

Forecasting Urban Dynamics with Mobility Logs by Bilinear Poisson Regression

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

Keisuke Maeda
The University of Tokyo
Tokyo, Japan
maeda@ics.t.u-tokyo.ac.jp

Takeshi Tsukiji
The University of Tokyo
Tokyo, Japan
tsukiji@ics.t.u-tokyo.ac.jp

Kota Tsubouchi
Yahoo! JAPAN Research
Tokyo, Japan
ktsubouc@yahoo-corp.jp

ABSTRACT

Understanding people flow in a city (*urban dynamics*) is of great importance in urban planning, emergency management, and commercial activity. With the spread of smart devices, many studies on urban dynamics modeling with mobility logs have been conducted. It is predictive analysis, not analysis of the past, that enables various applications contributing to a more prosperous society. To deal with the non-linear effects on urban dynamics from external factors, such as day of the week, national holiday, or weather, we propose a low-rank bilinear Poisson regression model, for a novel and flexible representation of urban dynamics predictive analysis. The results obtained from an experiment with one year's worth of mobility records suggest the high prediction accuracy of the proposed model. We also introduce the following applications: regional event detection via irregularities, visualization of urban dynamics corresponding to urban demographics, and extraction of urban demographics of unknown point of interests.

Author Keywords

urban dynamics; bilinear model; matrix factorization; Poisson regression; multi-task learning;

ACM Classification Keywords

H.5.m. Information Interfaces and Presentation (e.g. HCI): Miscellaneous

INTRODUCTION

Understanding people flow in a city is of great importance in urban planning, emergency management, and commercial activity [25, 28, 12, 32, 33, 29]. Modeling people transitions in a specific area leads to understanding the characteristics of the area. It can be assumed that people flow in business and residential districts obviously differ in both size and peak time.

In this paper, we define these dynamic behaviors as *urban dynamics*. With the spread of smart devices, a large amount of mobility logs, such as GPS or cell tower logs, has been accumulated. Therefore, many studies on urban dynamics modeling with such mobility logs have been conducted [28, 12, 32, 22, 21]. In these studies, models were designed to help in the analysis of the cities, a person's daily activities, effect of disasters and so on.

Studies focusing on improvement of city planning against disasters [28, 12, 32] and traffic congestion [25] have been explored over the past decade. Though these studies help us understand urban dynamics in a past disaster situation, they did not focus on future prediction. Song et al. [28] and Fang et al. [12], analyzed people flow during the Great East Japan earthquake by using mobility logs. The results of the analysis could be useful for reducing traffic congestion during a disaster. In another study [32], a population model in specific areas was constructed from mobility logs for real-time anomaly detection. The importance of monitoring dynamic changes in people flow is also described in [25]. However, it is more useful to predict people flow in advance for preventing accidents. If we know in advance where and when large groups of people will congregate, more police security officials can be strategically placed.

As noted above, it is clear that analysis of urban dynamics is essential in the design of infrastructure. It is noteworthy that the urban dynamics is affected by urban demographics, i.e., land-use and users in the area of interest. Therefore, analyses of urban dynamics would be helpful for economic applications. For example, if the urban dynamics in a commercial area where people tend to concentrate, such as stations or department stores, can be predicted in advance.

There have been studies on analyzing the demographics of urban areas from mobility logs [22, 21]. Reades et al. [22] divided Rome into 'pixels' and counted the number of access logs in each pixel to analyze when that pixel is more active. Subsequently, as urban demographics, business location data are extracted from the Yellow Pages. By comparing activity patterns and urban demographics, areas with insufficient public services or commercial premises were suggested. Phithakkintnukoon et al. [21] analyzed not only areas, but

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from Permissions@acm.org.
UbiComp '15, September 7-11, 2015, Osaka, Japan.

Copyright is held by the owner/author(s). Publication rights licensed to ACM.
ACM 978-1-4503-3574-4/15/09...\$15.00.
<http://dx.doi.org/10.1145/2750858.2807527>

also a person's transition patterns from one area to another. This paper indicates the fact that similar user categories and movement patterns can be found in similar urban dynamics. Predictably recovering such demographics specific to an area in advance may be useful in designing urban infrastructures or planning store role-outs.

From the mobility logs from taxicabs, Zhang et al. [33] constructed a model of refueling behavior and proposed a recommendation system to reduce waiting times at gas stations. The waiting times and the most crowded times of the day for each gas station were analyzed. Considering the balance of supply and demand, the location where to build new gas station became clearer. Analysis of data from IC train tickets [29] resulted in the understanding of people flow and demand of transportation in cities. These studies assume that human transition is affected by day of the week or weather conditions. However, the most important thing is to respond the demand immediately and properly. This requires predictive analysis of urban dynamics, not analysis of past data or on-line monitoring, enabling the development of various applications to promote a more prosperous society. However, predictive analysis of urban dynamics has not been established. To address the issues with predictive analytics in urban dynamics, the use of external factors, such as day of the week, national holiday, weather condition, or urban demographics of the area, plays an important role. In the following section, we describe the importance of the external factors for urban dynamics prediction, and that existing models cannot handle the external factors properly.

Related works and their limitations

Urban dynamics have been studied for many years. Analysis of a person's daily activity pattern is closely related to the field of mobile sensing or next place prediction [8, 2, 3, 7]. They both predict quantity or place based on a period of time. Therefore, statistical models have been extensively studied. However, they are inadequate for urban dynamics prediction.

One of the models leverages the periodicity of urban dynamics [5, 24, 13, 32]. It relies on the assumption that activities of people are repeated daily or weekly. Witayangkurn et al. [32], constructed seven hidden Markov models (HMMs) for each day of the week based on the idea that urban dynamics repeats, not only daily but also weekly. Though these models based on periodicity can predict daily or weekly activity, they cannot predict non-periodic activity related to national holidays and much longer periodic effects, i.e., seasonal effects. Indeed the irregularity-detection approach [32] shows that national holidays can be detected successfully, but it is not always suitable to regard national holiday as an irregularity but as a regular holiday. To construct custom HMMs specific to national holidays, the number of HMMs has to be doubled corresponding whether it is a national holiday. In other words, if other external factors are leveraged for richer representation, the number of HMMs exponentially grows by the combination of external factors. The increase in the number of models means a decrease in amount of data assigned to each model, which results in worse prediction accuracy.

Matrix and tensor factorization are well known methods for analyzing urban dynamics [10, 11, 12, 30, 33, 34, 30]. Factorization methods extract typical latent activity patterns and load factors of patterns at the same time, (see Fig. 2). These methods help us understand urban dynamics by visualizing latent patterns; however, these methods are not designed for predictive analysis for future urban dynamics. Specifically, these cannot predict urban dynamics in advance except when some of observation data are given to determine the load factors. Fan et al. [12] applied non-negative tensor factorization to mobility logs to extract basic activity components, such as commuting patterns or working patterns. Because of the sparseness of the target data, the feature matrix is also factorized collaboratively to fill in the target data tensor and achieve higher accuracy. A tensor factorization method with side information, such as SNS data [30], refueling logs of taxicabs [33], and noise complaint data [34], has been proposed. In contrast with past studies, they discussed the direction of future research project and how to take in account the factors related to the urban dynamics. Unfortunately, feature matrix or other factors are only used to help fill in the sparse data. The main point is still to analyze the past, not to predict the future.

Similar to the factorization method, some studies have focused on the fluctuation of human activity using mixture models [14, 15, 1, 27, 31, 19, 18]. Mixture models have high power of expression with non-linear distributions. Therefore, the target data in previous research project are quite varied, e.g., people-count data at a building entrance and traffic-count data on a freeway [14, 15], professional basket ball players' shot attempts [18], going-out behavior [27, 31] and number of pedestrians in a specific area [19]. However, mixture models are difficult to use for predictive analysis since the load factor to switch the appropriate latent patterns are not determined in priori. Similar to the factorization method, these models are not suitable for urban dynamics prediction.

When the load factor of each basic activity component is explicitly described by external factors, such as day of the week and weather conditions, it enables us to realize predictive analysis in urban dynamics. From this view-point, a discriminative model is used to model and predict the number of activities [6, 9, 17]. The crowd counting prediction [6], hazard index for urban roads [17], and traffic flow [9] were modeled using regression models in previous studies. The regression models can predict the future from related external factors. However, due to the non-linearity nature of urban dynamics, simple linear regression models are not suitable. The effects of external factors are fundamentally not independent. For example, an active population drastically increases during *commute time* on *weekday*. Due to the fact that simple linear regression models take into account each factor independently, they fail in representing the shift in peak time between weekdays and weekends. Similarly, generalized linear regression, e.g. Poisson regression, cannot handle the external factors properly.

Proposed bilinear Poisson regression model for urban dynamics prediction

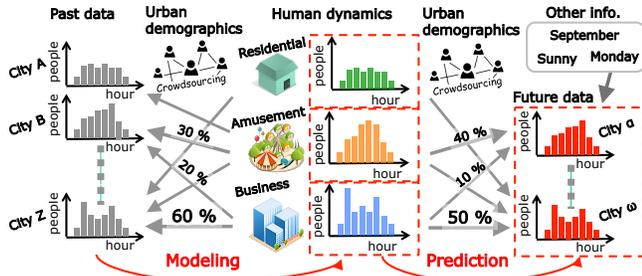


Figure 1. Urban dynamics prediction with proposed model

To address these non-linear effects from external factors, we propose a low-rank bilinear Poisson regression model, for flexible representation of urban dynamics for predictive analysis. The proposed model is not only useful for predicting future urban dynamics, but also enables us to understand the relationship between each external factors and urban dynamics. If the external factors of the target day are acquired in advance, the model can predict the future urban dynamics precisely according to this relationship. The concept of the proposed urban dynamics prediction with the proposed model is shown in Fig. 1. First, the basic activity patterns are extracted from a large amount of mobility logs. Subsequently, future urban dynamics is shaped according to the effect of external factors including demographics, such as land use and user information, in the corresponding area collected from market survey via crowd-sourcing.

We summarize the contribution of our model from a technical perspective. The current state of the art can be summarized as follows. 1) Almost all the research projects in point process or event process, the Poisson process or Poisson regression has been explored in the literature where no mixture effects are assumed. 2) In order to handle the changes in patterns due to the external factor changes, the mixture model and factorizing model have been focused in the field recently. However, these models cannot handle the external factors directly. 3) Even if we incorporate external factors into simple linear models, the external factors only affect the scale of predictive urban dynamics. Thus, simple linear models cannot be used in our application scenario. To the best of our knowledge, our model is the first that is able to consider not only the mixture effects but also incorporate multiple external factors to represent the changes in urban dynamics.

We also introduce the following prominent applications based on the proposed prediction model: 1) regional event detection via irregularities in the ordinary urban dynamics. Being based on passive and pervasive sensing on smart devices, our regional event detection can find various type of events. The event detection using SNS posting data [23] is suitable to detect the newsworthy events, but cannot detect unremarkable or nameless events, which can be detected by ours. 2) Visualization of urban dynamics corresponding to urban demographics. 3) Extraction of urban demographics of unknown point of interests (POIs). The extraction of urban demographics from mobility logs is also really meaningful because they suggest that we are released from conducting the market surveys of anywhere.

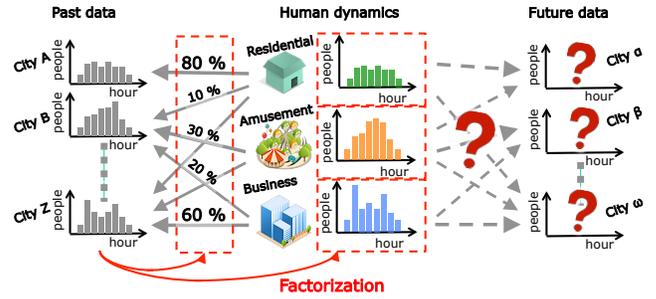


Figure 2. Analysis of urban dynamics with factorization

The contributions of this paper are as follows. For *ex ante* (ex. for the week ahead) prediction of urban dynamics, we propose a low-rank bilinear Poisson regression model. In contrast to current models, the proposed model incorporates external factors into urban dynamics. By evaluating the proposed model with a large amount of mobility logs and open-sourced data, we argue it achieves more precise prediction than comparative models. We also introduce the following applications based on the proposed prediction model: 1) regional event detection, 2) visualization of urban dynamics corresponding to each characteristic of an area, and 3) extraction of urban demographics from any POI, and confirmed the usefulness of the proposed model.

URBAN DYNAMICS PREDICTION BY LOW-RANK BILINEAR POISSON REGRESSION

Problem Settings

We modeled the daily transition of an active population in several areas of Tokyo as urban dynamics. We define the term POI as a focus spot and the target area as a rectangular area around the POI¹. Let $f(t)$ be the number of access logs per unit time (e.g. per minute) t ($0 \leq t \leq T$) in the target area, where T denotes time of day. Fig. 3 shows the transition of $f(t)$ in a target area in a day. Let $\Delta = T/S$ be the time interval, where S denotes the number of time segments in a day, then we define the total number of access logs observed in each time segment as an active population². We model this transition of an active population as a histogram, as shown in Fig. 4. Let $\mathbf{h} = \{h_1, h_2, \dots, h_S\}$ be a histogram that represents a one-day activity pattern, in which the s -th bin size $y = h_s$ is defined as $y = \int_{(s-1)\Delta}^{s\Delta} f(t)dt$. On the assumption that population density follows Poisson distribution, the likelihood of y is written as follows.

$$p(y) = \text{Pois}(y|\lambda) = \frac{\lambda^y \exp(-\lambda)}{y!} \quad (1)$$

Therefore, to predict an active population in each area, we estimate the mean parameter λ of the Poisson distribution. As a richer representation than Poisson distribution over histograms, Gaussian Cox processes or continuous Poisson processes [1], namely, doubly stochastic processes, vary the variance of λ across time as well as the mean of the active population; however, we do not use these point processes due to the high computational cost of inference.

¹We used 300 POIs and $900 \times 900m^2$ area; however, the scale of the experiment is extensible.

²[20] reported that active population could be inferred using a rebate method using access logs from smart devices.

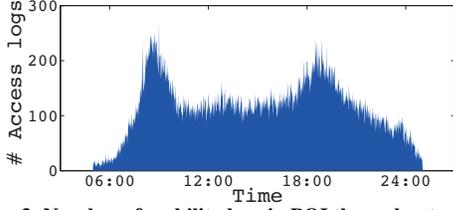


Figure 3. Number of mobility logs in POI throughout one day

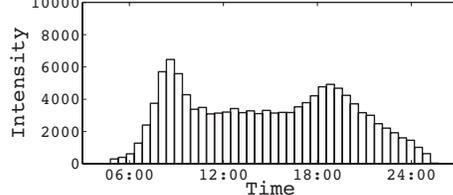


Figure 4. Active population histogram

Generalized linear model

We now argue that prediction using a naive generalized linear model can not work properly even if external factors are incorporated into the model. In the generalized linear model, we assume that the mean parameter λ is regressed by the time factor with weight parameters. Specifically, λ can be regarded as a function of a time feature vector \mathbf{t} formulated as follows, where β is a weight parameter vector.

$$\lambda(\mathbf{t}) = \exp(\mathbf{t}^\top \beta), \quad (2)$$

where \mathbf{t} is the S dimensional vector and its s -th element denotes the corresponding time segment. Inspired by the flexible representation of λ by mixture models, as in [14], we cast the time factor vector \mathbf{t} as an indicator function of normal distribution with mean parameter τ , that is,

$$\mathbf{t} = \{t_s | t_s = \mathcal{N}(s | \tau, \sigma), s = 1, \dots, S\}, \quad (3)$$

where $\mathcal{N}(\tau, \sigma)$ is the normal distribution with τ and variance parameter σ . In this case, σ is the hyper parameter that exhibits the effect of time and empirically determined by effective cross validations [26]. In addition to \mathbf{t} , we introduce \mathbf{d} , which contains various external factors that affect the transition of an active population, such as day of week, weather and urban demographics of the area. A naive approach by generalized linear regression to model transition in urban dynamics with these external factors may be thought of useful when external factor \mathbf{d} is incorporated as a parameter into the model. However, it is impossible to shift the peak of an active population because the correspondence between day of the week and time factor cannot be considered at the same time. In practice, λ in a generalized linear regression model should be formulated as a dot product by the extended features $[\mathbf{t}^\top \mathbf{d}^\top]$

$$\lambda(\mathbf{t}, \mathbf{d}) = \exp([\mathbf{t}^\top, \mathbf{d}^\top] \beta); \quad (4)$$

however, it cannot work properly. For example, Fig. 5 illustrates the transition on weekdays and weekends in Shinjuku, a downtown area of Tokyo, estimated by this representation. In this figure, two peaks can be found as the commute time. Fig. 5 also suggests that the weekday/weekend factor in affect active population size. On the other hand, the estimated transition on weekend also has peaks for *commute time*. This stems from the fact that the linear model cannot combine both factors, *commute time* and *weekend*, simultaneously.

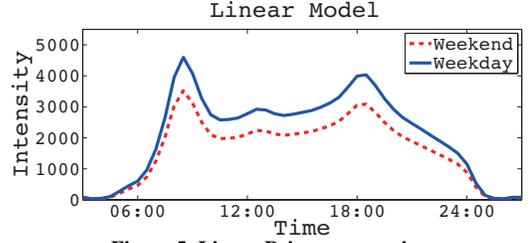


Figure 5. Linear Poisson regression

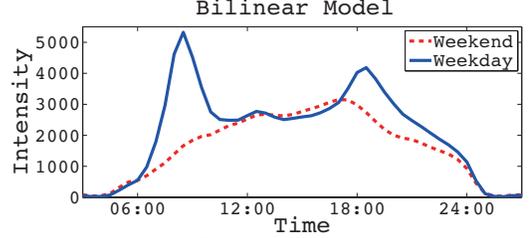


Figure 6. Bilinear Poisson regression

Proposed bilinear Poisson regression model

To model urban dynamics with both time and external factors in a flexible manner, we model λ , which denotes an active population in a time segment as

$$\lambda(\mathbf{t}, \mathbf{d}) = \exp(\mathbf{d}^\top \mathbf{B} \mathbf{t}), \quad (5)$$

where \mathbf{t} and \mathbf{d} denote the same feature, as in generalized linear regression, and $\mathbf{B} \in \mathbb{R}^{M \times S}$ denotes the weight parameter matrix. Fig. 6 illustrates the urban dynamics obtained using our bilinear regression model on weekdays and weekends in Shinjuku. On weekdays, there are two peaks in the morning and evening, which can be interpreted as *commute time*. Furthermore, the model took into account the fact that there is no obvious peak in weekends. Our model incorporates the active population increase on the condition that the day is a *weekday* and the time is *commute time*. Therefore, it is suitable for modeling urban dynamics compared to the naive linear model mentioned above.

In addition to formulating the proposed model with bilinear representation, we add the assumption that the rank of weight matrix \mathbf{B} prone to decrease, which makes both learning the parameters stable, and the results interpretable. To achieve rank reduction, the weight matrix \mathbf{B} in (5) is assumed to be a product with two low-rank matrices, $\mathbf{U} \in \mathbb{R}^{M \times K}$ and $\mathbf{V} \in \mathbb{R}^{S \times K}$. Note that K and M satisfy $K \ll M$.

$$\lambda(\mathbf{t}, \mathbf{d}) = \exp(\mathbf{d}^\top \mathbf{B} \mathbf{t}) = \exp(\mathbf{d}^\top \mathbf{U} \mathbf{V}^\top \mathbf{t}) \quad (6)$$

The row vectors of the weight matrix \mathbf{B} depict M types of patterns, which correspond to M types of external factors. In the low-ranked model, matrix \mathbf{V} relates to time factor \mathbf{t} and the k -th ($1 \leq k \leq K$) column vector denotes the k -th dynamics pattern. Matrix \mathbf{U} relates to external factor \mathbf{d} and its column vectors depict the relationship between K types of dynamics patterns and external factors, that is, several external factors share K types of dynamics patterns. This shrinkage helps us to understand the characteristics of urban dynamics. We optimize the model by maximizing the log likelihood against training data. Let us assume that we have N days of training data $\mathbf{D} = (y_{n,s}, \mathbf{d}_{n,s}, t_{n,s}), n = 1, \dots, N, s =$

Table 1. Example of POIs

	Sight-seeing spots	Stations
1	Tokyo Disneyland	Shinjuku
2	Tokyo Skytree	Ikebukuro
3	Tokyo DisneySea	Shibuya
4	Asakusa	Yokohama
5	Tokyo Tower	Kita-Senju
6	Narita Airport	Takadanobaba
7	Nihon-odori	Shinagawa
8	Kamakura	Tokyo
9	Nikko	Akihabara
10	Owakudani	Shinbashi

$1, \dots, S$, then the log likelihood of the data is written as

$$\begin{aligned} \ln L(\mathbf{U}, \mathbf{V}) &= \sum_n \sum_s \ln \text{Pois}(y_{n,s} | \lambda(\mathbf{t}, \mathbf{d})) \\ &\propto \sum_n \sum_s \{y_{n,s} \mathbf{d}_{n,s}^\top \mathbf{U} \mathbf{V}^\top \mathbf{t}_{n,s} \\ &\quad - \exp(\mathbf{d}_{n,s}^\top \mathbf{U} \mathbf{V}^\top \mathbf{t}_{n,s})\} \quad (7) \end{aligned}$$

Specifically, we plug in the log likelihood defined above with a regularization term into the following objective function

$$\hat{\mathbf{U}}, \hat{\mathbf{V}} = \arg \min_{\mathbf{U}, \mathbf{V}} \{-\ln L(\mathbf{U}, \mathbf{V}) + \Omega(\mathbf{U}, \mathbf{V})\} \quad (8)$$

where $\Omega(\mathbf{U}, \mathbf{V}) = \gamma \|\mathbf{U}\|_2^2 + \gamma \|\mathbf{V}\|_2^2$ is an L2 regularization term with a hyper parameter $\gamma (> 0)$. Although the problem in (8) is not bi-convex in \mathbf{U} and \mathbf{V} , optimizing one parameter is convex once the other is fixed. Thanks to this property, the model can be optimized in a similar way to the alternating least square algorithm [4], which is often used in non-negative matrix factorization (NMF). In each convex programming, we use Newton's method.

Design of features

In addition to vector \mathbf{t} , which denotes the time factor, we use vector \mathbf{d} , which denotes external factors in the proposed model. We describe the design of parameter \mathbf{d} . The external factors that seem to affect urban dynamics are as follows: day of the week, weekday/weekend/national holiday, month, season, weather, weather warning, locality, and number of stores, events and train lines near the target area. A factor that takes one state among M states, such as day of week and season, is denoted by an M -dimensional vector in which the corresponding ν -th element is 1 and all other elements are 0. Therefore, it is written as

$$\mathbf{d} = \{d_m | d_m = \delta_{m,\nu}, m = 1, \dots, M\}, \quad (9)$$

where $\delta_{m,\nu}$ is the Kronecker delta. When regarding a numerical value, such as the number of events and train lines, the quantitative values can be incorporated as a feature value.

EXPERIMENT

To evaluate predictive performance of the proposed low-rank bilinear Poisson regression model, we conducted an experiment with a large amount of mobility data.

Mobility data

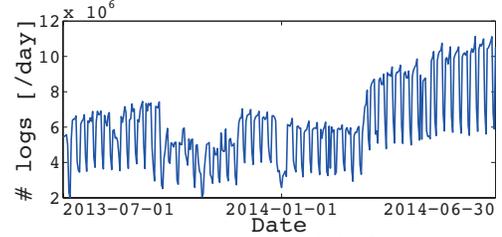


Figure 7. Number of logs over 300 POIs per day

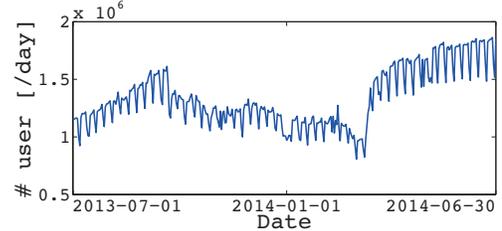


Figure 8. Number of users per day

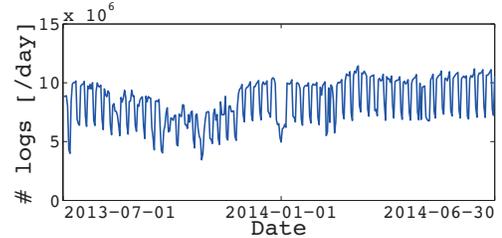


Figure 9. Number of normalized logs per day

As mentioned in the section for model description, the target of modeling is the active population transition in a specific area. In this experiment, we used massive mobility log records obtained from the disaster alert mobile application released from Yahoo! JAPAN³, which scales to the population size; therefore, the dataset could be used for predicting the active population⁴. Anonymized data were collected for one year (from July 1, 2013 to June 30, 2014), and consisted of 40 million records per day. Each record include three elements, timestamp, latitude, and longitude. A mobility log was recorded only when mobile devices were moving not stable. That is the reason the number of logs represents an *active* population. Three hundred specific city squares areas were selected as POIs. Some examples of the POIs are listed in Table 1.

The size of the target areas was set to $900 \times 900m^2$ (delta longitude = ± 0.004 degrees and delta latitude = ± 0.005 degrees), as in [12]. The number of GPS logs in each target area was counted and regarded as active populations of the POIs. The transition of the total number of data over the 300 POIs per day and the number of the users per day are shown in Fig. 7 and Fig. 8 respectively. Both numbers have weekly patterns which represent a more active population on weekdays and less so on weekends. The reason the number of total logs varied by period (Fig. 7) is not because of the variation in the active population but that in users (see Fig. 8). This stems from the fact that the users of the mobile application have been increasing gradually after the release in 2012.

³<http://emg.yahoo.co.jp/>

⁴[20] reported that active population can be inferred using a rebate method using access logs from smart devices.

To decrease this effect, we normalized the total number of logs according to the number of users on the last day, June 30, 2014. The number of logs after normalization (shown in Fig. 9) was used in the evaluation.

In this experiment, the fundamental period T was set to 24 hours. 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-hour notation). From empirical evaluations, the bin width of the histogram was set to 30 minutes; thus, the number of bins was 48.

Performance measure

The following three criteria were used as performance measure for evaluating our model and several others for comparison: mean absolute error (MAE), mean negative log likelihood (MNLL), and mean absolute peak error (MAPE). When the test data and the prediction value are represented as $y_{n,s}, \hat{\lambda}_{n,s}, n = 1 \dots N_t, s = 1 \dots S$, the peak time is defined as follows.

$$\tau_n = \arg \max_s y_{n,s} \text{ for } s \in \{1 \dots S\} \quad (10)$$

$$\hat{\tau}_n = \arg \max_s \hat{\lambda}_{n,s} \text{ for } s \in \{1 \dots S\} \quad (11)$$

With this expression, the performance measures can be formulated as

$$\text{MAE} = \mathbb{E}(|y_{n,s} - \hat{\lambda}_{n,s}|) \quad (12)$$

$$\text{MNLL} = \mathbb{E}(-\ln(\text{Pois}(y_{n,s}|\hat{\lambda}_{n,s}))) \quad (13)$$

$$\text{MAPE} = \mathbb{E}(|\tau_n - \hat{\tau}_n| \Delta), \quad (14)$$

where $\mathbb{E}(x)$ is the expectation of x , and Δ is the bin width of the histogram. To assess the model in a fair manner, five-fold cross validation was conducted. The model was trained with data of 30, 90, 180 days at each POI, and the next 180 days of data were used as test data. We evaluated two types of scenarios: 1) *ex ante* prediction: prediction only with external factors, and 2) *ex post* prediction: prediction with external factors and some data of the target day. The *ex ante* prediction is used when one wants to know future urban dynamics, for example to predict the urban dynamics for the week ahead with the external factors of that day (Wednesday, sunny and area is a business district). The urban dynamics under the selected external factors will be simulated by *ex ante* prediction. In contrast, *ex post* prediction is used when the urban dynamics of morning has already been observed and one wants to predict the afternoon with high accuracy.

External factor data

We leveraged day of the week, whether it is a national holiday, and the weather as the external factors. We experimentally select these external factors because they are intuitively assumed to affect dynamics patterns and can be accessible from open APIs. Taking into account the comparison models perform best when they are optimized to each POI, we evaluate predictive performance of each POI independently. Therefore urban demographics was excluded from external factors.

Weather data were collected from the Japan Meteorological Agency's website⁵. The weather of the nearest meteorological station from the POI was used. The candidate meteorological stations were Tokyo, Yokohama, Chichibu, Kumagaya, Chiba, and Utsunomiya. The category of the weather was 1 to 4, {sunny(1), cloudy(2), rainy(3), or rough weather(4)}.

Comparative models

A baseline model, linear Poisson regression model, Dirichlet-mixture Poisson model, and non-negative matrix factorization model were used for comparison.

The baseline model returns the average number of past data logs at each time segment, described as $\lambda(\mathbf{t}, \mathbf{d}) = \mathbb{E}(h_s)$. This model only takes into account the time factor, in other words, just modeling the average pattern over all days. The effect from external factors (ex. day of the week) could not be taken into account.

We also compared the linear Poisson regression formulated in (4). The parameters were optimized with Newton's method to maximize the log likelihood against the training data. As mentioned in the section describing generalized linear regression, this model reflects both time and external factors; however, the effect may result in scale instead of peak time shifts. In addition to time factor $\mathbf{t} \in \mathbb{R}^S$, day of the week (7 dim.), whether it is national holiday (2 dim.), weather (4 dim.), and the constant term (1 dim.) were used as external factors $\mathbf{d} \in \mathbb{R}^{14}$.

We also compared our proposed model with mixture models. Specifically, we used the Dirichlet-Poisson mixture model proposed by Shimosaka et al. [27]. Since it models the daily activity as a histogram and predicts the future, it is related to ours. With this model, it is assumed that the size of each histogram bin is distributed under Dirichlet-Poisson mixture distributions. The model is formulated as follows. Let z_n be the discrete random variable drawn from multinomial distribution represented by ω and express the n -th day belongs to which of the K distributions. $p(y_{n,s}|z_n = k)$ is the k -th distribution of $y_{n,s}$ trained by data. $\text{GEM}(\cdot)$ denotes the Griffiths-Engen-McCloskey distribution, and $\mathcal{M}(\cdot)$ denotes the multinomial distribution.

$$p(y_{n,s}) = \sum_k^K p(z_n = k)p(y|z_n = k) \\ p(z_n) = \mathcal{M}(z_n|\omega), \omega \sim \text{GEM}(\cdot) \quad (15)$$

The Dirichlet-mixture enables the Dirichlet-Poisson mixture model to estimate the proper K , i.e., the number of distributions to mix. It is true that this model has novel expressiveness with the framework of determining the proper number of components; however, when it comes to prediction, some of target day data must be observed. With no data, the model cannot determine the posterior of latent factor z , which results in low prediction accuracy. In the experiment, the prediction with no data is represented as DP0, the prediction with data of n hours is represented as DP n . DP0 corresponds to *ex ante* prediction and DP n corresponds to *ex post* prediction.

⁵<http://www.data.jma.go.jp/obd/stats/etrn>

In DP n , the probability that the target day belongs to the k -th category ω_k^* and posterior distribution $p(y)^*$ is as follows.

$$\omega_k^* = \pi \frac{p(y_{1:n}|z=k)}{\sum_k p(y_{1:n}|z=k)} \quad (16)$$

$$p(y)^* = \sum_k \omega_k^* p(y|z_n=k) \quad (17)$$

We used DP0 as a competitor in *ex ante* prediction and DP6 and DP12 in *ex post* prediction.

In addition to mixture modeling, we also compared our model with the non-negative matrix factorization (NMF) model. With the NMF model, the data matrix of each POI is factorized using the method proposed by Fan et al. [12]. Though the tensor factorization is originally used in [12], NMF model for each POI is suitable for prediction. The mobility data is represented as the matrix $Y \in \mathbb{R}^{S \times N}$. S, N are the number of time segments and the day respectively. The NMF method approximates this matrix as the linear combination of coefficient ω_k and the basic matrix Y_k , which is the Kronecker product of unit time vector \mathbf{u}_k and unit day vector \mathbf{v}_k .

$$Y \approx \sum_k \omega_k \mathbf{u}_k \otimes \mathbf{v}_k \quad (18)$$

In contrast to a previous study using the factorization method focused on extracting the latent pattern \mathbf{u}, \mathbf{v} and the load factor ω [12], we have to reuse this information for *ex post* prediction. Similarly to the mixture models mentioned above, this model also requires some of the target day data for high-accuracy prediction. When the data before time τ , $\mathbf{y}_{1:\tau}$ is already known, and the coefficient ω_k is optimized to minimize the square errors from the observed data. Then the rest of the data are predicted by mixing up the $\mathbf{u}_{k,\tau+1:S}$ according to $\hat{\omega}_k$. Though the NMF model was not originally designed to handle the distribution of data, we predicted $\hat{\mathbf{y}}_{\tau+1:S}$ on the assumption that it follows the Poisson distribution with the mean parameter $\hat{\lambda}_{\tau+1:S}$. The predicted parameter $\hat{\lambda}_{\tau+1:S}$ and $\hat{\omega}_k$ are written as follows.

$$\hat{\omega} = \arg \min_{\omega} \|\mathbf{y}_{1:\tau} - \sum_k \pi_k \mathbf{u}_{k,1:\tau}\|^2 \quad (19)$$

$$\hat{\lambda}_{\tau+1:S} = \sum_k \hat{\omega}_k \mathbf{u}_{k,\tau+1:S} \quad (20)$$

In NMF0, with no data, the load factor $\hat{\omega}$ is equal to ω .

Results

Fig. 10~Fig. 12 show the *ex ante* performance among the prediction models. We averaged the performance measure calculated in 300 POIs, and evaluated the performance by the averaged value. The horizontal axis shows the number of days of training data, and the vertical axis shows the mean and variance of each evaluation from five-fold cross validation. A smaller value denotes the better performance for each criterion. With respect to MAE, please note that the average number of active population in each time unit is approximately 600.

According to Fig. 10 and Fig. 11, the proposed low-rank bilinear Poisson regression model exhibited the best estimation

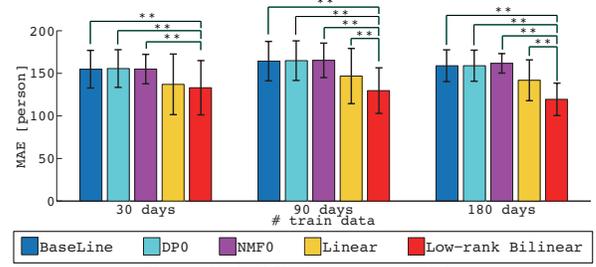


Figure 10. MAE of *ex ante* prediction

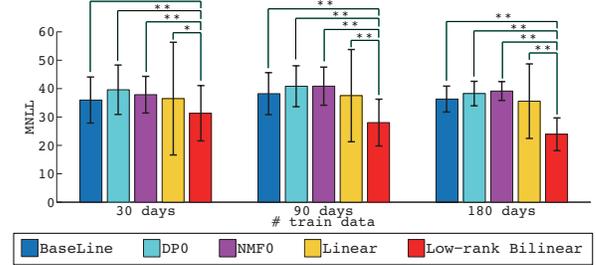


Figure 11. MNLL of *ex ante* prediction

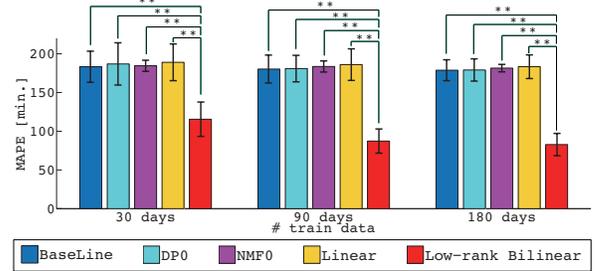


Figure 12. MAPE of *ex ante* prediction

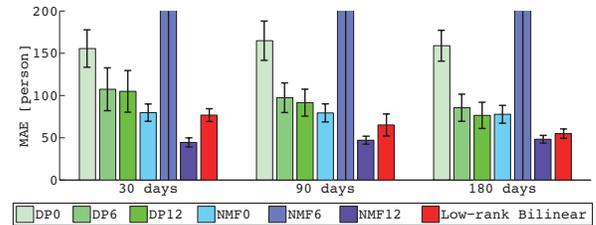


Figure 13. MAE of *ex post* prediction

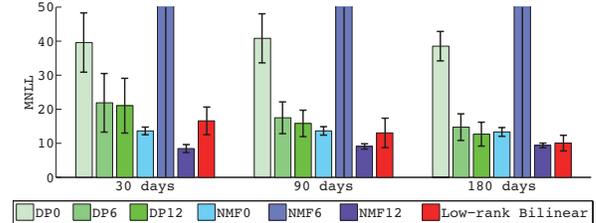


Figure 14. MNLL of *ex post* prediction

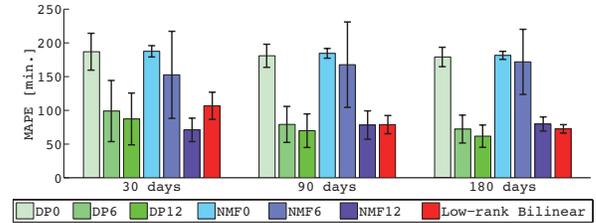


Figure 15. MAPE of *ex post* prediction

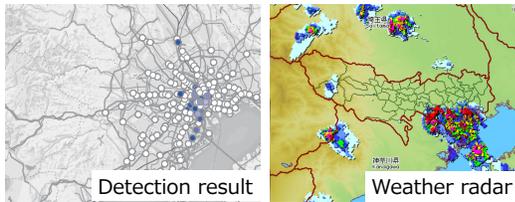


Figure 16. Unexpectedly heavy rain in Tokyo

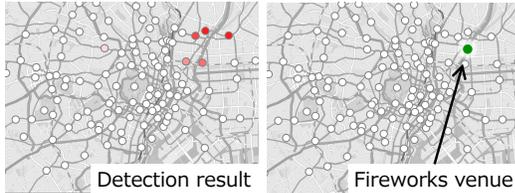


Figure 17. Large fireworks event on Sumida river

accuracy compared with the conventional models. In particular, the variance of the proposed model was small compared with that of the linear Poisson regression model, since the proposed method can reflect peak time shift with the external factors. Fig. 12 also shows that the proposed model drastically reduced peak time estimation errors.

Subsequently, we evaluated *ex post* prediction, i.e., prediction accuracy after some of the target data was observed. The data scale was normalized to decrease the effect of variation in data size over a period. Fig. 13 and Fig. 14 show that the proposed model performed well in urban dynamics estimation. In particular, it is noteworthy that the proposed model could estimate better than DP12 and NMF6⁶. This means that our proposed model can predict urban dynamics even better than the models that can take into account the actual urban dynamics of half or quarter of the target day. Though NMF12 exhibited good performance, the requirement of observation over 12 hours is not feasible for some applications. Our proposed model provides wide applicability, in that it can estimate urban dynamics without long observation.

APPLICATIONS OF OUR PROPOSED MODEL

This section shows some of prominent applications of the proposed model to detect regional event, visualize and extract demographics of urban dynamics.

Event detection from irregularities in urban dynamics

The proposed model learned from data with the same factors (place, day of week, weather, and time) as mentioned in the previous section. This means that the model can estimate urban dynamics in an *ordinary* case. If a regional event occurs, the active population drastically increases/decreases compared to that of the prediction. We define the irregularity index i as follows

$$i = \frac{y - \hat{\lambda}}{\hat{\lambda}}, \quad (21)$$

⁶The variance of NMF6 is extraordinary large. We found a time vector \mathbf{u}_k whose elements are almost 0 in the particular POI. In NMF6, the coefficient of this \mathbf{u}_k (i.e. $\hat{\omega}_k$) becomes big following (19) because the activity population during 03:00~09:00 ($\mathbf{y}_{1..6}$) tended to be 0. As a result, NMF6 had a bad performance in the prediction of the remaining time.

where $\hat{\lambda}$ indicates the estimated population for each time unit, defined in (6). A regional event can be detected using this simple irregularity score. If i exceeds the predetermined threshold (positive value), or the actual number is much larger than the predictive one, we define the event as irregular and describe it with red circle in the following figures. In contrast, if i falls below another predetermined threshold (negative value), or the actual number is much less than the predictive one, we also define the event irregular and describe it with blue circle.

We believe such an automatic irregularity detection tool for urban dynamics can be used in map services in next generation web services. As well as SNS data, photos with GPS metadata, search queries, reservation data, route data in car navigation systems, schedules, news, and detected events in terms of irregularities, will be useful for next generation web services.

Fig. 16 and Fig. 17 show the irregularity detection results of two cases: unexpectedly heavy rain in Tokyo and large fireworks event on the Sumida river. Each figure shows the degree of irregularity at the POI. White circles denote little difference between the actual and predicted urban dynamics. Vivid red circles denote that the actual size of the population was much greater than the predicted population, whereas the blues circles denote the opposite situation. Fig. 16 shows that we had sudden heavy rain in Tokyo on July 23rd, 2013, where many blue circles show that the active population greatly decreased due to the rain. Moreover, the area of influence of the rain is clearly shown in Fig. 16. There is no service that can monitor the affected area of a disaster in real time. Such a service will be really useful. In the right of Fig. 17 (Asakusa area), a large fireworks event was held on that day. Our model detected the venue regions since large crowd of spectators gathered to watch the fireworks.

The time series visualization of irregularities can provide us with interesting insights. Fig. 18 shows the urban dynamics on February 23rd, 2014. The largest marathon in Tokyo called the Tokyo Marathon was held on that day. About 10,000 runners started from Shinjuku (green point) at 9:00 and ran to finish line (orange point). In this event, many spectators stood on the side of the road and cheered the runners. The route could be reproduced from the detection results as shown in Fig. 18. The same approach will be useful for the detection of disaster areas, i.e., reproducing a typhoon route.

Visualization of demographics in urban dynamics

The proposed model is flexible in adding urban demographics of each region as external factors that characterize urban dynamics. To construct a model shared among all POIs, we add urban demographics to external factors. We considered two types of urban demographics, land use of an area (*residential, downtown*) and user information (*students, families*). For example, Shinjuku is a *downtown* with *business people*. In this section, we describe the method of adding these demographics into our model as external factor and aims to visualize demographics in urban dynamics. First, urban demographic information is added as a feature to the basic feature discussed



Figure 18. Tokyo Marathon

in *Design of feature*. Then, the urban dynamics of each demographic is visualized by activating the part of d corresponding to that demographic. In contrast to the conventional models with mixture/factorization methods, our model automatically visualizes the urban dynamics corresponding to specific or mixed demographics, whereas the latent patterns should be interpreted and manually labeled in the conventional models. Our model also contribute to the simulation of urban dynamics, which is essential to the urban planning.

Obtaining demographics by crowd sourcing

To obtain the demographics in each region, we used a questionnaire and a crowd sourcing service provided by Yahoo! JAPAN. The questionnaire considered of two types of questions. Regarding land use type, users were asked *Which land use type is the most appropriate to the (target area name)?* Ten options were offered: traffic node, downtown, night spot, amusement district, business district, tourist spot, sightseeing spot, residential district, temple/shrine area, and N/A. Six options were offered as a user information: students, business people, elderly, tourists, families, and N/A. To make answers more reliable, redundant questions were inserted into the queues to verify the crowdworker.

123,753 valid answers were gathered from the crowd sourcing in two weeks. Fig. 19 shows the questionnaire results in terms of land use and user information of five POIs: Shinjuku, Asakusa, Kichijoji, Tokyo Disney Land, Narita airport. Note that the questionnaire was normalized as a multinomial distribution. According to Fig. 19, the characteristics of land use type and user information at each POI followed our intuition (Shinjuku is well known to Japanese as a mixed area; downtown, night spot, and business district, and Asakusa is an area where there are many temples and shrines).

Visualization results

Fig. 20 shows the visualization results of demographics in terms of urban dynamics from land use type and user information viewpoints. The proposed model learned from 180 days of data and 100 POIs at the same time. The nine charts

on the left represent the urban dynamics from land use types, and the five charts on the right side from user information. Different colors show different days of the week.

The visualization results seem to reflect the characteristics of each urban demographic well. For the urban dynamics of each land use type, the following four results were evaluated as qualitatively-correct. (1) Amusement district is the only land use type where the population is larger on weekends than weekdays. (2) Night spot had little difference in population between weekends and weekdays. (3) Business district had a large difference in population between weekdays and weekends, especially during commute time. (4) The population of residential districts decreased during the day in weekdays.

If we obtain user demographics in addition to the location data, the daily rhythm of each user category (e.g. male / female) could be visualized. Two characteristics were observed from the results. (1) Though students and business people had similar dynamics on weekdays, they had totally different dynamics on weekends. Business people did not commute on weekends. On the other hand, students were active even on weekends. (2) In the urban dynamics of tourists, a moderate peak in the morning was found compared to a narrow peak found in the urban dynamics of business people.

These results imply that our model not only accurately predicts the active population but also analyzes the latent factors of urban dynamics both quantitatively and qualitatively.

Extraction of demographics of any POIs

In addition to the application of visualizing demographics, our model can automatically extract urban demographics from mobility records even when no questionnaire records are available. Urban dynamics data can be easily acquired from smartphone GPS logs even if the number of target POIs increases. However, it is not always easy to obtain crowd-sourcing results of new POIs due to financial costs. This application would drastically reduce such costs in investigating the demographics of the interest area. In order to estimate the demographics, we adopted a k-NN method on the basis of

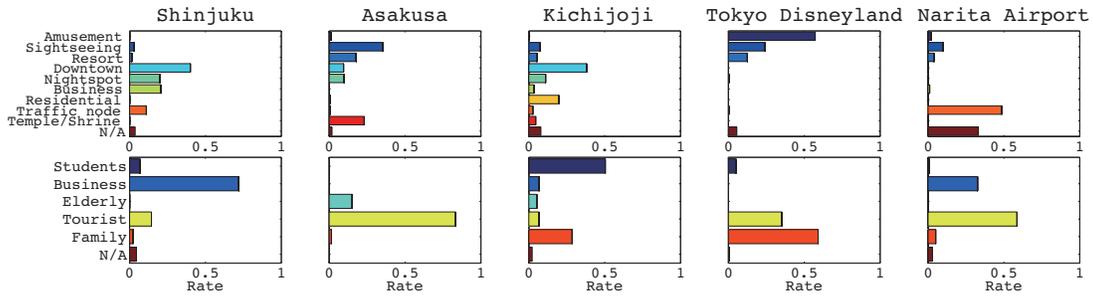


Figure 19. Questionnaire results for land use type and user segment from crowd sourcing. Every POI has different land use pattern and different user information.

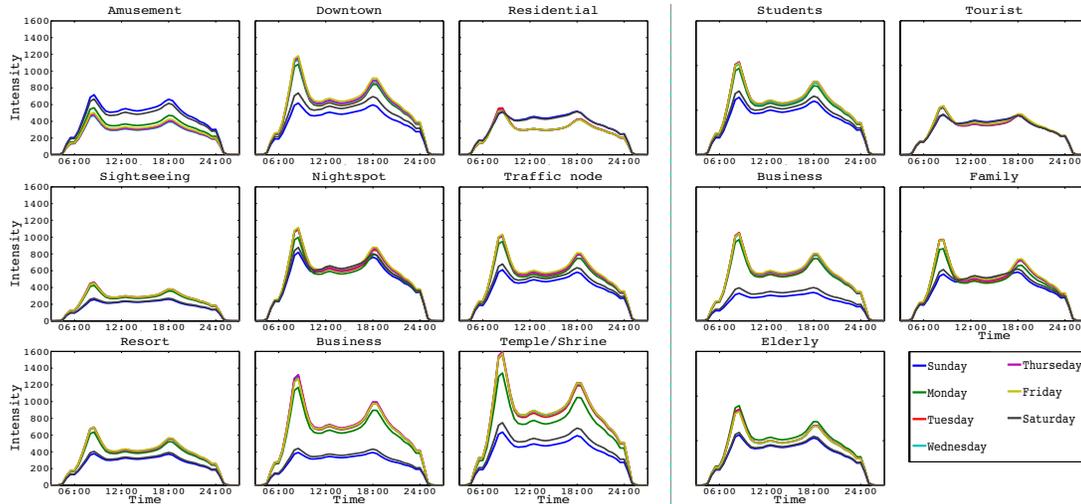


Figure 20. Urban dynamics of each land use type and user information.

the similarity of the urban dynamics. The k-nearest urban demographics are chosen from other POIs to retrieve the urban demographics of the target POI.

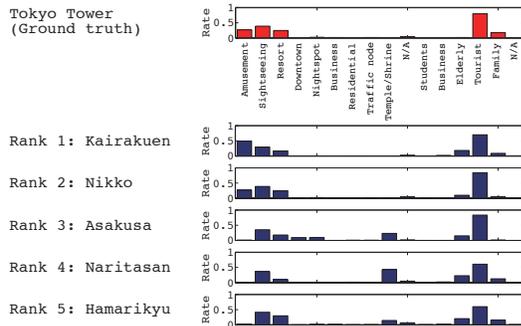


Figure 21. Extraction results of urban demographics in Tokyo Tower area

Fig. 21 shows the actual and top 5 retrieved urban demographics of the Tokyo Tower area obtained. Red bins show the actual urban demographics of the Tokyo Tower area from the result of crowd sourcing. Estimated urban demographics is close to the actual ones.

These extraction results are really meaningful because they suggest that we are released from conducting the market surveys of every POI. As discussed in this section, urban demographics can be reproduced from a massive amount of mobility logs.

CONCLUSION

We proposed a low-rank bilinear Poisson regression model for accurate *ex ante* prediction (ex. predicting urban dynamics for the week ahead). We discussed the usefulness of incorporating external factors, such as day of the week, national holiday, weather condition, urban demographics of the area, in urban dynamics prediction. The results obtained from our experiment with one year’s worth of mobility records confirms the high prediction accuracy of the proposed model. The estimation accuracy of the proposed model is as good as that of conventional models using mixture/factorization methods taking quarter/half of the target day in account. We also confirm the feasibility of three applications including regional event detection based on irregularity, visualization of urban dynamics, and extracting demographics automatically from unknown POIs.

For future work, issues remain on automatic selection or automatic design of the external factors, which is now done by humans. Extensions of the model based on the applications will also be considered. Expanding the weight parameter matrix to a high-order tensor, in other words, extension to a multi-linear model, will make the model more expressive. To stabilize the learning phase, recent nuclear norm minimization methods [16] are attractive as global optimization methods. However, the interpretability of the model should be considered and compared to low-rank approximation. Discussions with governments and industries on urban dynamics will also further promote ubiquitous computing community.

REFERENCES

1. Adams, R., Murray, I., and MacKay, D. Tractable nonparametric Bayesian inference in Poisson processes with Gaussian process intensities. In *Proc. of ICML '09* (2009), 9–16.
2. Alvarez-Lozano, J., García-Macías, J. A., and Chávez, E. Learning and user adaptation in location forecasting. In *Proc. of Ubicomp* (2013), 461–470.
3. Baumann, P., Kleiminger, W., and Santini, S. The influence of temporal and spatial features on the performance of next-place prediction algorithms. In *Proc. of the 2013 ACM International Joint Conference on Pervasive and Ubiquitous Computing* (2013), 449–458.
4. Berry, M., Browne, M., Langville, A., Pauca, P., and Plemmons, R. Algorithms and applications for approximate nonnegative matrix factorization. In *Proc. of CSDA '06* (2006), 155–173.
5. Candia, J., Gonzalez, M., Wang, P., Schoenharl, T., Madey, G., and Barabasi, A.-L. Uncovering individual and collective human dynamics from mobile phone records. *Journal of Physics A: Mathematical and Theoretical* 41 (2008), 1–11.
6. Chan, A., and Vasconcelos, N. Bayesian Poisson regression for crowd counting. In *Proc. of ICCV '09* (2009), 545–551.
7. Cheng, C., Yang, H., Lyu, M., and King, I. Where you like to go next: Successive point-of-interest recommendation. In *Proceedings of the Twenty-Third International Joint Conference on Artificial Intelligence, IJCAI'13* (2013), 2605–2611.
8. Chon, Y., Kim, S., Lee, S., Kim, D., Kim, Y., and Cha, H. Sensing wifi packets in the air: Practicality and implications in urban mobility monitoring. In *Proc. of Ubicomp* (2014), 189–200.
9. Dai, L., Qin, W., Xu, H., Chen, T., and Qian, C. Urban traffic flow prediction: A mapreduce based parallel multivariate linear regression approach. In *Proc. of ITSC '14* (2014), 2823–2827.
10. Eagle, N., and Pentland, A. Reality mining: Sensing complex social systems. *Personal Ubiquitous Comput.* (2006), 255–268.
11. Eagle, N., and Pentland, A. Eigenbehaviors: identifying structure in routine. *Behavioral Ecology and Sociobiology* 63, 11 (2009), 1689–1689.
12. Fan, Z., Song, X., and Shibasaki, R. Cityspectrum: A non-negative tensor factorization approach. In *Proc. of UbiComp '14* (2014), 213–223.
13. Gao, H., Tang, J., Hu, X., and Liu, H. Modeling temporal effects of human mobile behavior on location-based social networks. In *Proc. of CIKM '13* (2013), 1673–1678.
14. Ihler, A., Hutchins, J., and Smyth, P. Adaptive event detection with time-varying Poisson processes. In *Proc. of KDD '06* (2006), 207–216.
15. Ihler, A., and Smyth, P. Learning time-intensity profiles of human activity using non-parametric Bayesian models. In *Proc. of NIPS '07* (2007), 625–632.
16. jui Hsieh, C., and Olsen, P. Nuclear norm minimization via active subspace selection. In *Proc. of ICML '14* (2014), 575–583.
17. Ma, L., Yan, X., and Qiao, W. A quasi-Poisson approach on modeling accident hazard index for urban road segments. *Discrete Dynamics in Nature and Society* (2014), 1–8.
18. Miller, A., Bornn, L., Adams, R., and Goldsberry, K. Factorized point process intensities: A spatial analysis of professional basketball. In *Proc. of ICML '14* (2014), 235–243.
19. Nishi, K., Tsubouchi, K., and Shimosaka, M. Extracting land-use patterns using location data from smartphones. In *Proc. of URB-IOT '14* (2014), 38–43.
20. Nishi, K., Tsubouchi, K., and Shimosaka, M. Hourly pedestrian population trends estimation using location data from smartphones dealing with temporal and spatial sparsity. In *Proc. of SIGSPATIAL '14* (2014).
21. Phithakkitnukoon, S., Horanont, T., Di Lorenzo, G., Shibasaki, R., and Ratti, C. Activity-aware map: Identifying human daily activity pattern using mobile phone data. In *Proc. of HBU'10* (2010), 14–25.
22. Reades, J., Calabrese, F., and Ratti, C. Eigenplaces: analysing cities using the space time structure of the mobile phone network. *Environment and Planning B: Planning and Design* 36, 5 (2009), 824–836.
23. Sakaki, T., Okazaki, M., and Matsuo, Y. Earthquake shakes twitter users: Real-time event detection by social sensors. In *Proc. of WWW'10* (2010), 851–860.
24. Scott, J., Brush, B., Krumm, J., Meyers, B., Hazas, M., Hodges, S., and Villar, N. Preheat: Controlling home heating using occupancy prediction. In *Proc. of UbiComp '11* (2011), 281–290.
25. Sekimoto, Y., Shibasaki, R., Kanasugi, H., Usui, T., and Shimazaki, Y. PFlow: Reconstructing people flow recycling large-scale social survey data. In *Proc. of Pervasive '11* (2011), 27–35.
26. Shimazaki, H., and Shinomoto, S. Kernel bandwidth optimization in spike rate estimation. *J. Comput. Neurosci.* 29, 1-2 (2010), 171–182.
27. Shimosaka, M., Ishino, T., Noguchi, H., Sato, T., and Mori, T. Detecting human activity profiles with Dirichlet enhanced inhomogeneous Poisson processes. In *Proc. of ICPR '10* (2010), 4384–4387.
28. Song, X., Zhang, Q., Sekimoto, Y., Horanont, T., Ueyama, S., and Shibasaki, R. Modeling and probabilistic reasoning of population evacuation during large-scale disaster. In *Proc. of KDD '13* (2013), 1–13.

29. Sun, L., Lee, D.-H., Erath, A., and Huang, X. Using smart card data to extract passenger's spatio-temporal density and train's trajectory of MRT system. In *Proc. of UrbComp '12* (2012), 2154–2158.
30. Takeuchi, K., Tomioka, R., Ishiguro, K., Kimura, A., and Sawada, H. Non-negative multiple tensor factorization. In *Proc. of ICDM '13* (2013), 1199–1204.
31. Tominaga, S., Shimosaka, M., Fukui, R., and Sato, T. A unified framework for modeling and predicting going-out behavior. In *Proc. of Pervasive '12* (2012), 73–90.
32. Witayangkurn, A., Horanont, T., Sekimoto, Y., and Shibasaki, R. Anomalous event detection on large-scale GPS data from mobile phones using hidden Markov model and cloud platform. In *Proc. of UbiComp '13* (2013), 1219–1228.
33. Zhang, F., Wilkie, D., Zheng, Y., and Xie, X. Sensing the pulse of urban refueling behavior. In *Proc. of UbiComp '13* (2013), 13–22.
34. Zheng, Y., Liu, T., Wang, Y., Zhu, Y., Liu, Y., and Chang, E. Diagnosing New York city's noises with ubiquitous data. In *Proc. of UbiComp '14* (2014), 715–725.