

Neural Collective Graphical Models for Estimating Spatio-temporal Population Flow from Aggregated Data

Tomoharu Iwata and Hitoshi Shimizu

NTT Communication Science Laboratories

Kyoto, Japan

Abstract

We propose a probabilistic model for estimating population flow, which is defined as populations of the transition between areas over time, given aggregated spatio-temporal population data. Since there is no information about individual trajectories in the aggregated data, it is not straightforward to estimate population flow. With the proposed method, we utilize a collective graphical model with which we can learn individual transition models from the aggregated data by analytically marginalizing the individual locations. Learning a spatio-temporal collective graphical model only from the aggregated data is an ill-posed problem since the number of parameters to be estimated exceeds the number of observations. The proposed method reduces the effective number of parameters by modeling the transition probabilities with a neural network that takes the locations of the origin and the destination areas and the time of day as inputs. By this modeling, we can automatically learn nonlinear spatio-temporal relationships flexibly among transitions, locations, and times. With four real-world population data sets in Japan and China, we demonstrate that the proposed method can estimate the transition population more accurately than existing methods.

1 Introduction

Analyzing people flow is critical in a wide variety of applications, such as location-based advertisements (Dhar and Varshney 2011), marketing (Kuo, Chi, and Kao 2002), urban development (Ashworth and Voogd 1990), transportation system planning (Koopmans 1949), and evacuation guidance (Yi and Özdamar 2007). Using such sensor devices as GPS, WiFi, Bluetooth, and infrared beacons, we can obtain trajectories for each person. However, trajectory data are often aggregated to protect privacy. For example, mobile spatial statistics (Terada, Nagata, and Kobayashi 2013) are the hourly aggregated spatio-temporal population data of 500-meter grid squares that are generated based on operational data from mobile terminal networks so that individuals cannot be tracked.

In this paper, we propose a probabilistic model for estimating people flow given the aggregated data, which we call the neural collective graphical models. Here the aggregated data are the populations over time for each area, and the people flow is the transition populations between areas. Fig. 1

shows an example of the input aggregated spatio-temporal population data and the output people flow. In our task, the number of parameters to be estimated, which correspond to the people flow over time, is $(T - 1)LM$, where T is the number of time points, L is the number of areas, and M is the number of neighboring areas, where we assume that people move to the neighboring areas. On the other hand, the number of observations, which correspond to the spatio-temporal population data, is TL , which is much smaller than the number of parameters. Learning the population flow only from the aggregated data is an ill-posed problem.

To make the problem tractable, the proposed method assumes that the transition probabilities at different locations and time points are correlated to each other. For example, the transition probabilities at closely located areas or those at the same time of day resemble each other. Since the correlation is unknown, we automatically learn it with neural networks from the given aggregated data. The neural networks enable us to flexibly extract nonlinear spatio-temporal dependence. In the following discussion, we use the locations of the center of the origin and destination areas and the time of day as the input of a neural network. However, it can take other auxiliary information as input, such as the day of the week and the area's weather. We can reduce the effective number of parameters using a neural network whose parameter size is less than $(T - 1)LM$, where the neural network parameter size is constant with respect to the number of areas and time points.

Since no transition population data are given in our task, standard supervised learning methods are inapplicable because they require observations of target variables for training. The proposed model uses the framework of collective graphical models (Sheldon and Dietterich 2011; Sheldon et al. 2013) to estimate the population flow, where the locations of individual persons are analytically marginalized out, and a probabilistic model is investigated about their sufficient statistics. Sufficient statistics, which correspond to the transition populations between areas, are modeled by multinomial distributions, where the transition probability is defined by the neural network described above. By maximizing the likelihood, we simultaneously estimate the transition populations as well as the neural network parameters with gradient-based optimization methods.

The remainder of this paper is organized as follows. Sec-

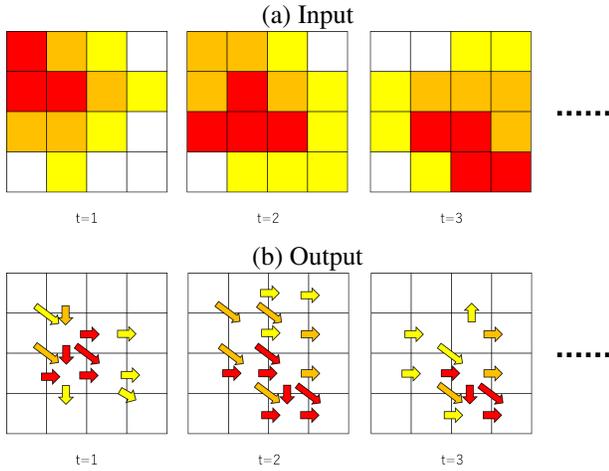


Figure 1: Example of input aggregated spatio-temporal population data and output population flow: (a) Population for each area over time and longitude and latitude of each area are given. (b) Our task estimates transition populations between areas over time.

tion 2 reviews related work. In Section 3, we formulate our task and propose a probabilistic model for estimating the population flow from the aggregated spatio-temporal population data. Section 4 demonstrates the effectiveness of our proposed model with experiments using real-world aggregated population data. Finally, we present concluding remarks and discuss future work in Section 5.

2 Related Work

Collective graphical models have been successfully used in a wide variety of applications, including modeling bird migration (Sheldon, Elmohamed, and Kozen 2007; Nguyen et al. 2016), information diffusion (Kumar, Sheldon, and Srivastava 2013), and contingency tables (Sheldon and Dietterich 2011). To the best of our knowledge, our work is the first attempt to automatically learn dependence among the parameters in collective graphical models with neural networks. Although this paper focuses on population flow, our framework is applicable to other applications of collective graphical models, where auxiliary information is available to effectively estimate parameters.

A number of methods have been proposed for estimating population flow from aggregated population data based on collective graphical models (Kumar, Sheldon, and Srivastava 2013; Du, Kumar, and Varakantham 2014; Sun, Sheldon, and Kumar 2015; Iwata et al. 2017; Tanaka et al. 2018; Akagi et al. 2018). The collective flow diffusion model (Kumar, Sheldon, and Srivastava 2013) assumed static transition probabilities that do not change over time. Iwata et al. (2017) modeled the temporal dependence of transition probabilities by clustering time points using mixture models. Akagi et al. (2018) modeled spatial dependence using the distance between the origins and destinations. However, these methods do not consider spatio-temporal dependence. In addition, they model it in a relatively simple parametric

Table 1: Notation

Symbol	Description
T	number of time points
L	number of areas
$z_{t\ell\ell'}$	population who moves from area ℓ to area ℓ' at time t
$y_{t\ell}$	population in area ℓ at time t
\mathcal{N}_ℓ	set of neighboring areas of area ℓ
\mathbf{x}_ℓ	longitude and latitude of area ℓ
$\theta_{t\ell\ell'}$	transition probability that a person at area ℓ moves to area ℓ' at time t , $\theta_{t\ell\ell'} \geq 0, \sum_{\ell' \in \mathcal{N}_\ell} \theta_{t\ell\ell'} = 1$
$f(\cdot; \phi)$	neural network with parameter ϕ

model. On the other hand, our proposed method can flexibly model nonlinear spatio-temporal dependence using neural networks.

Great interest is being shown in developing methods for predicting spatio-temporal populations (Zhang, Zheng, and Qi 2017; Hoang, Zheng, and Singh 2016; Zhang et al. 2016; 2018; Li et al. 2018; Cheng et al. 2017; Xie et al. 2010; Yu et al. 2016; Yao et al. 2018). For example, Zhang et al. (2018) proposed a deep learning method for forecasting future populations using spatio-temporal dependence as well as external conditions, such as weather and events. However, these methods do not estimate the transition populations between areas, which is our task. Xu et al. (2017) proposed a method to recover individual trajectories given aggregated population data and transition probabilities. This method requires the calculation of the transition probabilities in advance. On the other hand, the proposed method does not require transition probabilities, because both the transition probabilities and the transition populations are simultaneously estimated by taking the aggregated data as inputs.

3 Proposed Method

Task

Suppose that we are given spatio-temporal population data $\mathbf{Y} = \{\{y_{t\ell}\}_{\ell=1}^L\}_{t=1}^T$, where $y_{t\ell}$ is the population in area $\ell \in \{1, 2, \dots, L\}$ at time $t \in \{1, 2, \dots, T\}$. For simplicity, we assume that each area is a grid cell, but the proposed method is applicable to any forms of disjoint areas. Each area is associated with location information $\mathbf{x}_\ell \in \mathbb{R}^2$, e.g., the latitude and longitude of the area's center. Our task is to estimate the population flow between areas for each time point $\mathbf{Z} = \{\{z_{t\ell\ell'}\}_{\ell=1}^L\}_{t=1}^{T-1}$, $\mathbf{z}_{t\ell} = \{z_{t\ell\ell'}\}_{\ell' \in \mathcal{N}_\ell}$, where $z_{t\ell\ell'}$ is the number of people who move from area ℓ to area ℓ' at time t and \mathcal{N}_ℓ is the set of neighbors of area ℓ . Table 1 shows our notation.

Model

First, we consider a probabilistic model for individuals. Each person moves independently based on transition probability $\theta_{t\ell\ell'}$, which is the probability that a person in area ℓ moves to area ℓ' at time t , $\theta_{t\ell\ell'} \geq 0$, and $\sum_{\ell' \in \mathcal{N}_\ell} \theta_{t\ell\ell'} = 1$. Let $s_{tn} \in \{1, \dots, L\}$ be the location of person n at time t . Then

the probabilistic generative process of individual locations $\mathbf{S} = \{\{s_{tn}\}_{n=1}^N\}_{t=1}^T$, where N is the total population, is as follows:

- 1) For each time point $t = 1, \dots, T - 1$
- 2) For each person $n = 1, \dots, N$
- 3) $s_{t+1,n} \sim \text{Categorical}(\boldsymbol{\theta}_{ts_{tn}})$.

Here, $\text{Categorical}(\boldsymbol{\theta})$ represents a categorical distribution with event probability $\boldsymbol{\theta}$, and $\boldsymbol{\theta}_{t\ell} = \{\theta_{t\ell\ell'}\}_{\ell' \in \mathcal{N}_\ell}$. Given individual locations \mathbf{S} , area population \mathbf{Y} and transition population \mathbf{Z} are calculated by $y_{t\ell} = \sum_{n=1}^N I(s_{tn} = \ell)$, and $z_{t\ell\ell'} = \sum_{n=1}^N I(s_{tn} = \ell \wedge s_{t+1,n} = \ell')$, respectively, where $I(\cdot)$ is the indicator function, i.e., $I(A) = 1$ if A is true and $I(A) = 0$ otherwise. Therefore, we can estimate individual locations \mathbf{S} by maximizing the likelihood of observed area population \mathbf{Y} and then estimate transition population \mathbf{Z} using estimated individual locations $\hat{\mathbf{S}}$. However, estimating individual locations \mathbf{S} is too expensive to compute when the number of people is large.

To overcome this problem, we utilized the framework of the collective graphical models, where individual locations \mathbf{S} are analytically marginalized out and a probabilistic model of their sufficient statistics, i.e., transition population \mathbf{Z} , is obtained. Due to this marginalization, we do not need to explicitly estimate individual behaviors. In particular, the following is the generative process of transition population \mathbf{Z} of given area population \mathbf{Y} :

- 1) For each time point $t = 1, \dots, T - 1$
- 2) For each area $\ell = 1, \dots, L$
- 3) $\mathbf{z}_{t\ell} \sim \text{Multinomial}(\boldsymbol{\theta}_{t\ell}, y_{t\ell})$.

Here multinomial($\boldsymbol{\theta}, y$) is the multinomial distribution with event probability $\boldsymbol{\theta}$ and number of trials y . The multinomial distribution of transition population $\mathbf{z}_{t\ell}$ is as follows:

$$p(\mathbf{z}_{t\ell} | \boldsymbol{\theta}_{t\ell}, y_{t\ell}) = \frac{y_{t\ell}!}{\prod_{\ell' \in \mathcal{N}_\ell} z_{t\ell\ell'}!} \prod_{\ell' \in \mathcal{N}_\ell} \theta_{t\ell\ell'}^{z_{t\ell\ell'}}. \quad (1)$$

The following two relationships exist between area population \mathbf{Y} and transition population \mathbf{Z} :

$$y_{t\ell} = \sum_{\ell' \in \mathcal{N}_\ell} z_{t\ell\ell'}, \quad y_{t+1,\ell} = \sum_{\ell' \in \mathcal{N}_\ell} z_{t\ell'\ell}, \quad (2)$$

where the first equation indicates that the sum of the transition populations from an area equals the population in the area, and the second equation indicates that the sum of the transition populations to an area equals the population in the area at the next time point.

The number of parameters in transition probabilities $\boldsymbol{\Theta} = \{\{\boldsymbol{\theta}_{t\ell}\}_{\ell=1}^L\}_{t=1}^T$ is $O(TLM)$, where M is the number of neighbors. On the other hand, since the number of observations \mathbf{Y} is $O(TL)$, the transition probabilities cannot be determined when only these constraints are used. The proposed method imposes more constraints on the transition populations by incorporating spatio-temporal correlation among transition probabilities $\boldsymbol{\Theta}$. We model the transition probabilities with neural network $f(\cdot)$, which takes time t , location

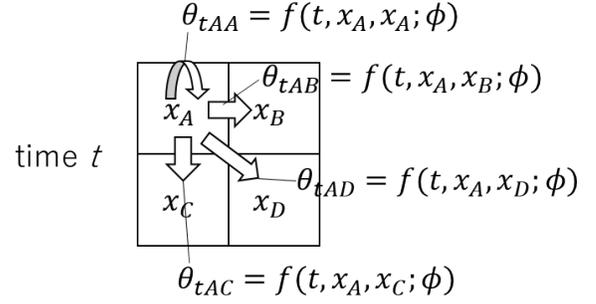


Figure 2: Example of proposed model with its four areas: A, B, C and D, and location information \mathbf{x}_A , i.e., latitude and longitude, is associated to each area. The transition probability between areas θ_{tAB} at time t is calculated by nonlinear function f , which is modeled by a neural network with parameter ϕ , taking time t , origin location \mathbf{x}_A , and destination location \mathbf{x}_B as input.

of origin area \mathbf{x}_ℓ , and location of destination area $\mathbf{x}_{\ell'}$ as the input, as follows:

$$\theta_{t\ell\ell'} = f(t, \mathbf{x}_\ell, \mathbf{x}_{\ell'}; \phi), \quad (3)$$

where ϕ is the parameters of the neural network. By this modeling, related transition probabilities, e.g., those with close time points, near locations and/or similar directions, are constrained to have similar values to each other. The non-linear correlation is automatically extracted from the given data by learning the neural network parameters. Fig. 2 shows an example of the proposed model.

We use the following input vector:

$$\mathbf{u}_{t\ell\ell'} = [\tilde{\tau}(t), \tilde{\mathbf{x}}_\ell, \tilde{\mathbf{x}}_{\ell'} - \tilde{\mathbf{x}}_\ell], \quad (4)$$

where $\tau(t)$ represents the time of day of time t , \tilde{x} indicates the normalized value of x to a range of -0.5 to 0.5 , $[a, b]$ indicates the concatenation of a and b , and $\mathbf{u}_{t\ell\ell'} \in \mathbb{R}^5$. Here the transition probability depends on time of day $\tau(t)$, origin location \mathbf{x}_ℓ , and the direction from the origin to destination $\mathbf{x}_{\ell'} - \mathbf{x}_\ell$. Note that our framework can straightforwardly incorporate other auxiliary information for modeling transition probabilities, such as the day of the week and the area's weather, by concatenating the information to input vector $\mathbf{u}_{t\ell\ell'}$.

We transform the input vector into a transition probability by the following three-layered, feed-forward neural network:

$$\mathbf{h}_{t\ell\ell'} = \tanh(\mathbf{W}_1 \mathbf{u}_{t\ell\ell'} + \mathbf{b}_1), \quad (5)$$

$$\theta_{t\ell\ell'} = \text{softmax}(\mathbf{w}_2 \mathbf{h}_{t\ell\ell'} + b_2), \quad (6)$$

where $\mathbf{h}_{t\ell\ell'} \in \mathbb{R}^H$ is the hidden unit of the neural network, $\mathbf{W}_1 \in \mathbb{R}^{H \times 5}$, $\mathbf{w}_2 \in \mathbb{R}^H$, $\mathbf{b}_1 \in \mathbb{R}^H$, $b_2 \in \mathbb{R}$ are the weights and bias parameters to be estimated, H is the number of hidden units, and $\phi = \{\mathbf{W}_1, \mathbf{w}_2, \mathbf{b}_1, b_2\}$. The proposed model can handle different numbers of neighbors across different locations since the number of neural network parameters does not depend on the number of neighbors. The softmax function in Eq. (6) outputs the transition probability normalized over the neighbors.

Estimation

We estimate transition population \mathbf{Z} and neural network parameters ϕ based on the maximum likelihood with regularizers for the constraints. Using Eqs. (1) and (3), the log likelihood is given by

$$L(\mathbf{Z}, \phi) = \sum_{t=1}^{T-1} \sum_{\ell=1}^L \sum_{\ell' \in \mathcal{N}_\ell} \log p(\mathbf{z}_{t\ell} | \theta_{t\ell}, y_{t\ell}) \quad (7)$$

$$\begin{aligned} &\propto \sum_{t=1}^{T-1} \sum_{\ell=1}^L \sum_{\ell' \in \mathcal{N}_\ell} -\log z_{t\ell\ell'}! + z_{t\ell\ell'} \log f(t, \mathbf{x}_\ell, \mathbf{x}_{\ell'}; \phi) \\ &\approx \sum_{t=1}^{T-1} \sum_{\ell=1}^L \sum_{\ell' \in \mathcal{N}_\ell} z_{t\ell\ell'} \left(1 - \log z_{t\ell\ell'} + \log f(t, \mathbf{x}_\ell, \mathbf{x}_{\ell'}; \phi)\right) \\ &\equiv L'(\mathbf{Z}, \phi), \end{aligned} \quad (8)$$

where Stirling's approximation, $\log z! \approx z \log z - z$, is used in the third line. We incorporate the constraints in Eq. (2) as regularizers and give the objective function to be maximized as follows:

$$\begin{aligned} G(\mathbf{Z}, \phi) &= L'(\mathbf{Z}, \phi) - \lambda \sum_{t=1}^{T-1} \sum_{\ell=1}^L \left\| y_{t\ell} - \sum_{\ell' \in \mathcal{N}_\ell} z_{t\ell\ell'} \right\|^2 \\ &\quad - \lambda \sum_{t=1}^{T-1} \sum_{\ell=1}^L \left\| y_{t+1,\ell} - \sum_{\ell' \in \mathcal{N}_\ell} z_{t\ell\ell'} \right\|^2, \end{aligned} \quad (9)$$

where the second and third terms correspond to the soft constraints in Eq. (2) and $\lambda > 0$ is the hyperparameter. By using the regularizers, we can estimate neural network parameters ϕ as well as transition population \mathbf{Z} with a gradient-based optimization method, since the objective function is differentiable with respect to the parameters. Here we relax transition population $z_{t\ell\ell'}$ to take non-negative real values instead of non-negative integers and parameterize it by $z'_{t\ell\ell'} = \log z_{t\ell\ell'}$ through which the non-negative constraint becomes unnecessary. Algorithm 1 shows the estimation procedure of the proposed model.

Hyperparameter λ is tuned using the following prediction error of the area population at the next time step:

$$E = \sum_{t=2}^T \sum_{\ell=1}^L \left\| \hat{y}_{t\ell} - y_{t\ell} \right\|^2. \quad (10)$$

The predictive value of area population $\hat{y}_{t\ell}$ is calculated using the transition probability and the population at the previous time step as follows:

$$\hat{y}_{t\ell} = \sum_{\ell' \in \mathcal{N}_\ell} \hat{\theta}_{t-1,\ell'\ell} y_{t-1,\ell'}, \quad (11)$$

where $\hat{\theta}_{t\ell\ell'} = f(t, \mathbf{x}_\ell, \mathbf{x}_{\ell'}; \hat{\phi})$ is the transition probability obtained by estimated neural network parameters $\hat{\phi}$.

4 Experiments

Data

We evaluated the proposed method using four real-world population data sets: Tokyo, Osaka, Nagoya, and Beijing (Iwata et al. 2017). Fig. 3 shows examples of them.

Algorithm 1 Estimation procedure for the proposed neural collective graphical model

Require: Spatio-temporal population data \mathbf{Y} , location information $\{\mathbf{x}_\ell\}_{\ell=1}^L$, neighbor information $\{\mathcal{N}_\ell\}_{\ell=1}^L$, hyperparameter λ ;

Ensure: Population flow \mathbf{Z} , estimated neural network parameters $\hat{\phi}$;

Initialize neural network parameters ϕ ;

repeat

Calculate transition probability $\theta_{t\ell}$ by neural network $f(t, \mathbf{x}_\ell, \mathbf{x}_{\ell'}; \phi)$ in (3) for $t = 1$ to $T - 1$, $\ell = 1$ to L ;

Calculate the objective function $G(\mathbf{Z}, \phi)$ in (9) and its gradient with respect to the population flow \mathbf{Z} and neural network parameters ϕ ;

Update the population flow \mathbf{Z} and neural network parameters ϕ using the gradient;

until End condition is satisfied

The Tokyo, Osaka, and Nagoya data are spatio-temporal aggregated population data in the areas of these three cities in Japan. They were generated from individual trajectory data and interpolated from geotagged tweets using railway and road information (Sekimoto et al. 2011)¹. The unit time was 30 minutes, the grid size was 10 km \times 10 km, and the number of grids was 16 \times 14. The Tokyo data contained the population data on July 1 and 7, October 7 and 13, and December 16 and 22 of 2013, where the total population of each day was 6,432, 9,166, 6,822, 10,134, 6,646, 10,338, respectively. The Osaka data contained the population data on August 8 and 11, September 16 and 22, and December 24 and 29 of 2013, where the total population of each day was 2,256, 3,034, 2,999, 3,569, 2,487, 3,480, respectively. The Nagoya data contained the population data on July 22 and 28, September 16 and 22, and December 24 and 29 of 2013, where the total population of each day was 929, 1,332, 1,148, 1,460, 975, 1,570, respectively.

The Beijing data are spatio-temporal aggregated population data in Beijing in China that were generated from T-Drive trajectory data (Yuan et al. 2010; 2011), which contained the trajectories of 10,357 taxis from February 3 to 7, 2008. The unit time was 15 minutes, the grid size was 2 km \times 2 km, and the number of grids was 20 \times 16.

With all of the data sets, the neighbors were the surrounding eight cells with the addition of the cell itself.

Measurement

For the evaluation measurement, we used the following normalized absolute error:

$$\text{NAE} = \frac{\sum_{\ell=1}^L \sum_{t=1}^{T-1} \sum_{\ell' \in \mathcal{N}_\ell} |z_{t\ell\ell'}^* - \hat{z}_{t\ell\ell'}|}{\sum_{\ell=1}^L \sum_{t=1}^{T-1} y_{t\ell}}, \quad (12)$$

¹We used the following sources: SNS-based People Flow Data, Nightley, Inc., Shibasaki & Sekimoto Laboratory, the University of Tokyo, Micro Geo Data Forum, People Flow project, and Center for Spatial Information Science at the University of Tokyo, <http://nightley.jp/archives/1954>

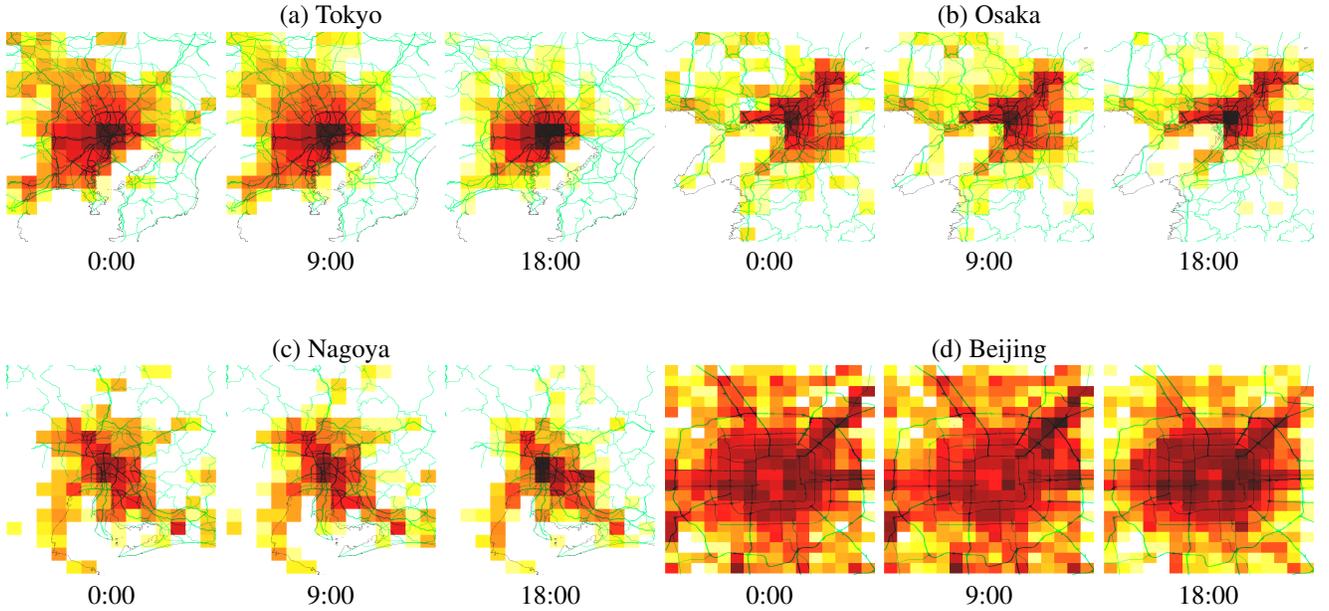


Figure 3: Aggregated spatio-temporal population data: (a) Tokyo, (b) Osaka, (c) Nagoya, and (d) Beijing data sets. Darker red colors represent higher populations in each area. A green line represents a main road.

which is the sum of the absolute errors between true transition population $z_{t\ell\ell'}^*$ and its estimation $\hat{z}_{t\ell\ell'}$ for each time step and for each neighbor area pair divided by the sum of the total number of populations $y_{t\ell}$. Note that we did not use true transition populations $z_{t\ell\ell'}^*$ for training because they are used only for evaluation.

Comparing methods

We compared the proposed method with the following four methods: VCGMM, ICGM, CGM, and STAY. The STAY method, which assumes that all people remain at the current area without moving to other areas, estimates the transition populations by $z_{t\ell\ell'} = y_{t\ell}$ if $\ell = \ell'$, $z_{t\ell\ell'} = 0$ otherwise. The CGM method is the collective graphical model, which assumes that the transition probability does not change over time $\theta_{t\ell\ell'} = \theta_{\ell\ell'}$. The ICGM method is the inhomogeneous transition probability collective graphical model, which assumes that transition probabilities $\theta_{t\ell\ell'}$ are different across different time points. VCGMM is the variational collective graphical mixture models (Iwata et al. 2017), which cluster the time of day using mixture models. With the proposed method, we used ten hidden units and maximized the objective function by ADAM (Kingma and Ba 2014).

The number of parameters to be estimated with the proposed method is $O(H)$, where H is the number of hidden units. Those with VCGMM, ICGM and CGM are $O(KLM)$, $O(TLM)$ and $O(LM)$, respectively, where K is the number of clusters, L is the number of locations, M is the number of neighbors, and T is the number of time points.

Results

Table 2 shows the normalized absolute error averaged over all the time points and all the area pairs. The proposed neural

Table 2: Normalized absolute errors on population flow estimation averaged over all time points

	Proposed	VCGMM	ICGM	CGM	STAY
Tokyo	0.148	0.167	0.176	0.208	0.192
Osaka	0.186	0.250	0.265	0.280	0.272
Nagoya	0.227	0.250	0.281	0.291	0.269
Beijing	0.408	0.470	0.500	0.479	0.532

CGM achieved the lowest error among all of the data sets. Fig. 4 shows the normalized absolute error for each time of day. The error with the proposed method was lower than the other methods, especially in the daytime, when the transition populations are large. This result indicates that the proposed method effectively learns the transition probabilities using spatio-temporal dependencies with neural networks. The error with the STAY method was high in the daytime since it cannot model transitions to different areas. The error with the CGM method was high at night since it assumes that the transition probability in the daytime and at night are the same. VCGMM achieved lower error than ICGM because it reduces the effective number of parameters by clustering the time points. However, since VCGMM cannot utilize location information nor flexibly model the temporal dependence, its performance was worse than the proposed method, which can model the nonlinear spatio-temporal dependence using both location and time information. With the proposed method, the error from 4:00 a.m. to 7:00 a.m. was relatively high, because the transition probability drastically changes in this period, and its switching was not learned properly. Interesting future work will extend our framework to model such switching transition probabilities using mixture of ex-

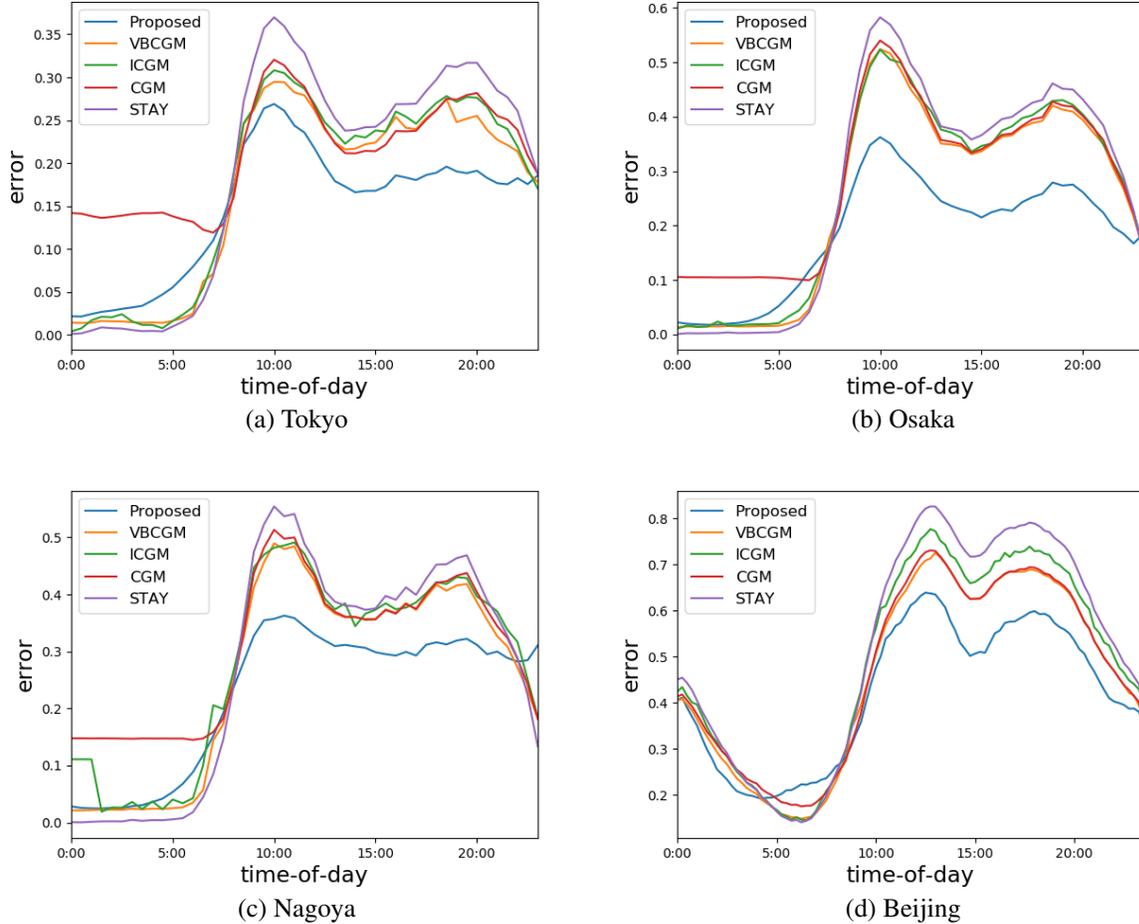


Figure 4: Normalized absolute errors on people flow estimation over time of day.

perts (Jacobs et al. 1991).

Figure 5 shows the transition populations estimated by the proposed method. Reasonable transitions were estimated with all of the data sets. There were few transitions around 2:00 a.m. Around 8:00 a.m., people commuted to the city centers from the suburbs. Around 2:00 p.m., many people moved around, especially in the city centers. Around 8:00 p.m., people left the city centers and returned to their homes in the suburbs.

5 Conclusion

We proposed neural collective graphical models for estimating spatio-temporal population flow given aggregated population data. With our proposed model, nonlinear spatio-temporal dependence on transition probabilities are automatically learned from the given data using neural networks. We experimentally confirmed that the proposed model achieved higher performance on transition population estimation than the existing methods. Although our results are encouraging, we must extend our approach in a number of directions. First, we want to apply our frame-

work to other collective graphical model applications, such as modeling contingency tables and information diffusion. Our framework, which models multinomial distribution parameters with neural networks, is straightforwardly applicable to them. Second, we plan to evaluate our proposed model with higher order Markov transition models that contain more parameters to be estimated. Our model would more effectively reduce the number of parameters with models that have many parameters. Third, we want to investigate using different types of neural networks, such as convolutional neural networks for incorporating images, and recurrent neural networks for forecasting future transitions.

References

- Akagi, Y.; Nishimura, T.; Kurashima, T.; and Toda, H. 2018. A fast and accurate method for estimating people flow from spatiotemporal population data. In *IJCAI*, 3293–3300.
- Ashworth, G. J., and Voogd, H. 1990. *Selling the city: Marketing approaches in public sector urban planning*. Belhaven Press.
- Cheng, X.; Zhang, R.; Zhou, J.; and Xu, W. 2017. Deep-

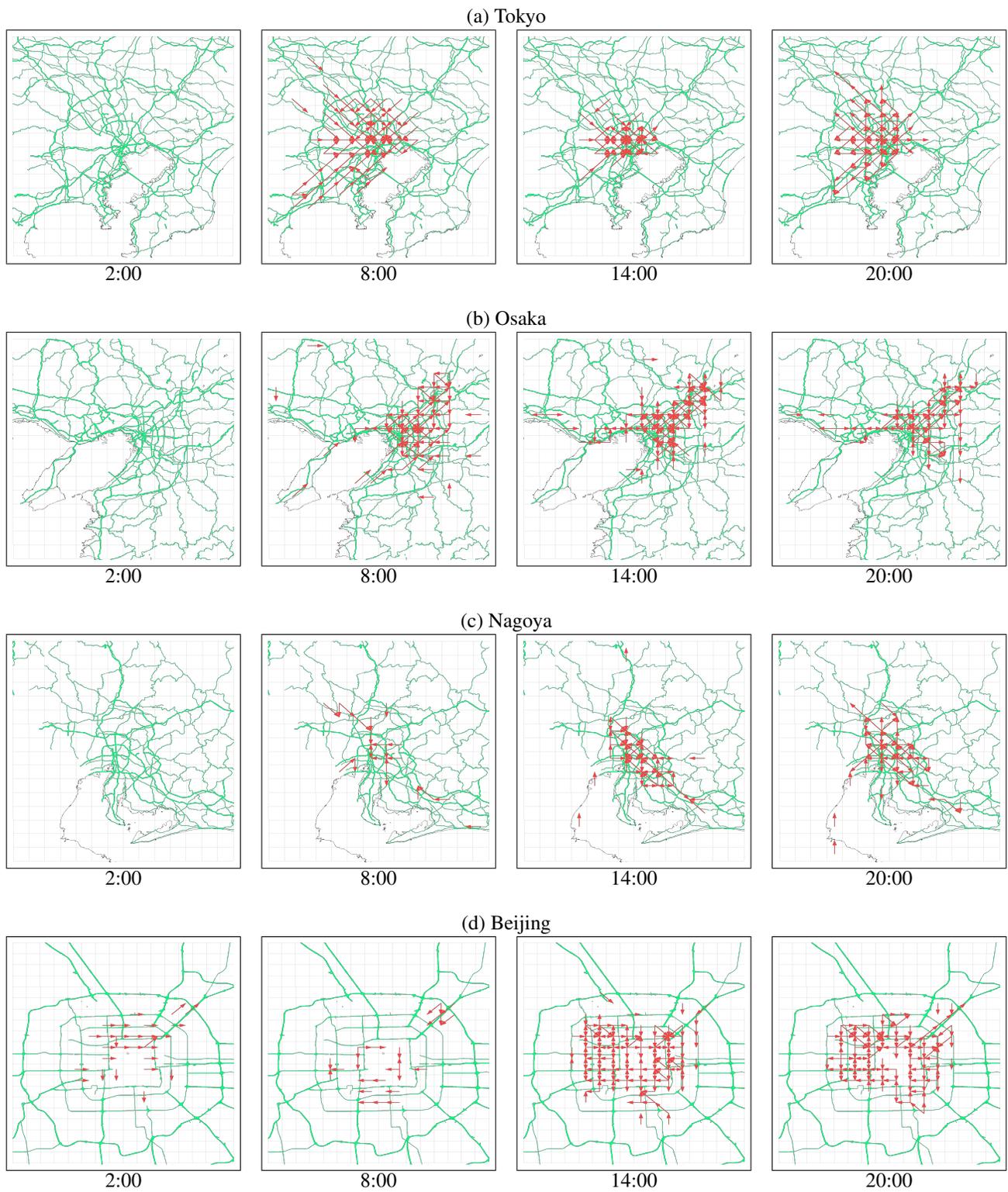


Figure 5: Transition populations estimated by proposed method: (a) Tokyo, (b) Osaka, (c) Nagoya and (d) Beijing data sets. When an estimated transition population exceeds a threshold, an arrow is drawn in that direction. A green line represents a main road.

- transport: Learning spatial-temporal dependency for traffic condition forecasting. *arXiv preprint arXiv:1709.09585*.
- Dhar, S., and Varshney, U. 2011. Challenges and business models for mobile location-based services and advertising. *Communications of the ACM* 54(5):121–128.
- Du, J.; Kumar, A.; and Varakantham, P. 2014. On understanding diffusion dynamics of patrons at a theme park. In *Proceedings of the International Conference on Autonomous Ggents and Multi-Agent Systems*, 1501–1502. International Foundation for Autonomous Agents and Multiagent Systems.
- Hoang, M. X.; Zheng, Y.; and Singh, A. K. 2016. FCCF: forecasting citywide crowd flows based on big data. In *Proceedings of the 24th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, 6. ACM.
- Iwata, T.; Shimizu, H.; Naya, F.; and Ueda, N. 2017. Estimating people flow from spatiotemporal population data via collective graphical mixture models. *ACM Transactions on Spatial Algorithms and Systems (TSAS)* 3(1):2.
- Jacobs, R. A.; Jordan, M. I.; Nowlan, S. J.; and Hinton, G. E. 1991. Adaptive mixtures of local experts. *Neural Computation* 3(1):79–87.
- Kingma, D. P., and Ba, J. 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.
- Koopmans, T. C. 1949. Optimum utilization of the transportation system. *Econometrica: Journal of the Econometric Society* 136–146.
- Kumar, A.; Sheldon, D.; and Srivastava, B. 2013. Collective diffusion over networks: Models and inference. In *Proceedings of International Conference on Uncertainty in Artificial Intelligence*.
- Kuo, R. J.; Chi, S.-C.; and Kao, S.-S. 2002. A decision support system for selecting convenience store location through integration of fuzzy AHP and artificial neural network. *Computers in industry* 47(2):199–214.
- Li, Y.; Yu, R.; Shahabi, C.; and Liu, Y. 2018. Diffusion convolutional recurrent neural network: Data-driven traffic forecasting. In *Proceedings of the 7th International Conference on Learning Representations*.
- Nguyen, T.; Kumar, A.; Lau, H. C.; and Sheldon, D. 2016. Approximate inference using dc programming for collective graphical models. In *Artificial Intelligence and Statistics*, 685–693.
- Sekimoto, Y.; Shibasaki, R.; Kanasugi, H.; Usui, T.; and Shimazaki, Y. 2011. PFlow: Reconstructing people flow recycling large-scale social survey data. *IEEE Pervasive Computing* 10(4):0027–35.
- Sheldon, D. R., and Dietterich, T. G. 2011. Collective graphical models. In *Advances in Neural Information Processing Systems*, 1161–1169.
- Sheldon, D.; Sun, T.; Kumar, A.; and Dietterich, T. G. 2013. Approximate inference in collective graphical models. In *Proceedings of the 30th International Conference on Machine Learning*.
- Sheldon, D.; Elmohamed, M.; and Kozen, D. 2007. Collective inference on Markov models for modeling bird migration. In *Advances in Neural Information Processing Systems*, 1321–1328.
- Sun, T.; Sheldon, D.; and Kumar, A. 2015. Message passing for collective graphical models. In *Proceedings of the 32nd International Conference on Machine Learning*, 853–861.
- Tanaka, Y.; Iwata, T.; Kurashima, T.; Toda, H.; and Ueda, N. 2018. Estimating latent people flow without tracking individuals. In *IJCAI*, 3556–3563.
- Terada, M.; Nagata, T.; and Kobayashi, M. 2013. Population estimation technology for mobile spatial statistics. *NTT DOCOMO Techn. J* 14:10–15.
- Xie, Y.; Zhao, K.; Sun, Y.; and Chen, D. 2010. Gaussian processes for short-term traffic volume forecasting. *Transportation Research Record* 2165(1):69–78.
- Xu, F.; Tu, Z.; Li, Y.; Zhang, P.; Fu, X.; and Jin, D. 2017. Trajectory recovery from ash: User privacy is not preserved in aggregated mobility data. In *Proceedings of the 26th International Conference on World Wide Web*, 1241–1250. International World Wide Web Conferences Steering Committee.
- Yao, H.; Wu, F.; Ke, J.; Tang, X.; Jia, Y.; Lu, S.; Gong, P.; Ye, J.; and Li, Z. 2018. Deep multi-view spatial-temporal network for taxi demand prediction. In *AAAI Conference on Artificial Intelligence*.
- Yi, W., and Özdamar, L. 2007. A dynamic logistics coordination model for evacuation and support in disaster response activities. *European Journal of Operational Research* 179(3):1177–1193.
- Yu, B.; Song, X.; Guan, F.; Yang, Z.; and Yao, B. 2016. k-nearest neighbor model for multiple-time-step prediction of short-term traffic condition. *Journal of Transportation Engineering* 142(6):04016018.
- Yuan, J.; Zheng, Y.; Zhang, C.; Xie, W.; Xie, X.; Sun, G.; and Huang, Y. 2010. T-drive: driving directions based on taxi trajectories. In *Proceedings of the 18th SIGSPATIAL International Conference on Advances in Geographic Information Systems*, 99–108. ACM.
- Yuan, J.; Zheng, Y.; Xie, X.; and Sun, G. 2011. Driving with knowledge from the physical world. In *Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 316–324. ACM.
- Zhang, J.; Zheng, Y.; Qi, D.; Li, R.; and Yi, X. 2016. DNN-based prediction model for spatio-temporal data. In *Proceedings of the 24th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, 92. ACM.
- Zhang, J.; Zheng, Y.; Qi, D.; Li, R.; Yi, X.; and Li, T. 2018. Predicting citywide crowd flows using deep spatio-temporal residual networks. *Artificial Intelligence* 259:147–166.
- Zhang, J.; Zheng, Y.; and Qi, D. 2017. Deep spatio-temporal residual networks for citywide crowd flows prediction. In *AAAI*, 1655–1661.