Simultaneous multiple POI population pattern analysis system with HDP mixture regression

Yuta Hayakawa[1(✉)], Kota Tsubouchi[2], and Masamichi Shimosaka[1]

[1] Tokyo Institute of Technology, Tokyo, Japan
{hayakawa,simosaka}@miubiq.cs.titech.ac.jp
[2] Yahoo Japan Corporation, Tokyo, Japan
ktsubouc@yahoo-corp.jp

Abstract. In recent years, the use of smartphone Global Positioning System (GPS) logs has accelerated the analysis of urban dynamics. Predicting the population of a city is important for understanding the land use patterns of specific areas of interest. The current state-of-the-art predictive model is a variant of bilinear Poisson regression models. It is independently optimized for each point of interest (POI) using the GPS logs captured at that single POI. Thus, it is prone to instability during fine-scale POI analysis. Inspired by the success of topic modeling, in this study, we propose a novel approach based on the hierarchical Dirichlet process mixture regression to capture the relationship between POIs and upgrade the prediction performance. Specifically, the proposed model enables mixture regression for each POI, while the parameters of each regression are shared across the POIs owing to the hierarchical Bayesian property. The empirical study using 32 M GPS logs from mobile phones in Tokyo shows that our model for large-scale finer-mesh analysis outperforms the state-of-the-art models. We also show that our proposed model realizes important applications, such as visualizing the relationship between cities or abnormal population increase during an event.

Keywords: First keyword · Second keyword · Another keyword.

1 Introduction

The analysis of urban dynamics is of significant importance in urban planning, especially in the placement of restaurants, shopping centers, or open spaces, and for local services such as transportation. Owing to the increase in smartphone usage, extensive Global Positioning System (GPS) logs, which are sufficient to reflect real-world population flow, are stored. Hence, several researchers have recently begun to analyze urban dynamics using these GPS logs.

For example, Fan et al. [4] assumed that population flow patterns represent the features of the cities and extracted patterns from GPS logs. This pattern extraction is an important approach in urban dynamics analyses, and has been studied widely [8, 13]. The extracted patterns help in understanding the urban characteristics of the city, and improve planning efficiency for store openings or local commercial distributions.

Population prediction is another important topic in urban dynamics research. Shimosaka et al. [12] predicted active populations in the areas including big stations and amusement parks using external variables such as weather or day of the week. These predictions enable practical applications such as traffic prediction [9], sales prediction based on external variables, and the detection of anomalies by evaluating the difference between a predicted and an actual population.

Moreover, predicting or interpreting urban dynamics in small individual meshes (e.g., $200\,\mathrm{m} \times 200\,\mathrm{m}$) enables effective responses to social needs, such as promotion of location-specific products and services . In the state-of-the-art research [12], they independently constructed a predictive model for each point of interest (POI). However, it is difficult for the model to predict urban dynamics accurately in the smaller meshes. This is because the observed number of logs can be affected by its noise and increasing the sample size can suppress the effects of the noise but it is not always possible to achieve such an increase.

However, similarity of functions in various cities can be used to overcome this limitation. Cities may have certain functions in common. For example, if a considerable part of one city is a business district and some parts of another city are also business districts, it means that these two cities have the same urban function. It can be assumed that such cities with certain similar functions share similar urban dynamics. Therefore, to effectively learn the predictive model for a city, data from other cities having partially similar functions to one of the functions of the target city can be utilized.

Inspired by the success of hierarchical Bayesian models that extract latent urban dynamics patterns across cities [8, 13], in this study, we propose a hierarchical Dirichlet process (HDP) regression mixture model that utilizes data of other areas having functions similar to those of the target city to achieve stable prediction in small areas This utilization of data from other areas virtually increases the sample size for learning a predictive model of the target area and stabilizes the learning even for small areas. The HDP regression mixture model also achieves parameter reduction. This is because the model for each city consists of a mixture of latent models; only the mixture coefficients and the parameters of latent models are learned.

The contributions of this work are as follows.

- We proposed the HDP regression model to make stable predictions of urban dynamics even for smaller areas by utilizing the data in the other areas and reduce the total parameter size by sharing the parameters among the cities.
- We conducted experiments using a real-world dataset, 32 M GPS logs from smartphones, to show that the proposed model predicts the urban dynamics more accurately than previous predictive models.
- We show two important industrial applications that can be realized using the model; detection of the abnormal congestion of an event and a visualization of the mixing coefficient of our proposed model to better understand the relationship.

2 Related Work

In this section, we describe work related to urban dynamics modeling. Research on extracting urban dynamics patterns have been conducted actively in recent years. In the present work, tensor factorization [21, 14, 4, 19, 18] or mixture modeling [8, 13] has been used to extract patterns. In the tensor factorization approaches, many researchers have modeled the urban dynamics data as a tensor with cities, time of the day, and date axes, and they extract latent patterns by factorization. For example, claims of city noise [21], check-in activities [14], and GPS logs [4] were modeled as a tensor to analyze their latent patterns. Nishi et al. [8] modeled the active population transition in one day using the Dirichlet mixture model and extracted the latent patterns shared by cities. Shimosaka et al. [13] also extracted patterns by mixture modeling. Moreover, they simultaneously clustered the cities using their proposed hierarchical Bayesian framework. This clustering provides an explicit understanding of the similarities between cities. However, these to extract latent patterns in past datasets do not predict future urban dynamics.

Several researchers have attempted to predict the activities in cities using external variables [17, 2, 12]. Wang et al. [17] proposed the negative binomial regression model by using external variables such as population or weather to predict traffic volume. Bogomolov et al. [2] constructed a model to predict the number of crimes using a random forest with demographic information. The research by Shimosaka et al. [12] is state-of-the-art research in urban dynamics prediction.

However, they constructed predictive models for each city using data specific to that city. Therefore, learning may be unstable, particularly for small areas. To solve these problems, some studies[20, 11] that utilized data from other cities to learn the model of the target area and improve accuracy. Zheng et al. [20] modeled the flow of people in the cities by a convolutional neural network (CNN). They trained the model while sharing the dataset between neighboring areas by convolution. However, the CNN only shared the data were among neighboring areas. Shimosaka et al. [11] also proposed a predictive model called SPF, which shared parameters between meshes while retaining spatial preservation and reduced the number of parameters using a factorization approach. In terms of model expression, however, the method proposed herein is a mixture regression and can model complex multimodal data more effectively than SPF, which is a single regression model.

3 Urban population pattern analysis system with HDP regression

This section presents an outline of the proposed urban population pattern analysis system with HDP regression using GPS logs from smartphones. Further, the applications realized by the proposed system are discussed. As shown on the left side of the system diagram in Fig. 1, the number of GPS logs from smartphone use is sufficient to represent the populations of a city, thus we regard the counting logs as the population in each area. These population counts, as well as external factors such as weather and holiday information, were used as the datasets for the urban dynamics prediction system. In particular, for the population count, one day was segmented into S parts; then, it was assumed that the number of GPS logs in area l within the τth time segment was an active population. We modeled the transition of the active population $y^{(l)}_{\boldsymbol{c},\tau,n}$ of the n-th day with conditions $\boldsymbol{c}$ (e.g., weather, day of the week, national holiday) through the time segments $\tau = \{1, ..., S\}$. The transition of the population is commonly used for the application of urban dynamics analyses. The proposed model realizes three main areas of urban dynamics: population prediction, anomaly detection, and inter-city relationship analysis that are widely studied by researchers. In the following subsections, we describe the detailed setting of each application.

Active population prediction is the problem of predicting the active population $y^{(l)}_{\boldsymbol{c},\tau,n}$ in the area l, in the τ-th time segment on the n-th day using external factors $\boldsymbol{c}$. Active population prediction helps to provide real-world applications such as predictions of sales or traffic volume. As can be expected, prediction accuracy is quite important for the quality of the application. However, prediction accuracy can be unstable owing to the lack of a sufficient sample size for training, particularly in fine-grained meshes. Thus, it is necessary to share a dataset over the meshes to enable improved prediction.

Anomaly detection is one of the hot topics in urban dynamics research [10, 6, 12]. It is also useful in several practical applications, including the detection of city events and traffic obstacles. One method for detecting anomalies in the population is evaluating a difference between a predicted and an actual population. Some studies [6, 12] evaluated the anomalous population counts in a city based on the irregularity index, $\frac{1}{\hat{y}^{(l)}_{\boldsymbol{c},\tau}}(y^{(l)}_{\boldsymbol{c},\tau,n} - \hat{y}^{(l)}_{\boldsymbol{c},\tau})$,, where $y^{(l)}_{\boldsymbol{c},\tau,n}$ is an actual population and $\hat{y}^{(l)}_{\boldsymbol{c},\tau}$ is the prediction. Owing to the accurate population estimate by the bilinear Poisson regression, they successfully detected or predict a large event and unexpected heavy rain. Note that the anomaly detection significantly depends on the prediction accuracy; unstable prediction in the fine-grained mesh leads to impractical anomaly detection. Thus, stable prediction is essential for anomaly detection in fine-grained meshes.

Understanding the relationships between the different areas of a city is another important aspect of urban dynamics analysis. Finding areas that have similar features within a region is helpful in deciding the placement of new commercial entities such as restaurants or stores. As mentioned above, the relationships between areas were investigated based on pattern extraction using generative models in previous research [8, 13], and these aspects of generative models that is extracting latent pattern is helpful for updating predictive performance. We proposed the combination model of a generative and discriminative model described in the following section.

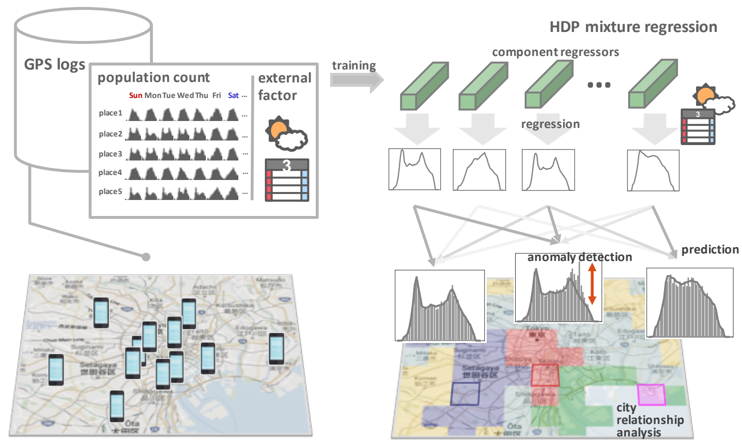

Fig. 1. Urban population pattern analysis system proposed in this paper.

4 Definition of proposed method: HDP mixture regression

In this research, we proposed the regression mixture model for accurate prediction in fine-grained meshes such as $200\,\mathrm{m} \times 200\,\mathrm{m}$ mesh while the existing state-of-the-art model [12] predicts in $900\,\mathrm{m} \times 900\,\mathrm{m}$ mesh. A model for an area consists of a mixture of latent models shared by all cities; each area has a specific mixing coefficient. LDA [1] and HDP [15] are widely known methods for latent allocation. We proposed a predictive model using HDP mixture regression, which uses the HDP as the prior of the mixture model.

In HDP mixture regression, the model is learned by Bayesian inference. However, it is difficult to analytically infer the Poisson regression under Bayesian inference. To infer the Poisson regression in the Bayesian framework, the Poisson distribution can be approximated by a Gaussian distribution [3]. Approximation around zero is inadequate and can result in poor prediction. Here, Gaussian regression, which can be analytically applied to Bayesian inference, is used. We confirm the prediction accuracy of the Gaussian regression, as well as the Poisson regression in the experiment described in section V. This section first explains prediction by the state-of-the-art model (bilinear Poisson regression) and then describes prediction by the proposed HDP mixture regression.

4.1 Urban dynamics prediction by bilinear Poisson regression

As mentioned earlier, the predictive model using bilinear Poisson regression[12] is the state-of-the-art model in urban dynamics prediction research. They assumed that the active population $y^{(l)}_{\boldsymbol{c},\tau,n}$ follows the Poisson distribution $\mathcal{P}(y^{(l)}_{\boldsymbol{c},\tau,n}|\lambda^{(l)}_{\boldsymbol{c},\tau})$, where $\lambda^{(l)}_{\boldsymbol{c},\tau}$ is the mean parameter of the Poisson distribution. They considered a combination of the time feature, which drastically influences the active population, with other features. This enabled them to model the differences between the patterns under different conditions, such as weekday and weekend.

In the bilinear Poisson regression, the parameter of the Poisson distribution $\lambda_{\boldsymbol{c},\tau}^{(l)} > 0$ is represented using the weight matrix $\boldsymbol{W}_l \in \mathbb{R}^{M \times S}$, the time feature $\boldsymbol{\phi}(\tau) \in \mathbb{R}^S$, and $\boldsymbol{\varphi}(\boldsymbol{d}) \in \mathbb{R}^M$ as $\ln \lambda_{\boldsymbol{c},\tau}^{(l)} = \boldsymbol{\varphi}(\boldsymbol{d})^\top \boldsymbol{W}_l \boldsymbol{\phi}(\tau)$. Shimosaka et al. [12] reduced the rank of the weight matrix by decomposing the weight matrix as $\boldsymbol{W}_l = \boldsymbol{U}_l \boldsymbol{V}_l^\top$, where $\boldsymbol{U}_l \in \mathbb{R}^{M \times K}$ and $\boldsymbol{V}_l \in \mathbb{R}^{S \times K}$ are the decomposed matrices satisfying the condition $K \ll M, K \ll S$. Although this low-rank matrix reduces the risk of overfitting, the model only uses the dataset in a single area. This results in an insufficient sample size as shown in the next subsection. Moreover, the total number of parameters across all areas is $K(M+S)L$ where L is the number of areas. The parameters will increase linearly with the number of areas, and this large number of parameters causes the instability in learning.

4.2 Definition of HDP mixture regression

We formulate herein the HDP mixture regression to utilize the data from other areas to learn the model of the target area while suppressing the increase in the number of parameters. Following the work by Wang et al. [16], we used Sethuraman's construction was used to represent the stick-breaking process that realizes the HDP. The latent variable $z_{l,n,m}$ represents an active population $\boldsymbol{y}_{\boldsymbol{c}}^{(n,l)}$ as defined in the previous chapter, and is assigned to the cluster m. It follows the condition $z_{l,n,m} \in \{1,0\}, \sum_m z_{l,n,m} = 1$ and m is an area-level cluster. This area-level cluster is defined to analytically apply the variational inference as the document-level cluster in the previous work [16]. $z_{l,n,m} = 1$ indicates that the active population in the area l on the day n is assigned to cluster m. The latent variable $r_{l,m,k}$ represents the correspondence between an area-level cluster m of the area l and a global cluster k. The global clusters are shared between areas; $r_{l,m,k}$ also follows the requirement $r_{l,m,k} \in \{1,0\}, \sum_k r_{l,m,k} = 1$. $r_{l,m,k} = 1$ indicates the area-level cluster m of the area l, and corresponds to the global cluster k.

To utilize the data from other areas focusing on the peak of the urban dynamics pattern rather than the volume of the active population, we normalize the active population in the learning process as $\tilde{y}_{\boldsymbol{c},\tau}^{(n,l)} = \frac{1}{\eta_l} y_{\boldsymbol{c},\tau,n}^{(l)}$, where η_l is calculated by $\eta_l = \frac{1}{N}\sum_{n=1}^{N}\sum_{\tau=1}^{T} y_{\boldsymbol{c},\tau,n}^{(l)}$ using the training dataset. The joint distribution $\tilde{y}_{\boldsymbol{c},\tau}^{(n,l)}$, $z_{l,n,m}$, $r_{l,m,k}$ is represented as follows,

$$p(\tilde{y}_{\boldsymbol{c},\tau}^{(n,l)}, z_{l,n,m}, r_{l,m,k}) = \prod_m \pi_{l,m}^{z_{l,n,m}} \prod_k \rho_k^{r_{l,m,k}} \mathcal{N}(\tilde{y}_{\boldsymbol{c},\tau}^{(n,l)} | \lambda_{\boldsymbol{c},\tau,k}, \sigma_k^2)^{z_{l,n,m} r_{l,m,k}} \tag{1}$$

where $\pi_{l,m}$ is the mixing coefficient for the area l, and ρ_k is the global mixing coefficient. $\lambda_{\boldsymbol{c},\tau,k} = \boldsymbol{\varphi}(\boldsymbol{c})^\top \boldsymbol{W}_k \boldsymbol{\phi}(\tau)$, where $\boldsymbol{\phi}(\tau)$ is the time feature and $\boldsymbol{\varphi}(\boldsymbol{c})^\top$ is the feature with other conditions. $\pi_{l,m}$ and ρ_k are generated by the stick-breaking process as follows,

$$p(\pi'_{l,m}) = \mathcal{B}(1, \beta_0), p(\rho'_k) = \mathcal{B}(1, \gamma_0)$$
$$\pi_{l,m} = \pi'_{l,m} \prod_{s=1}^{t-1} (1 - \pi'_{l,s}), \rho_k = \rho'_k \prod_{j=1}^{k-1} (1 - \rho'_j) \tag{2}$$

The priors of $z_{l,n,m}$, $r_{l,m,k}$, and the weight parameter $\boldsymbol{W}_k$ are represented as follows,

$$p(\boldsymbol{z}_l|\boldsymbol{\pi}_l) = \prod_{n,m} \pi_{l,m}^{z_{l,n,m}}, p(\boldsymbol{r}) = \prod_{l,m,k} \rho_k^{r_{l,m,k}}$$
$$p(\boldsymbol{W}_k) = \mathcal{N}(\text{vec}(\boldsymbol{W}_k)|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k), p(\sigma^2) = \text{Gamma}(\sigma^2|a_0, b_0), \tag{3}$$

where vec$(\cdot)$ is the vectorization of the matrix. For example, for the $A = [\boldsymbol{a}_1, \boldsymbol{a}_2, ...\boldsymbol{a}_K]$, $\text{vec}(A) = [\boldsymbol{a}_1^\top, \boldsymbol{a}_2^\top, ...\boldsymbol{a}_K^\top]^\top$. Generally, the posterior of the parameters or latent variables cannot be analytically estimated in the model using HDP. A sampling method such as Gibbs sampling [7, 5] or variational inference [16] is used to estimate the posterior approximately. In this research, we use variational inference to estimate the approximated posterior of the parameters and latent variables.

Our proposed mixture model using HDP as a prior reduces the number of parameters compared to the previous method. As mentioned earlier, in the previous predictive method [12], the number of parameters is $K(M+S)L$ where K is the dimension of the low-rank matrix, M is the dimension of the time feature, and S is the dimension of the other feature. If we set $L = 100 \times 100 = 10000$, $M = 48$, $S = 28$, and $K = 5$, the number of all the parameters is 3.8M. In our proposed method, the number of parameters is dependent on the number of global clusters B. Considering the number of mixing coefficients, the total number of parameters is $BMS + LT + B$ where B is the maximum number of global clusters and T is the maximum number of area-specific clusters. In the aforementioned setting, the number of parameters is suppressed to 567K with the setting $B = T = 50$.

4.3 Prediction by HDP mixture regression

It is assumed that the posterior estimated with the dataset $\boldsymbol{Y} = \{\tilde{y}_{\boldsymbol{c},\tau}^{(n,l)}\}_{n,l,\tau}$ in all areas is approximated as follows,

$$p(\boldsymbol{\rho}, \boldsymbol{\pi}, \boldsymbol{r}, \boldsymbol{z}, \boldsymbol{W}, \boldsymbol{\sigma}|\boldsymbol{Y}) = q(\boldsymbol{\rho})q(\boldsymbol{\pi})q(\boldsymbol{r})q(\boldsymbol{z})q(\boldsymbol{W}, \boldsymbol{\sigma}). \tag{4}$$

The predictive distribution of the HDP mixture regression is a mixture and is multimodal. The predictive distribution $p^*(\tilde{y}_{\boldsymbol{c},\tau}^{(l)*}|\boldsymbol{Y})$ and the prediction value of the model $\hat{y}_{\boldsymbol{c},\tau}^{(l)*}$ are defined as follows

$$p^*(\tilde{y}_{\boldsymbol{c},\tau}^{(l)*}|\boldsymbol{Y}) \simeq \sum_m \mathbb{E}_{q(\boldsymbol{\pi})}[\pi_{l,m}] \sum_k \mathbb{E}_{q(\boldsymbol{r})}[r_{l,m,k}] \mathbb{E}_{q(\boldsymbol{W},\boldsymbol{\sigma})}[p(\tilde{y}_{\boldsymbol{c},\tau}^{(l)*}|\lambda_{\boldsymbol{c},\tau,k}, \sigma_k^2)], \tag{5}$$
$$\hat{y}_{\boldsymbol{c},\tau}^{(l)*} = \eta_l \arg\max_{\tilde{y}_{\boldsymbol{c},\tau}^{(l)*}} (p^*(\tilde{y}_{\boldsymbol{c},\tau}^{(l)*}|\boldsymbol{Y})). \tag{6}$$

4.4 Urban dynamics prediction systems using HDP mixture regression

We describe the outline of the urban dynamics prediction system using the proposed model, the HDP mixture regression shown in Fig. 1. The proposed method shares the datasets of all areas of interest during the training (see the upper part of the figure). Because only the component regressors (except for the mixing coefficients) have parameters, the proposed method can suppress the number of parameters (see the upper right part of the figure). As shown in the right lower part of Fig. 1, the HDP mixture regression model realizes important applications. The model provides a prediction for each area because each area has its own mixing coefficient (see the right part of the figure). Consequently, owing to the prediction, the model can also realize anomaly detection for each area. The mixture coefficients represent the manner in which the model for each area depends on each component regressor; thus, users can find the similarities between cities in terms of active population transition by visualizing the value of one of the mixture coefficients.

5 Experimental results

To evaluate the performance of the proposed urban dynamics prediction method, we conducted experiments comparing the proposed method with existing methods used to model urban dynamics.

5.1 Dataset

This experiment utilized the GPS logs obtained by the smartphone app *Bosai Sokuho*[1] released by Yahoo Japan Corporation. The GPS logs were collected from users across Japan who consented to providing their location information, and they have all been anonymized. In the Kanto region alone, 15 M logs were collected per day. Each GPS log includes a time stamp, longitude, and latitude and is collected when the user moves. Thus, the logs represent human activities. Logs collected from July 1st, 2013 to June 30th, 2014 were used in this experiment. The number of logs in each mesh in the 3 km $\times$ 3 km square area shown in Fig. 2 were counted at 30 minute intervals. We used the dataset for this experiment. In this experiment, we divided the target area into two different sizes of meshes: one was 600 m $\times$ 600 m mesh and the other was 200 m $\times$ 200 m mesh as shown in Fig. 2.

5.2 Evaluation metric

We used Mean Negative Log Likelihood (MNLL), Mean Absolute Error (MAE) as evaluation metrics. They were also used in the existing state-of-the-art research [12]. MNLL is defined as $\mathrm{MNLL} = \frac{1}{NT}\sum_{n=1}^{N}\sum_{\tau=1}^{T}(-\ln p(y^{(l)}_{\mathbf{c},\tau,n}|\lambda^{(l)}_{\mathbf{c},\tau}))$, MAE is defined as $\mathrm{MAE} = \frac{1}{NT}\sum_{n=1}^{N}\sum_{\tau=1}^{T}|y^{(l)}_{\mathbf{c},\tau,n} - \hat{y}^{(l)}_{\mathbf{c},\tau}|$.

[1] https://emg.yahoo.co.jp/

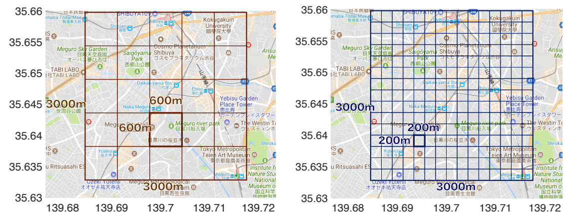

Fig. 2. Left: First mixing coeffieint in each mesh l, right: Second mixing coefficient in each mesh l

5.3 Comparison methods

We compared the proposed HDP regression mixture model (HDP-reg) with the bilinear Poisson regression model (BP) [12], and the bilinear Gaussian regression model (BG) , and SPF [11] in terms of MAE and MNLL. The bilinear Gaussian regression model used in this experiment, models urban dynamics with Gaussian distribution as the HDP-reg does, and it was used to compare with MNLL fairly. We also verified that the bilinear Gaussian regression model has accuracy equivalent to that of the bilinear Poisson regression model to confirm the validity of the comparison between HDP-reg and the bilinear Gaussian regression model.

In the bilinear Gaussian regression setting, we assumed population count $y^{(l)}_{\boldsymbol{c},\tau,n}$ was sampled from a Gaussian distribution, $\mathcal{N}(y^{(l)}_{\boldsymbol{c},\tau,n}|\mu^{(l)}_{\boldsymbol{c},\tau},\sigma^2)$ and the mean parameter $\mu^{(l)}_{\boldsymbol{c},\tau}$ was estimated by the bilinear form, $\hat{\mu}^{(l)}_{\boldsymbol{c},\tau} = \boldsymbol{\varphi}(\boldsymbol{d})^\top \boldsymbol{W}^{(l)}\boldsymbol{\phi}(\tau)$. We conducted the experiments using two types of bilinear regression (Poisson or Gaussian) as shown below,

1. **BP / BG 1 for All**: One regressor for all meshes. These models can be learned using all the datasets in all meshes, and this can stablize the learning. However, these models provide the same predictive results for All meshes, which implies that they cannot represent the differences between cities.
2. **BP / BG 1 for 1**: One regressor for each mesh. Each model can model the characteristics of the urban dynamics for each mesh; however, the dataset used during the learning is specific to the respective mesh, and this small dataset can cause overfitting.

We use the expectation of the model as the prediction value for the bilinear Poisson/Gaussian regression model. In this experiment, we used two types of features; one is the time feature and the other is the weekday feature.

5.4 Comparison with previous predictive methods

We compared our proposed model with previous predictive models using datasets in two different meshes, as shown in Fig. 2. MAE and MNLL were used as metrics on the five-fold cross-validation in this experiment. We did not compare the MNLL between the bilinear Gaussian regression and the bilinear Poisson

Table 1. Comparison on predictive metrics

mesh size	600m × 600m		200m × 200m	
	MAE	MNLL	MAE	MNLL
BG 1 for All	26.0 ± 8.6	1.55 ± 0.25	3.73± 0.83	2.14 ± 0.19
BP 1 for All	26.0 ± 8.6	[1] 10.0 ± 3.4	3.73 ±0.83	[1] 2.88 ± 0.45
BG 1 for 1	24.2 ± 8.7	1.44 ± 0.27	3.45 ± 0.83	1.94 ± 0.17
BP 1 for 1	23.8 ± 8.9	[1] 8.73 ± 3.3	3.42 ± 0.86	[1] 2.67 ± 0.42
SPF	24.4 ± 8.7	[1] 9.68 ± 3.1	3.44 ± 0.83	[1] 2.80 ± 0.39
Proposed	**23.1 ± 7.2**	**1.37 ± 0.25**	**3.38 ± 0.68**	**1.92 ± 0.24**

regression because each measurement in each model was different. We used the 30-day dataset for training and the 180-day dataset for testing.

The experimental results are shown in Table 1. With this experiment, we confirmed that the performance of BP and BG on MAE was almost equivalent. This result ensured the validity of comparison of HDP-reg and BG. The performance of the proposed model was better than any of the bilinear regression models in MAE The smaller mesh size made the number of datasets in each mesh small, and made the MAE smaller. Proportionally, the performance difference in the 200 m meshes was bigger than for the 600 m meshes. The performance of HDP-reg in the MNLL was better than those of BG 1 for '1' and BG 1 for All.

5.5 Application using HDP mixture regression

In this section, we show that the proposed model has the potential to realize the applications mentioned in Section III, especially, Anomaly detection and City relationship analysis.

We evaluated the congestion caused by cherry blossom viewing using prediction by HDP mixture regression. The actual populations and predictions every Saturday from March 8 to April 12 around Nakameguro Station (Tokyo, Japan) are shown in Fig. 3. It is famous for cherry blossom viewing along the Meguro River. Cherry blossoms come into full bloom from the end of March to the beginning of April and the figure also shows that the congestion on April 5 is the heaviest. This abnormal congestion can be automatically detected by setting the threshold of the anomaly metric shown in (3). This anomaly detection also can be used for evaluating the effect of the events in terms of the increase or decrease of visits, compared to normal.

We also visualized the mixing coefficients for understanding the relationship between meshes. Mixing coefficients represent how each mesh area relies on each component regressor, and the cities that have similar mixing coefficients have similar active population transition. The mixing coefficient for the visualization was calculated as $\zeta_{l,k} = \sum_m \mathbb{E}_{q(\boldsymbol{\pi})}[\pi_{l,m}]\mathbb{E}_{q(\boldsymbol{r})}[r_{l,m,k}]$. $\zeta_{l,k}$ represents how a mesh area l relies on k-th component regressor. The left figure of Fig. 4 represents the $k = 1$-st and $k = 2$-nd patterns from $K = 12$ patterns. The right figure of Fig. 4

[1] The performance of bilinear Poisson regression in MNLL should be compared only between BP 1 for '1' and BP 1 for All because the measurement of a Poisson distribution and that of a Gaussian distribution are different.

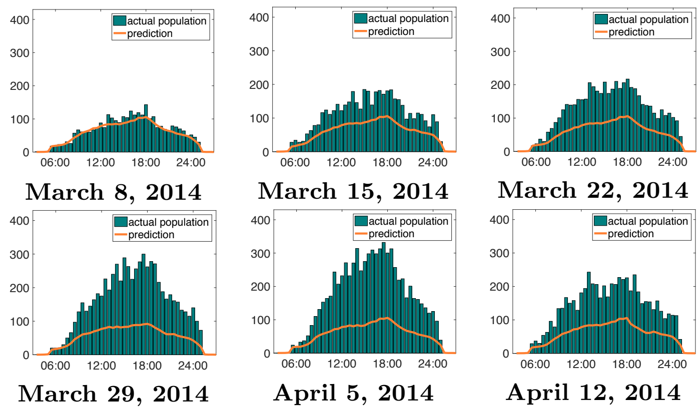

Fig. 3. The actual populations and predictions in cherry blossom season in Tokyo.

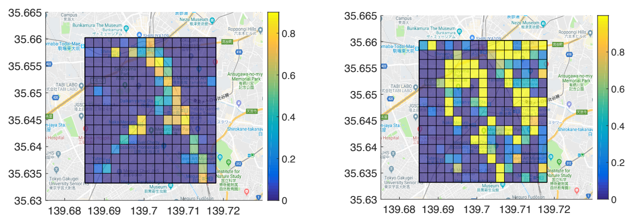

Fig. 4. left: 1-st mixing coeffieint in each mesh l, right: 2-nd mixing coefficient in each mesh l

shows a large number of models because the areas on the railway or station area rely on the $k = 1$-st component regressor. The $k = 2$-nd components are relied on by the downtown areas around stations. From these visualizations, it can be said that the proposed model can represent the relationship between mesh areas.

6 Conclusion

In this research, we modeled urban dynamics on large-scale high-resolution meshes using GPS logs of mobile phones for urban dynamics analysis systems. To predict future population stably even in smaller areas, we proposed using an HDP mixture regression model that uses the datasets in all the meshes for the training phase and predicted the urban dynamics for each mesh stably. We conducted the experiments using smartphone GPS logs in 3 km $\times$ 3 km squares in the Tokyo region, and the number of datasets was 32 M. The proposed model was compared with an state-of-the-art model and achieved an MAE improvement of 1.3. We also showed the two types of applications: evaluating anomalous congestion caused by the cherry blossom viewing activities and visualizing the coefficients to understand the relationship between mesh areas.

References

1. Blei, D.M., Ng, A.Y., Jordan, M.I.: Latent dirichlet allocation. Journal of machine Learning research (2003)
2. Bogomolov, A., Lepri, B., Staiano, J., Oliver, N., Pianesi, F., Pentland, A.: Once upon a crime: towards crime prediction from demographics and mobile data. In: Proc. of ICMI (2014)
3. Chan, A.B., Vasconcelos, N.: Bayesian poisson regression for crowd counting. In: Proc. of ICCV (2009)
4. Fan, Z., Song, X., Shibasaki, R.: CitySpectrum: A non-negative tensor factorization approach. In: Proc. of UbiComp (2014)
5. Ishwaran, H., James, L.F.: Approximate dirichlet process computing in finite normal mixtures: smoothing and prior information. Journal of Computational and Graphical statistics **11**(3) (2002)
6. Konishi, T., Maruyama, M., Tsubouchi, K., Shimosaka, M.: CityProphet: city-scale irregularity prediction using transit app logs. In: Proc. of UbiComp (2016)
7. MacEachern, S.N., Müller, P.: Estimating mixture of dirichlet process models. Journal of Computational and Graphical Statistics **7**(2) (1998)
8. Nishi, K., Tsubouchi, K., Shimosaka, M.: Extracting land-use patterns using location data from smartphones. In: Proc. of the First International Conference on IoT in Urban Space (2014)
9. Okawa, M., Kim, H., Toda, H.: Online traffic flow prediction using convolved bilinear poisson regression. In: Proc. of MDM (2017)
10. Pan, B., Zheng, Y., Wilkie, D., Shahabi, C.: Crowd sensing of traffic anomalies based on human mobility and social media. In: Proc. of SIGSPATIAL (2013)
11. Shimosaka, M., Hayakawa, Y., Tsubouchi, K.: Spatiality preservable factored Poisson regression for large scale fine grained GPS-based population analysis. In: Proc. of AAAI (2019)
12. Shimosaka, M., Maeda, K., Tsukiji, T., Tsubouchi, K.: Forecasting urban dynamics with mobility logs by bilinear poisson regression. In: Proc. of UbiComp (2015)
13. Shimosaka, M., Tsukiji, T., Tominaga, S., Tsubouchi, K.: Coupled hierarchical Dirichlet process mixtures for simultaneous clustering and topic modeling. In: Proc. of ECML-PKDD (2016)
14. Takeuchi, K., Tomioka, R., Ishiguro, K., Kimura, A., Sawada, H.: Non-negative multiple tensor factorization. In: Proc. of ICDM (2013)
15. Teh, Y.W., Jordan, M.I., Beal, M.J., Blei, D.M.: Hierarchical dirichlet processes. Journal of the American Statistical Association (2006)
16. Wang, C., Paisley, J., Blei, D.: Online variational inference for the hierarchical dirichlet process. In: Proc. of AISTATS (2011)
17. Wang, X., Lindsey, G., Hankey, S., Hoff, K.: Estimating mixed-mode urban trail traffic using negative binomial regression models. Journal of Urban Planning and Development **140**(1) (2013)
18. Yuan, J., Zheng, Y., Xie, X.: Discovering regions of different functions in a city using human mobility and pois. In: Proc. of KDD (2012)
19. Zhang, F., Yuan, N.J., Wilkie, D., Zheng, Y., Xie, X.: Sensing the pulse of urban refueling behavior: A perspective from taxi mobility. In: ACM Trans. on Intelligent Systems and Technology (2015)
20. Zhang, J., Zheng, Y., Qi, D.: Deep spatio-temporal residual networks for citywide crowd flows prediction. In: Proc. of AAAI (2017)
21. Zheng, Y., Liu, T., Wang, Y., Zhu, Y., Liu, Y., Chang, E.: Diagnosing new york city's noises with ubiquitous data. In: Proc. of UbiComp (2014)