar{\nu} = 10.0$		$\bar{\nu} = 15.0$		$\bar{\nu} = 20.0$	
	Score	Density	Score	Density	Score	Density	Score	Density	Score	Density	Score	Density
CityProphet [6]	3.684	210.135	3.684	210.488	3.694	210.586	25.749	234.710	29.530	165.390	33.751	129.372
SCP (Baseline) [2]	**0.566**	**91.951**	**0.570**	**92.004**	**0.575**	**92.060**	15.695	109.380	23.871	144.344	34.063	140.510
CityOutlook	0.796	93.881	0.801	93.921	0.805	93.968	**14.482**	**102.163**	**23.132**	**138.420**	**32.698**	**132.301**

both normal dynamics and anomalous crowd gathering forecasting. For the irregularity score forecast, we adopt normal sample (NS)-MAE and anomalous sample (AS)-MAE defined as NS-MAE = $\frac{1}{DT}\sum_{d=1}^{D}\sum_{t=1}^{T}|v_{d,t}^{(l)} - \hat{v}_{d,t}^{(l)}|$, where $v_{d,t}^{(l)} < \bar{\nu}$, and AS-MAE = $\frac{1}{DT}\sum_{d=1}^{D}\sum_{t=1}^{T}|v_{d,t}^{(l)} - \hat{v}_{d,t}^{(l)}|$, where $v_{d,t}^{(l)} \geq \bar{\nu}$, where D is the number of days. Similarly, for the crowd dynamics forecasting, we used NS-MAE = $\frac{1}{DT}\sum_{d=1}^{D}\sum_{t=1}^{T}|y_{d,t}^{(l)} - \hat{y}_{d,t}^{(l)}|$ where $v_{d,t}^{(l)} < \bar{\nu}$, and AS-MAE = $\frac{1}{DT}\sum_{d=1}^{D}\sum_{t=1}^{T}|y_{d,t}^{(l)} - \hat{y}_{d,t}^{(l)}|$ where $v_{d,t}^{(l)} \geq \bar{\nu}$, where $\hat{y}_{d,t}^{(l)}$ is the predicted urban dynamics at POI l on date d and time t and $y_{d,t}^{(l)}$ is the ground truth. Note that, if NS-MAE is evaluated with a small threshold $\bar{\nu}$, this means that the performance is evaluated only for normal dynamics observed on a daily basis. Furthermore, if the AS-MAE is evaluated with a large threshold $\bar{\nu}$, this means that performance is assessed on the exceptional anomalous crowd gathering, which is a significant deviation from daily life.

CityProphet [6] and Supervised-CityProphet (**SCP**) [2] were used as comparative models. CityProphet is a method that uses context information and scheduled crowd dynamics as input data. In CityProphet, two prediction models, schedule-based population (SP) and descriptor-based population (DP), are proposed to predict the number of search queries, and the anomaly score is computed based on the comparison between the two models. Supervised-CityProphet is discussed in Section 2.

4.3 Experimental Results

Table 1 shows the overall evaluation in irregularity score forecast (Score) and crowd density forecast (density). The results show that the proposed method CityOutlook achieves the best performance in anomalous crowd dynamics forecasts compared to the baseline approaches, simultaneously providing the same level of accurate forecasting in normal dynamics as SCP. From the results of the irregularity score forecast in AS-MAE, the proposed method outperforms the baseline model SCP [2] by 7.7% for $\bar{\nu} = 10.0$, 3.0% for $\bar{\nu} = 15.0$, and 4.0% for $\bar{\nu} = 20.0$. From the results of crowd dynamics forecast in AS-MAE, the proposed method provides a reduction of 6.6% on $\bar{\nu} = 10.0$, 4.1% on $\bar{\nu} = 15.0$, and 5.8% on $\bar{\nu} = 20.0$ relative to SCP. On the contrary, baselines cannot forecast normal dynamics and abnormal dynamics simultaneously. We can confirm that SCP also has severe performance drawbacks in that forecasting performance deteriorates in AS-MAE, in contrast to the performance in NS-MAE. This indicates that SCP was not able to capture the anomaly patterns well and sticking to the ordinary forecast output.

The evaluation of CityOutlook compared to SCP indicates that IW produces effective measures for anomalies, while ensuring the performance in normal dynamics forecast. The proposed method achieved a dramatic performance improvement in forecasting anomalous crowd dynamics, i.e., 8.20 improvement in MAE, in contrast to a small inferiority in performance as in normal dynamics forecast, i.e., 1.90 difference in MAE, on the threshold of $\bar{\nu} = 20.0$. The numerical degradation in the performance of normal dynamics forecasting in comparison with SCP is not problematic when the forecasting system is deployed to the real-world scenario, because it is a small error in prediction in a situation where there is no risk of an accident due to a crowding.

5 CONCLUSION

In this paper, we proposed CityOutlook for early crowd dynamics forecast one week in advance. Compared with the recent advances in forecasting systems based on mobility logs and people's schedule patterns, the proposed method provides an effective learning strategy of anomalies, addressing the problem of data imbalance and scarcity of anomalies by the importance-based reweighting with anomaly-aware anotation scheme tailored with heterogeneous data. The experimental results on massive real datasets demonstrate the superiority of our model over the existing methods. Our approach shows better results in the forecasting performance of an anomaly, outperforming the baseline approach SCP by 6.6% on crowd anomaly forecasts, while providing accurate normal crowd dynamics forecasts at the same level as comparative models. For our future work, forecasting in the typical environments, such as subways or rural areas, could be addressed.

REFERENCES

[1] Hirotugu Akaike. 1969. Fitting autoregressive models for prediction. *Annals of the institute of Statistical Mathematics* (1969).
[2] Soto Anno et al. 2020. Supervised-CityProphet: Towards Accurate Anomalous Crowd Prediction. In *Proc. of SIGSPATIAL*.
[3] Yen-Chi Chen. 2017. A tutorial on kernel density estimation and recent advances. *Biostatistics & Epidemiology* (2017).
[4] Charles Elkan. 2001. The Foundations of Cost-Sensitive Learning. In *Proc. IJCAI*.
[5] Renhe Jiang et al. 2019. DeepUrbanEvent: A System for Predicting Citywide Crowd Dynamics at Big Events. In *Proc. of SIGKDD*.
[6] Tatsuya Konishi et al. 2016. CityProphet: City-Scale Irregularity Prediction Using Transit App Logs. In *Proc. of UbiComp*.
[7] Hidetoshi Shimodaira. 2000. Improving predictive inference under covariate shift by weighting the log-likelihood function. *Journal of statistical planning and inference* (2000).
[8] Masamichi Shimosaka et al. 2015. Forecasting Urban Dynamics with Mobility Logs by Bilinear Poisson Regression. In *Proc. of UbiComp*.
[9] Tong Xia and Yong Li. 2019. Revealing Urban Dynamics by Learning Online and Offline Behaviours Together. *Proc. of IMWUT* (2019).
[10] Makoto Yamada et al. 2011. Relative Density-Ratio Estimation for Robust Distribution Comparison. In *Advances in NIPS*.
[11] Yu Zheng et al. 2014. Urban Computing: Concepts, Methodologies, and Applications. *ACM Trans. on Intelligent Systems and Technology* (2014).