Article Type: Description (see Introduction for more detail)

CityOutlook+: Early Crowd Dynamics Forecast through Unbiased Regression with Importance-based Synthetic Oversampling

Soto Anno, *Tokyo Institute of Technology, Tokyo, Japan*

Kota Tsubouchi, *Yahoo Japan Corporation, Tokyo, Japan*

Masamichi Shimosaka, *Tokyo Institute of Technology, Tokyo, Japan*

Abstract—This article studies crowd dynamics forecast one week in advance to detect irregular urban events, which plays an important role in infection prevention and crowd control. Previous approaches have failed to deal with the scarcity of anomalous events, resulting in a large model bias, and could not quantify the number of visitors in anomalous crowding. We proposed an unbiased regression using importance weighting (IW), called CityOutlook [1], and successfully reduced the model bias and showed promising results. However, the straightforward weighting of the scarce data risks leading to the instability of the model due to the increase in model variance. To address this issue, we propose a non-trivial extension of our prior work called CityOutlook+ that realizes unbiased and less-variant regression by performing synthetic minority oversampling based on the importance. We evaluate CityOutlook+ using real datasets and demonstrate the superiority of our model to CityOutlook and state-of-the-art approaches.

This study forecasts crowd dynamics one week in advance to detect regular and irregular events and counter anomalous people movements. Crowd dynamics, i.e., the crowd density changes over time, significantly increase during unusual events and pose a tremendous threat to public safety (e.g., accidents or epidemics due to surging crowds). Early forecasting of crowd dynamics enables us to facilitate strategic planning for infection prevention and crowd control, such as allocating personnel for crowd management and medical resources. However, forecasting becomes challenging when it comes to both the normal dynamics (i.e., daily changes) and abnormal dynamics (i.e., changes under irregular events).

The rise of pervasive and ubiquitous computing, characterized by the seamless incorporation of technology into everyday life, is vividly demonstrated in our use of GPS-based mobility logs. These logs enable real-time analysis of crowd dynamics [2], [3], and simulating the crowd flows using regressive models in an online learning manner is one of the prominent methods [4] for anomalous crowds forecast; 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. Alternatively, given 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 [5], [6].

However, as yet, there are no methods to successfully 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. Fig. 1 shows the crowd dynamics over several days and its degree of irregularity, which we call the irregularity score. Most of the data become normal, and the number of anomalous records is limited. This leads to significant estimator bias in a regression model of anomalous crowd dynamics, as the model cannot fit to anomalous data, although it well represents normal patterns of dynamics.

Our prior work [1] proposed CityOutlook to over-

Digital Object Identifier 10.1109/XXX.0000.0000000

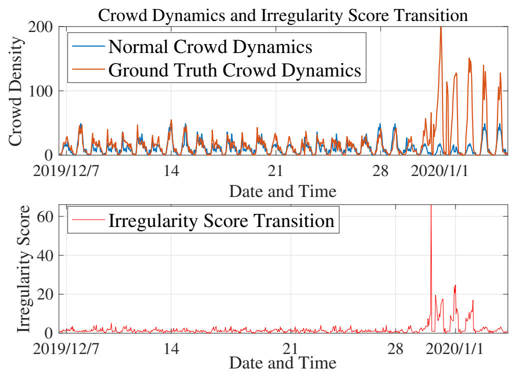


FIGURE 1: Crowd dynamics and irregularity scores in Meiji Jingu Shrine.

come the challenges and limitations of related work by using density ratio-based importance weighting (IW) [7] for an unbiased estimate of the distribution of anomalous data. However, weighting the anomalous data causes prediction instability due to the model becoming more sensitive to the noise in the anomaly data, leading to overfitting and increased model variance. Solving this problem is inherently challenging due to the trade-off between bias and variance.

Motivated by this challenge, we propose a non-trivial extension of our previous work, **CityOutlook+**, which uses synthetic sampling of data to achieve unbiased and *less-variant* regression. We combine the IW with synthetic minority oversampling (SMOTE) [8] to reduce model variance. Although SMOTE and its family [9] aim to suppress overfitting to minority, i.e., scarce data by synthesizing internally dividing points of existing data, a way of theoretically determining the number of sampled data to reduce the bias has not been established. We address this issue by developing a new resampling algorithm that calculates the number of samples based on the importance.

Our contributions are summarized as follows:

- We extend our prior work of unbiased regression [1] to realize both unbiased and *less-variant* prediction for early crowd dynamics forecast.
- We propose a novel regression framework with importance estimation-based resampling, called *CityOutlook+*, to robustly model both normal and abnormal crowd dynamics.
- The proposed method is evaluated on a large-scale real dataset, and experimental results and case studies demonstrate that it outperforms CityOutlook and state-of-the-art approaches in predicting abnormal crowd dynamics while maintaining accuracy in forecasting normal crowd dynamics.

Related Work

Crowd Dynamics Forecast. Mobile device-based location history has facilitated the forecasting of crowd dynamics. Fan et al. [10] and Jiang et al. [4] proposed online learning-based systems using human-mobility logs to predict crowd flow, but they lacked long-term prediction of anomalies. Researchers explored using schedule-based features like transit search applications [5], [6]; however, due to the rarity of anomalous events, these methods cannot provide accurate early forecasts (e.g., one week ahead).

Imbalanced Learning tackles learning patterns from imbalanced data [11] and is extensively researched in various machine learning fields. Resampling [8] and cost-sensitive learning [12] approaches are used for classification problems. A few studies focused on regression problems. Prior studies have proposed sigmoid-like relevance [13] and extended SMOTE [8] method for regression. However, these methods are unsuitable for forecasting normal and anomalous crowd dynamics due to the limited applicability of the output probability distribution.

Importance Weighting (IW) [7], [14], which has been conventionally used for penalizing the loss function under covariate shift [15], is formulated as a shift in the distribution of explanatory variables between source and target data. The importance is used to reweight source data as an unbiased estimate for target data functions [7]. However, dividing the dataset makes it difficult to apply IW to the regression task, and penalizing scarce patterns increases the variance of the estimator, which is problematic.

Our prior work [1] was established to address these issues, and provided an effective measure for the relevance of anomalies by applying IW to the crowd dynamics modeling.

Preliminaries

Variable Definition and Problem Statement

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.

Definition 1 (**Ground-Truth Crowd Density**): GPS-based mobility logs are used to address crowd density patterns. The number of mobility records in an area at a specific time is the ground-truth crowd density. The observed crowd density at a POI l on the date d and time segment t is denoted by $y^{(l)}_{d,t}$.

Definition 2 (**Scheduled Crowd Density Set**): The study defines the scheduled crowd density for future crowd dynamics and uses transit search logs to obtain the scheduled crowd density patterns. The transit search history consists of query records with scheduled date d, time t, searching date d', and destination POI l. $s^{(l)}_{d,t|d-i}$ denotes the number of logs for scheduled date d and time t, searched i days before the date d. Then, we define the scheduled crowd density set $\mathcal{S}^{(l)}_{d,t}$ as $\mathcal{S}^{(l)}_{d,t} = \{s^{(l)}_{d,t|d-i} \mid i = p_d,\ p_d+1,\ ...,\ p_d+p_w\}$, where p_d is the earliest day before the scheduled date, and p_w denotes the range of days.

Definition 3 (**Crowd Dynamics Irregularity Score**): We define *irregular crowd dynamics* as abnormally high crowd density. Hence, the irregularity score should be large for anomalous crowds to reflect the high degree of congestion and close to zero for no crowds. The crowd dynamics irregularity score $\nu^{(l)}_{d,t}$ represents the deviation of ground-truth crowd density $y^{(l)}_{d,t}$ from normal dynamics $\bar{y}^{(l)}_{d,t}$ as $\nu^{(l)}_{d,t} = (y^{(l)}_{d,t} - \bar{y}^{(l)}_{d,t})/\bar{y}^{(l)}_{d,t}$.

Definition 4 (**Early Crowd Dynamics Forecast**): One week before events, using scheduled crowd density $\mathcal{S}^{(l)}_{d,t}$, where $p_d = 7$, and normal crowd dynamics $\bar{y}^{(l)}_{d,t}$, we forecast normal and abnormal crowd dynamics at a POI l on date d and time t.

Supervised-CityProphet: Baseline Approach

Drawing on the previous work [6], we design a predictive model of the irregularity score $\hat{\nu}$. 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 $\hat{\nu}^{(l)}_{d,t} = f(\boldsymbol{\xi}^{(l)}_{d,t};\boldsymbol{\theta})$, where $\boldsymbol{\theta}$ is the learning parameter. $\boldsymbol{\xi}^{(l)}_{d,t} \in \mathbb{R}^{p_w}$ 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 follows:

$$\boldsymbol{\xi}^{(l)}_{d,t} = \left\{ \xi^{(l)}_{d,t-j|d-i} \;\middle|\; \xi^{(l)}_{d,t-j|d-i} = \frac{s^{(l)}_{d,t-j|d-i} - \bar{s}^{(l)}_{d,t}}{\bar{s}^{(l)}_{d,t}} \right\}, \quad (1)$$

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 $f(\boldsymbol{\xi}^{(l)}_{d,t};\boldsymbol{\theta})$ using an autoregressive model [16] as follows:

$$\begin{aligned}\hat{\nu}^{(l)}_{d,t} &= f(\boldsymbol{\xi}^{(l)}_{d,t};\boldsymbol{\theta}) \\ &= [1, {\boldsymbol{\xi}^{(l)}_{d,t}}^{\top}]\boldsymbol{\theta} = \sum_{i=p_d}^{p_d+p_w} \sum_{j=-1}^{1} \theta_{i,j}\xi^{(l)}_{d,t-j|d-i} + \theta_c, \qquad (2)\end{aligned}$$

where $\boldsymbol{\theta} \in \mathbb{R}^{3p_w+1}$, whose elements are denoted by $\theta_{i,j}$ and θ_c. 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, $\mathcal{L}(\nu, f(\boldsymbol{\xi};\boldsymbol{\theta}))$, as follows:

$$\min_{\theta\in\Theta}\left[\frac{1}{N}\sum_n \mathcal{L}(\nu, f(\boldsymbol{\xi};\boldsymbol{\theta}))\right] = \min_{\theta\in\Theta}\left[\frac{1}{N}\sum_n (\nu_n - f(\boldsymbol{\xi}_n;\boldsymbol{\theta}))^2\right], \quad (3)$$

where Θ is the parameter space, and N and n denote the number of data and its index, respectively. Bilinear Poisson regression [3] is used to estimate normal crowd dynamics $\bar{y}^{(l)}_{d,t}$ and $\bar{s}^{(l)}_{d,t}$ by predicting it from contextual factors such as holidays, weekdays/weekends, and weather. This model is described in detail in the experiments section.

However, as discussed in [1], minimizing OLS loss fails to robustly capture anomalous crowd dynamics patterns This is because minimizing OLS loss in Eq. (3) can be regarded as learning normal crowd dynamics, and there is still a significant bias in the regression model on the abnormal patterns. For this reason, we consider defining the criteria for the relevance of data and penalizing the OLS loss of each training data.

CityOutlook+: Proposed Method

In this section, we present our proposed method CityOutlook+, which is an extension of CityOutlook leveraging the concept of synthetic minority oversampling (SMOTE). Firstly, we review the research challenges of the baseline approach and illustrate the framework overview of the proposed method. Then, we describe data pre-processing and the importance-weighting-based unbiased regression proposed in our prior work [1]. Finally, we extend our prior work for realizing the unbiased and less-variant regression, i.e., CityOutlook+.

Research Challenges and Basic Concept

We focus on IW as a powerful indicator to penalize the loss; However, it is non-trivial to employ the IW technique for two reasons: (1) Dividing the set of input data into normal and abnormal data is challenging because the abnormality of the input $\boldsymbol{\xi}$ complicatedly depends on the contextual factors such as weekday-or-not and schedules for 1-week-ahead or 10-days-ahead. This issue makes it difficult to applying the IW to crowd dynamics forecasting. (2) Straightforward weighting of abnormal data results in model instability owing to the large variance of the estimator. Weighting the loss of the abnormal data makes the model become more sensitive to the inherent data noise. This means that the estimator overfits the data noise and increases the

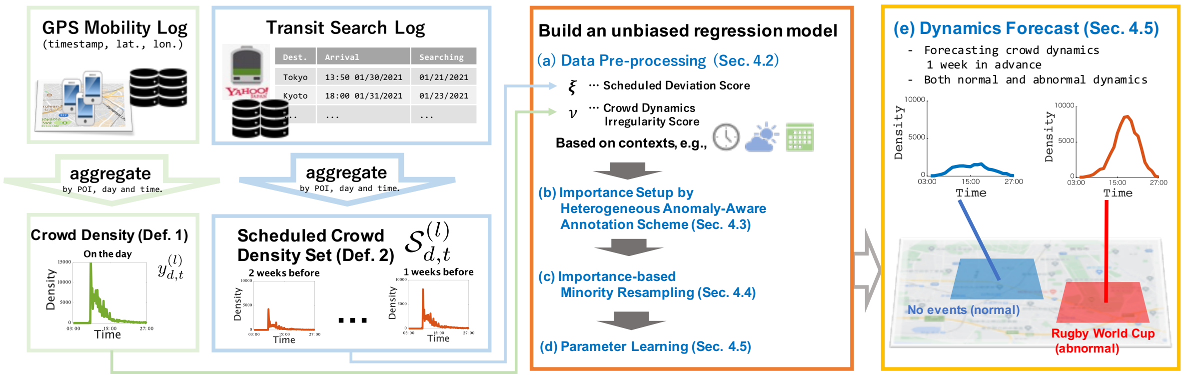


FIGURE 2: Overall framework of CityOutlook+.

variant error on the prediction. Therefore, forecasting a crowd anomaly becomes much more difficult even if the importance is established.

To address the aforementioned issues, we focus on the heterogeneous properties of mobility logs and data augmentation by synthetic minority oversampling [8]. The basic concept of CityOutlook+ is (1) to design a data anomality annotation strategy with the heterogeneous property of mobility data, and (2) to build an importance-based resampling approach to augment data and mitigate the learning instability.

The overall framework of CityOutlook+ is illustrated in Fig. 2. It uses the dataset of schedule deviation score and crowd dynamics irregularity score based on the mobility logs and transit search histories, as discussed in the preliminary section. To build an unbiased regression model, we firstly review the data pre-processing (Fig. 2(a)), describe a heterogeneous anomaly-aware annotation scheme for setting up an importance as an anomaly indicator (Fig. 2(b)), introduce an importance-based minority resampling algorithm (Fig. 2(c)), and finally present parameter learning and dynamics forecast (Fig. 2(d), (e)).

Data Pre-Processing

We preprocess the crowd density $y^{(l)}_{d,t}$ and scheduled crowd density set $\mathcal{S}^{(l)}_{d,t}$ as [6] to obtain (ν, ξ), where ν is the irregularity score of crowd dynamics as defined in *Definition 3*, and schedule deviation score based on (1). To estimate the normal crowd dynamics $\bar{y}^{(l)}_{d,t}$ and $\bar{s}^{(l)}_{d,t}$, we use bilinear Poisson regression [3]. This method assumes normal crowd dynamics can be modeled from external contextual factors and a time factor. For these factors, we used holiday-or-not, weekday-or-weekend, and weather information. Based on one-hot encoding, holiday-or-not, and weekday-or-weekend features are two-dimensional vectors. Weather information is a four-dimensional vector: sunny, cloudy, rainy, and the others. We use the tensor product to compose these features into one input vector and regress the normal density by forming a bilinear representation with a time factor.

Importance Setup by Heterogeneous Anomaly-Aware Annotation Scheme

As discussed in the preliminaries section, minimizing OLS loss suffers from a significant bias on abnormal patterns. To address this issue, we define the relevance of data and penalize the OLS loss. As proposed in our prior work [1], we used the density-ratio-based importance for defining the relevance. With the importance $w(\xi) = \frac{p(s=1|\xi)}{p(s=0|\xi)}$, importance-weighted least squared loss can be minimized as follows:

$$\min_{\theta \in \Theta} \left[\frac{1}{N} \sum_{n} \frac{p(s_n = 1|\xi)}{p(s_n = 0|\xi)} (\nu_n - f(\xi_n; \theta))^2 \right]. \tag{4}$$

In our prior work [1] to calculate the importance effectively, we proposed the heterogeneous anomaly-aware annotation scheme tailored for penalizing the data by importance-based relevance and anomalous crowd dynamics learning. This scheme refers to the crowd dynamics irregularity score ν which is defined based on the number of mobility logs, and explicitly defines the anomaly labels for the input ξ. We spuriously separate the input dataset by using the upper bound of the normality $\bar{\nu}_{\text{thre}}$. The normal input dataset $\mathcal{D}_{no}$ and anomalous input dataset $\mathcal{D}_{ano}$ are defined as follows:

$$\mathcal{D}_{no} = \{\xi \mid (\nu, \xi), \nu < \bar{\nu}_{\text{thre}}\}, \tag{5}$$

$$\mathcal{D}_{ano} = \{\xi \mid (\nu, \xi), \bar{\nu}_{\text{thre}} \leq \nu\}, \tag{6}$$

We estimate the density $p(s = 0, \xi)$ and $p(s = 1, \xi)$ respectively in a non-parametric manner by using kernel density estimation [17] with a Gaussian kernel as follows:

$$p(s=0,\xi) = \frac{1}{|\mathcal{D}_{no}|}\sum_{\tilde{\xi}_i \in \mathcal{D}_{no}} \frac{1}{(2\pi h^2)^{D/2}} \exp\left\{-\frac{||\xi-\xi_i||^2}{2h^2}\right\}, \tag{7}$$

$$p(s=1,\xi) = \frac{1}{|\mathcal{D}_{ano}|}\sum_{\xi_j \in \mathcal{D}_{ano}} \frac{1}{(2\pi h^2)^{D/2}} \exp\left\{-\frac{||\xi-\xi_j||^2}{2h^2}\right\}, \tag{8}$$

where h denotes the Gaussian kernel width, and $D = 3p_w + 2$. In practice, we use relative importance [14] to prevent learning instability caused by importance explosion and allow the model to learn both normal and abnormal patterns. This is defined as follows:

$$\begin{aligned}\tilde{w}(\xi) &= \frac{p(s=1|\xi)}{\beta p(s=1|\xi) + (1-\beta)p(s=0|\xi)} \\ &= \frac{p(s=1,\xi)}{\beta p(s=1,\xi) + (1-\beta)p(s=0,\xi)}.\end{aligned} \tag{9}$$

where $\beta \in [0, 1]$ is a hyper-parameter.

Importance-based Minority Sampling

IW provides an efficient and quantitative learning norm for the data imbalance issue; however, the scarcity of anomalous patterns due to the limited sample size of abnormal data makes it difficult to robustly learn the anomaly. This is because the small sample size of anomalous data results in a large variance of the estimator, especially on the weighted loss function.

We overcome this issue by leveraging another perspective of a data augmentation strategy, called synthetic minority oversampling [8]. In this strategy, the anomalous data is synthetically oversampled by generating a dividing point which divides the data and its neighbors internally. Intuitively, the model learns noise that enhances its representative power through synthetic oversampling-based data augmentation.

However, the sampling criteria for the regression approach have not been discussed in previous studies because they have mainly focused on the resampling of classification tasks, whose oversampling criteria can be easily obtained based on the number of samples belonging to minority classes.

Therefore, we extend the algorithm of synthetic minority oversampling for our regression task with the sampling criteria based on the importance. We present a new sampling algorithm for the regression task in Algorithm 1. We focus on learning a training sample with the importance of 10 is the same as learning the ten identical samples. To implement this principle in the algorithm, the importance is estimated for each training sample (line 11), and the training sample is oversampled by the number of importance (lines 12-26)

Algorithm 1 Importance-based Minority Resampling

Input: $\mathcal{D}, k$

1: // $\mathcal{D}$ - dataset of (ξ, ν)
2: // k - number of nearest neighbors
3: $new\mathcal{D} \leftarrow \{\}$, $\mathcal{D}_{no} \leftarrow \{\}$, $\mathcal{D}_{ano} \leftarrow \{\}$.
4: Divide $\mathcal{D}$ into $\mathcal{D}_{no}$ and $\mathcal{D}_{ano}$.
5: Estimate $p(s=0,\tilde{\xi})$ by Eq. (7).
6: Estimate $p(s=1,\tilde{\xi})$ by Eq. (8).
7: **for all** $(\xi_n, \nu_n) \leftarrow \mathcal{D}$ **do**
8:     $newCases \leftarrow \{\}$
9:     $case \leftarrow (\xi_n, \nu_n)$
10:     // estimate the importance
11:     $\tilde{w}_n \leftarrow \tilde{w}(\xi_n)$ by Eq. (9).
12:     **if** $\tilde{w}_n \geq 2$ **then**
13:         $nns \leftarrow \mathrm{KNN}(k, case, \mathcal{D}\backslash\{case\})$ // k-nearest neighbors
14:         **for** $i \leftarrow 1$ to $\lfloor \tilde{w}_n \rfloor$ **do**
15:             // importance-based synthetic minority oversampling
16:             $(\xi_{nns}, \nu_{nns}) \leftarrow$ randomly choose one of the nns
17:             **for all** $j \in$ indices of ξ_{nns} **do**
18:                 $diff \leftarrow \xi_{nns}[j] - \xi_n[j]$
19:                 $\xi_{new}[j] \leftarrow \xi_n[j] + \mathrm{RANDOM}(0,1) \times diff$
20:             **end for**
21:             $d_1 \leftarrow \mathrm{DIST}(\xi_{new}, \xi_n)$ // Euclidean distance
22:             $d_2 \leftarrow \mathrm{DIST}(\xi_{new}, \xi_{nns})$ // Euclidean distance
23:             $\nu_{new} \leftarrow \frac{d_2 \times \nu_n + d_1 \times \nu_{nns}}{d_1 + d_2}$
24:             $new \leftarrow (\xi_{new}, \nu_{new})$
25:             $newCases \leftarrow newCases \cup \{new\}$
26:         **end for**
27:     **else**
28:         // importance weighting for least square loss.
29:         $new \leftarrow (\sqrt{\tilde{w}_n} \cdot \xi_n, \sqrt{\tilde{w}_n} \cdot \nu_n)$.
30:         $newCases \leftarrow newCases \cup \{new\}$
31:     **end if**
32:     $new\mathcal{D} \leftarrow new\mathcal{D} \cup \{newCases\}$
33: **end for**

Output: $new\mathcal{D}$ - resampled dataset

Parameter Learning and Dynamics Forecast

In the learning process of the proposed model, we minimize the least squared loss based on the resampled dataset. The learned parameter $\hat{\theta}$ can be obtained by solving the following optimization problem:

$$\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}\in\Theta} \frac{1}{N'}\left[\sum_{n=1}^{N'} \mathcal{L}(\nu_n, f(\boldsymbol{\xi}_n; \boldsymbol{\theta}))\right] + \gamma||\boldsymbol{\theta}||_2^2, \quad (10)$$

where $\mathcal{L}$ is the least squared loss, N' is the size of oversampled dataset $new\mathcal{D}$, 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 $\hat{\nu}_{d,t}^{(l)}$ as $\hat{y}_{d,t}^{(l)} = (1 + \hat{\nu}_{d,t}^{(l)})\bar{y}_{d,t}^{(l)}$, where $\bar{y}_{d,t}^{(l)}$ is the normal dynamics.

Experiments

Dataset

We evaluated the models based on two real datasets: the GPS-based mobility logs and transit search logs.

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. Hence, we did not use any dataset including personally identifiable information for analyzing the data and building the model.

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. The destination mainly denotes train stations, but it is sometimes set to places where events occur. Therefore, we used such records as the transit search logs. We added the number of search records per stations and time segment; therefore, we did not use any personal information for model learning.

Over six months (from October 1, 2019, to March 31, 2020), we used 58 POIs and their corresponding stations, including the Greater Tokyo Area, stadiums, shrines, and fireworks venues. Each POI is $600 \times 600m^2$, and we counted the mobility logs at each time segment as crowd dynamics.

Experimental Setups

T is set to 24 (i.e., one time segment denotes 1-h period). Following previous research [3], 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). We used the scheduled crowd dynamics observed one week in advance; thus, $p_d = 7$. We also set $p_w = 7$ to consider the people's schedule patterns specified two weeks before the day of the event. For the regularization term, we set $\gamma = 0.01$. For the hyper-parameter settings of the evaluation, we set $\beta = 0.1$, $\bar{\nu}_{\text{thre}} = 6.0$, $h = 5.0$.

We adopted a mean absolute error (MAE)-based metric to evaluate our model. The robustness of ordinary MAE to outliers prevents proper evaluation of the forecasting performance during anomalous crowding. To address this, we used the MAE conditioned by the irregularity score-based threshold $\bar{\nu}$ to measure the performance on normal and anomalous crowding, respectively. We defined normal sample (NS)-MAE, which calculates the MAE with samples whose irregularity score is less than $\bar{\nu}$, and anomalous sample (AS)-MAE, which calculates the MAE with samples whose irregularity score is more than $\bar{\nu}$. If NS-MAE is evaluated with a small threshold $\bar{\nu}$, the performance is evaluated only for daily normality. Furthermore, if the AS-MAE is evaluated with a significant threshold $\bar{\nu}$, performance is assessed on the exceptional anomalous crowding.

Comparative Models

CityProphet [5], Supervised-CityProphet (**SCP**) [6], bilinear Poisson regression [3] (**BPReg**), **CityOutlook** [1] served as comparison models. BPReg is dedicated to forecasting normal dynamics, as stated in the model setting section. CityProphet inputs context information and scheduled crowd dynamics and proposes two models: schedule-based population (SP) and descriptor-based population (DP). Supervised-CityProphet is introduced in the preliminary section. CityOutlook's parameters were optimized via importance-weighted least square loss; it is the proposed CityOutlook+ without importance-based synthetic minority oversampling.

Experimental Results

Overall Performance Comparison. Table 1 shows the overall evaluation in irregularity score forecast (Score) and crowd density forecast (density).

CityOutlook+ outperforms CityProphet and SCP by up to 9.82% and provides the same level of accurate forecasting in normal dynamics as SCP. It shows a 4.06%, 2.81%, and 5.84% improvement over SCP for $\bar{\nu} = 10.0$, 15.0, and 20.0, respectively, in the irregularity score forecast. The proposed method improves 6.64%, 4.41%, and 9.82% on crowd dynamics forecast for $\bar{\nu} = 10.0$, 15.0, and 20.0, relative to SCP. CityOutlook+

[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.

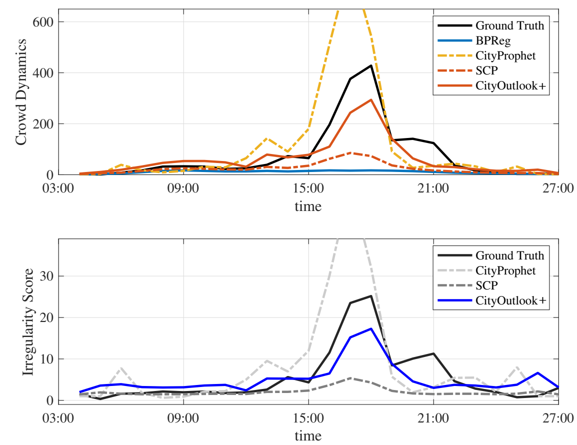
Model	NS-MAE						AS-MAE					
	$\bar{\nu}$ = 10.0		$\bar{\nu}$ = 15.0		$\bar{\nu}$ = 20.0		$\bar{\nu}$ = 10.0		$\bar{\nu}$ = 15.0		$\bar{\nu}$ = 20.0	
	Score	Density	Score	Density	Score	Density	Score	Density	Score	Density	Score	Density
CityProphet [5]	3.684	210.135	3.684	210.488	3.694	210.586	25.749	234.710	29.530	165.390	33.751	129.372
BPReg [3]	—	**84.749**	—	**84.811**	—	**84.873**	-	119.883	—	155.295	—	152.359
SCP (Baseline) [6]	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 [1]	0.796	93.881	0.801	93.921	0.805	93.968	**14.482**	102.163	**23.132**	138.420	32.698	132.301
CityOutlook+	0.772	98.854	0.778	98.901	0.783	98.948	15.057	**102.122**	23.199	**137.983**	**32.073**	**126.718**

outperforms CityOutlook by 4.2% in crowd density with $\bar{\nu}$ = 20.0 and the performance improvement on AS-MAE became larger with higher anomalous thresholds. We can conclude synthetic sampling approach was shown to be superior to just weighting by importance as CityOutlook, providing more forecasting robustness on anomalous patterns.

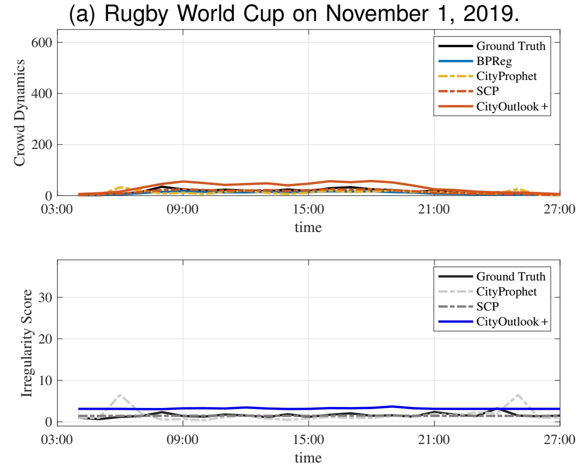
On the contrary, existing methods fail to forecast normal and abnormal dynamics simultaneously. SCP has performance drawbacks in AS-MAE, while BPReg cannot accurately forecast anomalous crowds, and CityProphet exhibits instability in both forecasts.

Case Study in Rugby World Cup 2019. We present crowd dynamics forecast examples for an actual event, i.e., Rugby World Cup 2019 third-place playoff, held in Ajinomoto Stadium, Tokyo, Japan[3]. Fig. 3a and Fig. 3b depict the results of BPReg, CityProphet, SCP, and CityOutlook+ against the ground truth on the event (November 1, 2019) and non-event day (November 8, 2019). The top graph shows crowd dynamics transitions, while the bottom graph presents the irregularity scores. The forecasting was conducted one week in advance using the scheduled crowd dynamics observed until 7 days before the event.

In the Rugby World Cup third-place playoff held in Tokyo Stadium, Ajinomoto Stadium, with a large number of spectators, the proposed CityOutlook+ produced a quantitative forecast on both crowd dynamics and irregularity scores and detected the occurrence of events, whereas CityProphet was unstable, and SCP failed to capture congestion. Results are shown in Fig. 3a. We also visualized the forecasting on November 8, 2019, where no events were held, and confirmed that the proposed method provided an accurate forecast for normal crowd dynamics similar to comparative models.



(a) Rugby World Cup on November 1, 2019.



(b) No events on November 8, 2019.

FIGURE 3: Forecasting on Ajinomoto Stadium.

Discussion

Uncertainty of Early Forecasting. Our method involves uncertainty in the early forecast due to using the following external information: transit search logs and weather information. As shown in prior work [6], the earlier is from the event date, the lower the search volume and the lower the indicator power of congestion. Therefore, earlier prediction (e.g., two weeks in advance) might result in lower prediction performance. We also used weather information to model the normal crowd density, but the weather forecast for the event day one week in advance may not be accurate. In a practical scenario, it is advisable to calculate multiple forecasts for possible weather patterns and plan coun-

[3] https://www.ajinomotostadium.com/

termeasures comprehensively.

Conclusion

In this study, we proposed CityOutlook+ for early crowd dynamics forecast one week in advance. Compared with the recent advances in crowding forecasting systems, the proposed method provides an effective learning strategy for anomalies, addressing the problem of data imbalance and scarcity of anomalies by the importance-based minority resampling. The experimental results on massive real datasets demonstrate the superiority of our model over the existing methods. Our predictive methodologies will enhance the accuracy of real-world crowd congestion forecasting, contribute significantly to improving crowd security and infection control measures, and stimulate further research within the community utilizing GPS-loggable pervasive devices.

REFERENCES

1. S. Anno *et al.*, “Cityoutlook: Early crowd dynamics forecast towards irregular events detection with synthetically unbiased regression,” in *Proc. of SIGSPATIAL*, 2021.
2. T. Xia and Y. Li, “Revealing urban dynamics by learning online and offline behaviours together,” *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, 2019.
3. M. Shimosaka *et al.*, “Forecasting urban dynamics with mobility logs by bilinear poisson regression,” in *Proc. of UbiComp*, 2015.
4. R. Jiang *et al.*, “Deepurbanevent: A system for predicting citywide crowd dynamics at big events,” in *Proc. SIGKDD*, 2019.
5. T. Konishi *et al.*, “Cityprophet: City-scale irregularity prediction using transit app logs,” in *Proc. of UbiComp*, 2016.
6. S. Anno *et al.*, “Supervised-cityprophet: Towards accurate anomalous crowd prediction,” in *Proc. of SIGSPATIAL*, 2020.
7. H. Shimodaira, “Improving predictive inference under covariate shift by weighting the log-likelihood function,” *Journal of statistical planning and inference*, 2000.
8. N. V. Chawla *et al.*, “Smote: synthetic minority over-sampling technique,” *Journal of artificial intelligence research*, 2002.
9. P. Branco *et al.*, “SMOGN: a pre-processing approach for imbalanced regression,” in *Proc. of the First International Workshop on Learning with Imbalanced Domains: Theory and Applications*, 2017.
10. Z. Fan *et al.*, “Citymomentum: An online approach for crowd behavior prediction at a citywide level,” in *Proc. of UbiComp*, 2015.
11. H. He and Y. Ma, *Imbalanced learning: foundations, algorithms, and applications*. Wiley-IEEE Press, 2013.
12. C. Elkan, “The foundations of cost-sensitive learning,” in *Proc. of IJCAI*, 2001.
13. L. Torgo and R. Ribeiro, “Utility-based regression,” in *Proc. of PKDD*, 2007.
14. M. Yamada *et al.*, “Relative density-ratio estimation for robust distribution comparison,” in *Proc. of NIPS*, 2011.
15. Y.-L. Yu and C. Szepesvári, “Analysis of kernel mean matching under covariate shift,” in *Proc. of ICML*, 2012.
16. H. Akaike, “Fitting autoregressive models for prediction,” 1969.
17. Y.-C. Chen, “A tutorial on kernel density estimation and recent advances,” *Biostatistics & Epidemiology*, 2017.

Soto Anno received his BE and ME from the Tokyo Institute of Technology in 2020, 2022. He is currently a Ph.D. candidate student at the Tokyo Institute of Technology. His research focuses on urban computing. Contact him at anno@miubiq.cs.titech.ac.jp.

Kota Tsubouchi received the PhD degree from the University of Tokyo, Japan, in 2010. Until March 2012, he did research about on-demand traffic systems with the University of Tokyo. Since April 2012, he has been a data scientist and a senior researcher with Yahoo JAPAN Research. His research interest focuses on data analysis including human activity logs, such as location information, search logs, shopping history, and sensor data. Contact him at ktsubouc@yahoo-corp.jp.

Masamichi Shimosaka received his BE, ME, and Ph.D. from the University of Tokyo in 2001, 2003, and 2006, respectively. He joined Tokyo Institute of Technology as an associate professor in July 2015. Prior to joining Tokyo Institute of Technology, he was a faculty member at the University of Tokyo from 2006 to 2015. While his Ph.D. candidate, he was funded by the Japan Society for the Promotion Science as a research fellow. His research interests include machine intelligence and ubiquitous computing. He is a member of the ACM and IEEE. Contact him at simosaka@miubiq.cs.titech.ac.jp.