IoT based Trustworthy Demographic Dynamics Tracking with Advanced Bayesian Learning

Peiran Li, Haoran Zhang, Wenjing Li, Keping Yu, Ali Kashif Bashir, Ahmad Ali AlZubi, Jinyu Chen, Xuan Song, and Ryosuke Shibasaki

***Abstract*—Tracking demographic dynamics for the built environment is important for a smart city. As a kind of ubiquitous Internet of Things (IoT) device, portable devices (e.g., mobile phones) afford a great potential to achieve this goal. Tracking the demographic dynamics illuminates two things: population's mobility (where do people go) and the related demographics (who are they). Many past studies have investigated the tracking of population dynamics; however, few of them tried tracking the demographic dynamics. In this context, our study proposed a ubiquitous IoT based trustworthy approach for built environment demographic dynamics tracking. First, we employed a meta-graph-based data structure to represent users' life patterns and projected them into a low-dimension space as uniform features. Then, based on the life-pattern features, we derived a variation-inference-based advanced Bayesian model to infer the demographics. Finally, taking a region in Tokyo as a case study, we compared our methods with baseline methods (heuristic algorithm, deep learning), and the result proved a superior accuracy (the MAPE improved by 0.07 to 0.28) as well as reliability (0.78 Pearson correlation coefficient with survey data).**

***Index Terms*—IoT, demographics, variation inference, GPS trajectory.**

I. Introduction

A. Background

TRACKING demographic dynamics for the built environment is important in many fields, such as smart building, railway station planning, placement of commercial advertising, emergency management and so on [1, 2]. As a kind of ubiquitous Internet of Things (IoT) [3, 4], portable devices (e.g., mobile phones) afford a great potential to instantly track the built environment demographic dynamics, especially GPS records. To achieve this goal, two factors have to be known: population's mobility (where do people go) and the related demographic information (who are they). Many past studies have investigated the tracking of population dynamics (which only reflects people's mobility) but few of them tried tracking the corresponding demographic dynamics (which also contains people's demographic information). In this context, our study proposed a ubiquitous IoT based trustworthy approach for built environment demographic dynamics tracking. As Figure 1 shows, the underlying problem of this study is labeling the anonymous users with demographic information (age and gender) based on GPS trajectory data and census data so that further tracking of fine-time-interval and variable-range built environment demographic dynamics could be achieved.

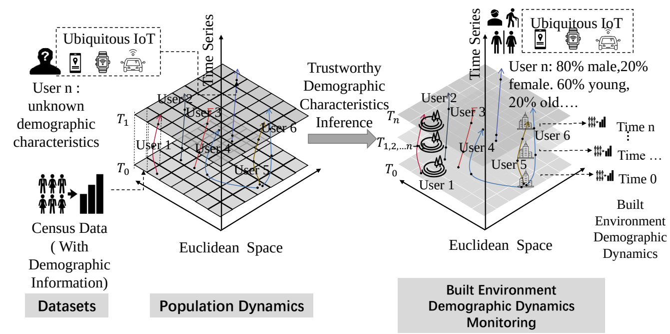


Fig. 1: Ubiquitous IoT based Instant Monitoring for Built Environment Demographic Dynamics.

The GPS sensor on portable IoT devices could conveniently provide the mobility information but barriers were posed on how to infer the demographic information (mainly the age/gender characteristics) based on GPS trajectory data. Although there have been many GPS trajectory data-based studies, mining demographic information from GPS trajectory data is not trivial since a single data source could barely afford enough information [5]. Most past studies chose to combine other types of data, such as mobile usage data[6] (e.g., social media [7]) and individual-level user profile data, to enhance the performance in data inference. However, those additional data, especially some highly sensitive data such as user profiles, is not readily accessible due to privacy risks, etc.

To address the above barriers, we proposed a trustworthy approach based on variation inference theory to perform the demographic inference with only GPS trajectory data and census data.

B. Literature Review

Understanding the mobility patterns of different demographic groups is important; furthermore, it could be more

Acknowledgment: This work was supported by the Researchers Supporting Project (No. RSP-2021/395), King Saud University, Riyadh, Saudi Arabia. The author Peiran Li thanks for the CSC (China Scholarship Council) scholarship.

Peiran Li, Horan Zhang, Wenjing Li, Jinyu Chen, and Ryosuke Shibasaki are with the Center for Spatial Information Science, The University of Tokyo

Xuan Song is currently with Department of Computer Science and Engineering, Southern University of Science and Technology, and CSIS, The University of Tokyo

Keping Yu is with Global Information and Telecommunication Institute, Waseda University, Shinjuku, Tokyo 169-8050, Japan. (e-mail: keping.yu@aoni.waseda.jp)

Ali Kashif Bashir is with the Department of Computing and Mathematics, E-154, John Dolton, Chester Street, M15 6H, Manchester Metropolitan University, Manchester, United Kingdom. (email: dr.alikashif.b@ieee.org)

Ahmad Ali AlZubi is with Computer Science Department, Community College, King Saud University, P.O. Box 28095, Riyadh 11437, Saudi Arabia

Haoran Zhang is the corresponding author. (email:zhang_ronan@csis.u-tokyo.ac.jp)

significant to capture the differences of mobility patterns from different demographic groups. Lu et al. analyzed the correlation between travel behavior and socio-demographics [8], and many studies proved thereafter the high correlation between human mobility behavior and the corresponding demographic information [9, 10]. However, inferring demographic information only through mining GPS trajectory data can be nontrivial; as an alternative, most studies employed multi-source data to infer demographic information for better performance. Wang et al. developed a tensor factorization-based method to infer the demographic information of mobile device users from their AP-trajectories by utilizing prior knowledge such as users' social networks [11]. Wu et al. inferred demographic information from GPS trajectories combined with the geographic context [12]. They extracted spatiotemporal features from GPS trajectory data and obtained the land-use data associated with the GPS trajectories. Then both spatiotemporal and semantic features were integrated as the input to supervised classification models implemented to infer demographic information. Xu et al. proposed a deep learning model named Semantic-enhanced Urban Mobility Embedding (SUME) which learns density vectors for user demographic inference that was achieved through jointly modeling physical mobility patterns (from GPS trajectories data) and semantics of urban mobility (from POI data) [7]. Studies that attempted to infer demographic information using single-source data on individual level also exist though in a limited number. Roy et al. tried to adopt a machine learning approach to predict demographic information [13]; Solomon et al. employed the word2vec approach to construct a GPS-trajectory-based demographic information inference model [5]. Moreover, multi-source data (which is utilized by most studies) is not easy to collect; for some highly sensitive information such as individual-level user profiles, even a single-source dataset is often not accessible. As a consequence, it is imperative that a trustworthy inference method [14, 15] utilizing less sensitive data, such as census data, is developed. This gives rise to the proposal of our approach.

C. Motivation

It has been reported that people's life patterns are related to their demographic information [16]; in the meantime, mining users' life-patterns solely from GPS trajectory data could be feasible [17]. Therefore, a basic idea for our goal - trustworthy demographic dynamics tracking - is to model the relationship between life-pattern features (mined from GPS trajectory data) and demographic information.

There are two key subjects to be addressed: 1) How to mine useful life-pattern features from heterogeneous GPS trajectory data? In this regard, we employed a meta-graph structure to homogenize GPS trajectory data and reduce the high-dimension meta-graph to a low-dimension space so as to form life-pattern features (high-dimension data structure is difficult to deal with). 2) How to model the relationship between life-pattern features and demographic information? For an inference-type problem, statistical maximum likelihood estimation and deep learning (mainly neuron network) are normally used; however, a neuron network (being a high-parameter method) is not suitable due to the limited amount of census data. On the other hand, if low-parameter methods are to be considered, since it is a complex conditional probability distribution problem (see equation (3)), solving the posterior possibility could also be troublesome using maximum likelihood estimation e.g., a heuristic optimization algorithm. Therefore, variation inference is alternatively an appropriate way to deal with the task.

D. Contributions

The goal of this paper is to formulate a trustworthy ubiquitous IoT based instant tracking approach for built environment demographic dynamics. To this end, an effective variation-inference-based model was derived to estimate the demographics solely from GPS trajectory data; then a trustworthy tracking of demographic dynamics was performed (no individual-level information was utilized but precise aggregated information could be obtained in any scale). Taking a region in Tokyo as a case study, we compared the results of our methods with those of baseline methods (heuristic algorithm, deep learning) and proved a superior accuracy (the MAPE improved by 0.07 to 0.28) as well as reliability (0.78 Pearson correlation coefficient with survey data). To summarize, the contributions of this study are:

1) the first trustworthy tracking of built environment demographic dynamics based on large-scale GPS trajectory data,
2) demonstrating the strong correlation between user's life-pattern and demographics, and
3) deriving an effective demographics inference model based on variation inference theory.

E. Organization

The remaining sections of this paper are organized in the following structure: in Section II, a mathematical formulation of our problem was introduced. Section III proposed the methodology framework, variation inference theory, and the model derivation based on the theory. Taking a region in Tokyo as a case study, we conducted and evaluated our method in Section IV. Lastly, we summarized our work in Section V.

II. Preliminary

A. Definition

First, we developed the mathematical formulation of this inference problem. Suppose the number of mobile phone users is n; the number of grids obtained through dividing the target area is a; the number of time intervals obtained through dividing the period under consideration is T; the number of demographic groups is m. Define the number of ground-truth demographic samples (of an individual user) $r = a \times T$.

Definition 2.1 (Mobility Record): A mobility record for a user is a triplet (u, t, l), which denotes that user u visits location l at time t, where l stands for the gridded location (each gird is a 500m*500m square) determined by latitude and longitude.

Definition 2.2 (User Spatio-Temporal Matrix): The mobility sequence for a user is a vector $\{s_0, \ldots, s_r \mid s \in \{0,1\}\}$, which indicates u_i is present or absent at a specific time t and location l. Then, the user spatio-temporal mobility matrix could be defined as $\boldsymbol{S}_{r \cdot n} = \{\overrightarrow{s_0}, \ldots, \overrightarrow{s_n}\}$ which includes all users' spatio-temporal sequence.

Definition 2.3 (User Demographic Characteristics Matrix): The user demographic characteristics could be defined as $\{p_0, \ldots, p_m \mid p \in [0,1]\}$, which indicates the possibility that a user belongs to a specific age/gender group. Then, the user demographic characteristics matrix is a matrix $\boldsymbol{A}_{m \cdot n} = \{\vec{p_0}, \ldots, \vec{p_n}\}$ which includes all users' demographic characteristics.

Definition 2.4 (Built Environment Demographics Matrix): Built environment demographics is a vector $\vec{sa} = \{sa_0, \cdots, sa_m \mid sa \in [0,1]\}$ which indicates the proportion of different demographic groups in a certain area (i.e., the built environment) at certain time. Built environment demographics matrix is a matrix $\boldsymbol{SA}_{r \cdot m} = \{\overrightarrow{sa_0}, \ldots, \overrightarrow{sa_r}\}$ including all areas and all time.

Definition 2.5 (Built Environment Demographics Dynamics): Given a built environment area (a range of location, denoted as la), and the monitoring period T, the built environment demographics dynamics is a matrix $\boldsymbol{D}_{la \cdot T} = \left\{\overrightarrow{SA_{\text{tstart,la}}}, \ldots, \overrightarrow{SA_{\text{tend,la}}}\right\}$ including built environment demographics in la during $t_{start}, \ldots, t_{end}$.

B. Solving Barriers

Thus, $\boldsymbol{SA}_{r \cdot m}$ could be calculated by $\boldsymbol{A}_{m \cdot n}$ and $\boldsymbol{S}_{r \cdot n}$:

$$\boldsymbol{SA}_{r \cdot m} = \boldsymbol{S}_{r \cdot n} \cdot {\boldsymbol{A}_{m \cdot n}}^T \tag{1}$$

The ground-truth demographic data can afford the exact value of $\overline{\boldsymbol{SA}}_{r \cdot m}$. Left multiply the inverse matrix of $\boldsymbol{S}_{r \cdot n}$ simultaneously on both sides of equation (1), we obtain:

$${\boldsymbol{A}_{m \cdot n}}^T = {\boldsymbol{S}_{r \cdot n}}^{-1} \cdot \overline{\boldsymbol{SA}}_{r \cdot m} \tag{2}$$

At first glance, $\boldsymbol{A}_{m \cdot n}$ could be solved by simply utilizing equation (2);however, daunting barriers stand in the way of computing ${\boldsymbol{S}_{r \cdot n}}^{-1}$. Specifically, the barriers embody two subproblems:

Problem 2.1: In most cases, $\boldsymbol{S}_{r \cdot n}$ should be an ill-conditioned matrix that has a large condition number – most individuals' GPS information is sparsely distributed in a large spatio-temporal space. As a result, the solution could be highly unrobust due to the sensitive inverse operation of the matrix $\boldsymbol{S}_{r \cdot n}$.

Problem 2.2: A purely mathematical approach limits the generality – it is prone to yield an over-fitted result: solving a specific formula can well afford the demographic information of specific people within the current GPS trajectory dataset, but it may degenerate when applied with another dataset that is unseen.

Based on the above considerations, solving $\boldsymbol{A}_{m \cdot n}$ through brute force should not be a good choice. Hence, our goal is redirected to find:

$$\begin{aligned} \boldsymbol{A}_{opt} \mid MAE\left(\overline{\boldsymbol{SA}}_{r \cdot m}, \boldsymbol{SA}_{r \cdot m} \mid \boldsymbol{A}_{opt}\right) \leq \\ MAE\left(\overline{\boldsymbol{SA}}_{r \cdot m}, \boldsymbol{SA}_{r \cdot m} \mid \boldsymbol{A}\right) \end{aligned} \tag{3}$$

Then, the built environment demographic dynamics is:

$$\boldsymbol{D}_{l \cdot T} = \left\{\boldsymbol{S}_{r \cdot n} \cdot {\boldsymbol{A}_{opt}}^T \mid \text{ where } S \text{ in } T \text{ and } l\right\} \tag{4}$$

III. Methodology

A. Framework

First, users' life-pattern features were mined from heterogeneous GPS trajectory data. We identified significant places (e.g., home, workplaces and others.) by clustering from trajectories for each user and generated an individual graph that reflected his/her location by each hour within one day. Then we constructed support trees with a uniform structure for all users - in the support graph, each edge is assigned with a unique index in an ascending manner from top to bottom and from left to right. Finally we generated a topology-attribute matrix (T-A matrix) to incorporate the user's life-pattern feature (Figure 2).

Second, based on the extracted life-pattern features, a variation-inference-based demographics inference method was derived, which only requires the GPS trajectory data and census data. For each user, we reduce the dimension of the T-A matrix by NFM (Non-negative matrix factorization) to project it into a 3-dimension space (a meta-graph space), so that the spatial locations in the meta-graph space could represent the user's life-patterns. We could assume that the possibility distribution of a certain demographic group in the meta-graph space is a Gaussian distribution (life-pattern is related to one's demographic information). Then, we employed variation inference theory to infer the optimal parameters of the Gaussian distribution of all demographic groups. Once the parameters have been determined, we could infer each user's demographic characteristics by calculating the joint probability of all demographic Gaussian distributions.

Lastly, taking a region in Tokyo as a case study, we compared the accuracy of variation inference with those achieved by other baseline methods (heuristic algorithm, deep learning). Also, by comparing the daily time-use estimation from our results with survey data, and by comparing the regional demographic dynamics from our results with the ground-truth data, we evaluated the effectiveness of the variation inference method against other baseline methods.

B. Variation Inference Theory

In Bayesian statistics, unknown quantities inference could be considered as the calculation of posterior probabilities, which are often difficult to conduct. To tackle this problem, one of the solutions is MCMC (Markov Chain Monte Carlo), but it works slowly when confronted with a large size of data (which is expected in our task). Alternatively, Variation Inference (VI) method could serve as a powerful tool for achieving approximate possibility inference from a large size of big data [18]. The implementation of the VI method is discussed as follows.

Suppose the input observation variables are $\vec{x} = x_1, x_2, \ldots, x_n$, and the latent variables inside the model are $\vec{z} = z_1, z_2, \ldots, z_n$. Approximating a conditional density of latent variables $\vec{z}$ is the aim of VI. The basic idea of VI is to

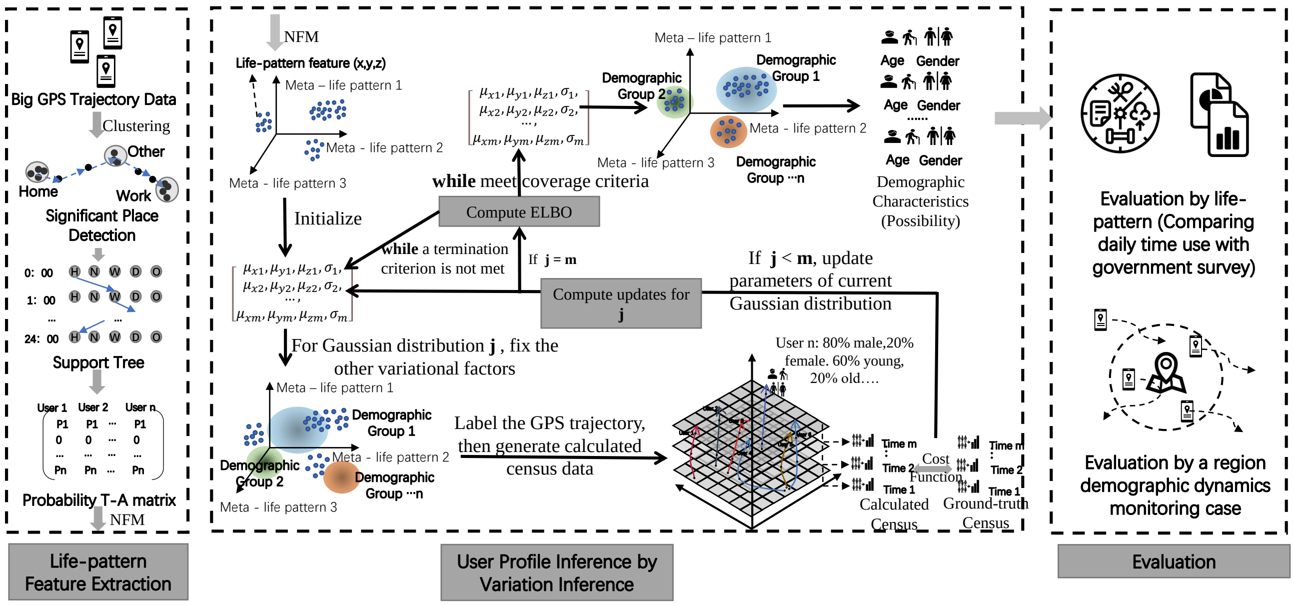


Fig. 2: The framework of methodology.

take it as an optimization problem: propose at first a family of approximate probability distribution Q which is related to latent variables $\vec{z}$, and the goal is to find a distribution among Q, which has a minimal KL-Divergence (Kullback-Leibler Divergence) from the true posterior possibility distribution [18], that is:

$$q^*(\vec{z}) = \operatorname{argmin} \cdot KL(q(\vec{z})||p(\vec{z} \mid \vec{x})) \qquad q(\vec{z}) \in Q \tag{5}$$

where the optimized distribution $q^*(\vec{z})$ could be regarded as the approximate posterior possibility distribution $p(\vec{z}|\vec{x})$.

Since $KL(q(\vec{z})||p(\vec{z}|\vec{x}))$ is difficult to compute directly, it is usually replaced by a constructed item named Evidence Lower Bound (ELBO) within the VI method:

$$ELBO = E[\log p(\vec{z}, \vec{x})] - E[\log q(\vec{z})] \tag{6}$$

It could be demonstrated that minimizing the KL-Divergence $KL(q(\vec{z})||p(\vec{z}|\vec{x}))$ is equivalent to maximizing the ELBO. Therefore, equation (5) could be transformed to be (7):

$$q^*(\vec{z}) = \operatorname{argmax} ELBO(qz, pz, x) q(z) \in Q \tag{7}$$

Choosing an appropriate form of approximate probability distribution Q could facilitate the optimization. A common, simple and effective variational family is the *mean-field variational family* which assumes that the latent variables are independent of each other:

$$q(\vec{z}) = \prod_{j=1}^{m} q_j(z_j) \tag{8}$$

Based on equation (6) and (8), and by combining the CAVI (Coordinate Ascent Variational Inference) method [18, 19], the rule of coordinate ascent could be derived to be (the derivation is shown in Appendix):

$$q^*(\vec{z_k}) \propto e^{E_{-k}\left(\log^p\left(\vec{z_k}|Z_{-k},\vec{x}\right)\right)} \tag{9}$$

According to expression (9), fixing other coordinates of $\vec{z}$ allows computing the updates of current parameters, as is shown in the following procedure:

Input: A model $p(\vec{z}|\vec{x})$, observations x
Output: A variational density $q(\vec{z}) = \prod_{j=1}^{m} q_j(z_j)$
Initialize: Variational factors $q_j(z_j)$
While the ELBO do not meet the termination criterion do
  for $j \in \{1, \ldots, m\}$ **do**
    set $q_j(z_j) \propto e^{E_{-j}\left(\log^p\left(\vec{z_j}|Z_{-j},\vec{x}\right)\right)}$
  end
  Compute $ELBO(q) = E(\log p(z, x)) - E(\log q(z))$
end
return $q(z)$

C. Variation Inference Model Construction

Parameters to be optimized. There are so numerous users that taking all users' demographic characteristics as the input parameters could be impractical. To construct an input parameter form that could be solvable, we assume that each demographic group is Gaussian distributed in the life-pattern space (as is shown in Figure 2) so that only four parameters will be needed to describe each demographic group (totally $4 \times m$ parameters for m demographic groups). Once one set of Gaussian distribution parameters is determined, all users' demographic characteristics could be calculated and could then be used to iterate based on the census data to obtain a new set of Gaussian distribution parameters until the convergence of demographic characteristics.

Suppose there are m demographic groups, the parameters to be optimized could be defined as:

$$\boldsymbol{x}_i = \begin{bmatrix} \mu_{x1}, \mu_{y1}, \mu_{z1}, \sigma_1, \mu_{x2}, \mu_{y2}, \mu_{z2}, \sigma_2, \\ \ldots, \mu_{xm}, \mu_{ym}, \mu_{zm}, \sigma_m \end{bmatrix} \tag{10}$$

where μ_x, μ_y, μ_z stands for the coordinate of the center of a demographic group's Gaussian distribution in the life-pattern space; σ denotes the standard error of this Gaussian distribution. Then the possibility that a specific user belongs to each demographic group could be calculated as:

$$\vec{p_i} = \begin{bmatrix} \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-\left((x-\mu_{x1})^2+(y-\mu_{y1})^2+(z-\mu_{z1})^2\right)}{2\delta^2}}, \cdots, \\ \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-\left((x-\mu_{xm})^2+(y-\mu_{ym})^2+(z-\mu_{zm})^2\right)}{2\delta^2}} \end{bmatrix} \tag{11}$$

and the demographic characteristics matrix of all users $\mathbb{U}$ should be:

$$A = \{\overrightarrow{p_0}, \ldots, \overrightarrow{p_n}\} \tag{12}$$

Considering equation (3), we define the cost function f as:

$$f(\boldsymbol{x}_i) = \mathrm{MAE}\left(\overline{\boldsymbol{SA}}, \boldsymbol{S} \cdot \boldsymbol{A}^T \mid \boldsymbol{A}\right) \tag{13}$$

Mathematical Derivation. Based on the VI theory and CAVI algorithm shown above, we derived the algorithm for our task as follows:

Let

$$p(x, z) = e^{f(x,z)} \tag{14}$$

where:

- f is the same as the cost function f in equation (13)
- $\boldsymbol{x}$ stands for the observations, i.e., the life-pattern coordinate (x, y, z) of each user in the life-pattern space.
- $\boldsymbol{z}$ stands for the parameters for mixed Gaussian distribution, which are the same as the parameters used in equation (5)

where:

- $\boldsymbol{q_j}$ stands for a Gaussian possibility density function, and m stands for the number of demographic groups.

For equation (11), let q_j stands for a Gaussian possibility density function (similar as equation (7)), and m stands for the number of demographic groups. Hence, $q_j(z_j)$ should be:

$$\begin{aligned} & q_j(z_j) \propto e^{E_{-j}(\log^p(Z_j|Z_{-j},x))} \\ =& e^{\left(E_{-j}\left[\log e^{f(x,z_j)|z_{-j}}\right]\right)} \\ =& e^{(E_{-j}[f(x,z_j)|z_{-j}])} \propto E_{-j}\left[f(x, z_j) \mid z_{-j}\right] \end{aligned} \tag{15}$$

And ELBO should be:

$$ELBO(q) = f(z, x) \tag{16}$$

The overall flow is shown as below:

Input: A model $p(\vec{z}|\vec{x})$, observations x
Output: A variational density $q(\vec{z}) = \prod_{j=1}^{m} q_j(z_j)$
Initialize: Variational factors

$$q_j(z_j) = \frac{1}{\sigma j\sqrt{2\pi}} e^{\frac{-\left((x-\mu_{xj})^2+(y-\mu_{yj})^2+(z-\mu_{zj})^2\right)}{2\delta j^2}}$$

where x, y, z belongs to $\vec{x}$

While the ELBO do not meet the termination criterion do
  for $j \in \{1, \ldots, m\}$ **do**
    Set $q_j(z_j) \propto E_{-j}\left[f(x, z_j) \mid z_{-j}\right]$
  end
  Compute $ELBO(q) = f(z, x)$
end
return $q(z)$

With the optimized $q(z)$, we can calculate $\boldsymbol{A}_{opt}$ and formulate the built environment demographic dynamics as:

$$\boldsymbol{D}_{la \cdot T} = \left\{\boldsymbol{S}_{r \cdot n} \cdot \boldsymbol{A}_{opt}^T \mid \text{ where } S \text{ in } T \text{ and } la\right\} \tag{17}$$

IV. Case Study: Experiment in Tokyo, Japan

A. Data Description

This study employed only 2 datasets for demographic (age/gender) inference: a large-scale GPS trajectory dataset and a census population data. In addition, a life-pattern statistical dataset is used for evaluation:

Human Mobility Data.

Human Mobility Data. This study employed a human mobility dataset named "Konzatsu-Tokei (R)" Data. "Konzatsu-Tokei (R)" Data refers to people flows data collected by individual location data sent from mobile phone under users' consent, through Applications provided by NTT DOCOMO, INC. Those data is processed collectively and statistically in order to conceal the private information. Original location data is GPS data (latitude, longitude) sent in about every a minimum period of 5 minutes and does not include the information to specify individual. ※Some applications such as "docomo map navi" service (map navi · local guide). In this study, we selected users who passed Tokyo in the corresponding period with demographic dataset. In this study, we selected users who passed Tokyo (23 wards) in the corresponding period with demographic dataset.

Time-series Demographic Dataset. A demographic data named 'Mobaku data' was taken as the ground-truth statistical data. It is generated by the DoCoMo (i.e., NTT DoCoMo, Inc.) cell phone network - the number of cell phones can be counted and the population can be estimated considering the penetration rate of DoCoMo within each area. Since its users covered 80 million among the total of 126 million population in Japan, the estimation could achieve a significantly high statistical precision. The demographic data was selected to be consistent with the human mobility data in terms of the period as well as areas of record. As shown in figure 1, every grid (500m*500m) contains the population of different demographic groups: male/female with age falling into 8 groups: 0∼15, 15∼20, 20∼30, 30∼40, 40∼50, 50∼60, 60∼70, 70∼80.

Dataset of Time Use and Leisure Activities. To further evaluate the result of our inference, the '*Survey on Time Use and Leisure Activities*' (conducted by the Statistics Bureau of Japan) statistical data were utilized. This survey is conducted once every five years to observe the daily time use of different activities for Japanese people [20]. It contains the average time use in a single day on different types of activities (e.g., working, studying, sleeping and etc.) for populations with different ages (including four age groups: $<$ 35, 35∼44, 45∼64 and $>$ 65, male and female).

B. Baseline Settings

For an inference problem, statistical maximum likelihood estimation (usually a low-parameter method) and deep learning (mainly neuron network, a high-parameter method) are usually used. Among low-parameter methods, we also employed an optimization method – PSO (Particle Swarm Optimization) as a baseline in addition to the proposed VI-based method in Section 3.4. Among high-parameter methods, we used a deep learning approach to fit the life-pattern feature with demographic characteristics i.e., to construct a point-to-point relationship between one's life-pattern feature and one's demographic characteristics. Specifically, we implemented two models: one was Fully Connected Neuron Network (FCN) which directly yielded demographic characteristics for each user; the other one was Multi-Task Fully Connected Neuron Network (Multi-Task FCN) – branches of different demographic groups were generated, with each branch representing the possibility for the user of belonging to that demographic group.

Base Baseline: Persistence Algorithm To further confirm the effectiveness of all inference models, we introduced a common baseline - persistence algorithm (the "naive" forecast) where regional demographics of the previous day are used for the estimation of the value of the current day.

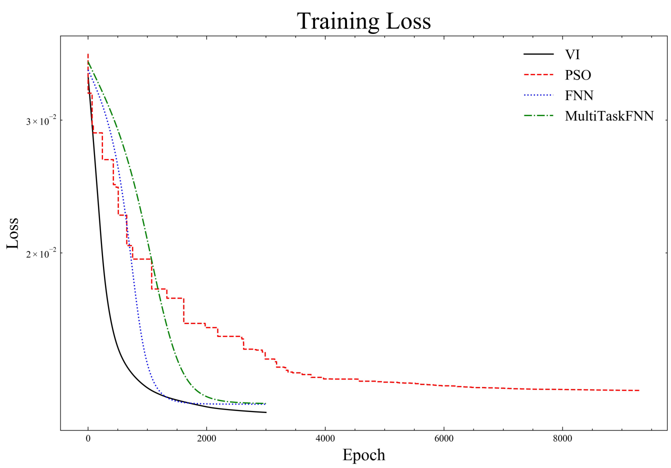


Fig. 3: Loss-epoch of model training.

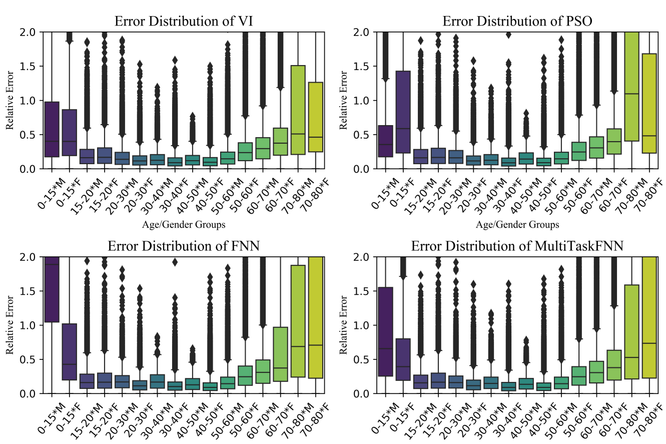


Fig. 4: Relative error distribution comparison of different methods.

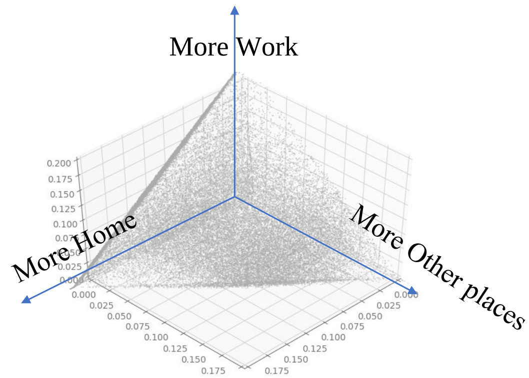


Fig. 5: Illustration of life-pattern space.

C. Evaluation Metrics

Evaluation by Daily Time Use. We estimated daily time use from the inferred demographic characteristics combined with human mobility data and compared the result with the statistical data '*Survey on Time Use and Leisure Activities*' (conducted by the Statistics Bureau of Japan).

First, for each demographic group (e.g., male aged 30 40), we performed min-max normalization for each demographic group aggregating all the users; then we selected users whose normalized possibility was above 0.8 as the representatives of this demographic group. Second, we exploited the average 'daily home time', 'daily work time', 'daily other time' for every demographic group from GPS trajectory data (we have identified users' homes, workplaces, and other significant places by clustering GPS trajectory data). Third, we counted the occurrence of the same values in the dataset '*Survey on Time Use and Leisure Activities*'. Then, we resampled the results into the same age interval. Lastly, we made a regression between survey's daily time use result and our GPS-derived daily time use result. The Pearson correlation coefficient of regression was taken as the metric to evaluate the model performance.

$$r = \frac{\sum (x - m_x)(y - m_y)}{\sqrt{\sum (x - m_x)^2 \sum (y - m_y)^2}} \tag{18}$$

Evaluation by Built Environment Demographics. Taking a built environment as a case study, we evaluated our results from another view. As the Figure 8 shows, we choose an exhibition hall area with about 700m length and 300m width from Tokyo, which has 3-day ground-truth built environment demographic dynamics. The ground-truth data were utilized to evaluate the estimation of built environment demographic dynamics through VI and other 4 baselines. First, we detected stay points from GPS trajectory by clustering; then, we labeled the GPS trajectory with the inferred demographic characteristics for each user to obtain the built environment demographic dynamics of the period consistent with the ground-truth data. Lastly, we calculated the MAPE (Mean Absolute Percentage Error) to evaluate the performance of different methods.

$$\mathrm{MAPE} = \frac{100}{n} \sum_{t=1}^{n} \left| \frac{A_t - F_t}{A_i} \right| \tag{19}$$

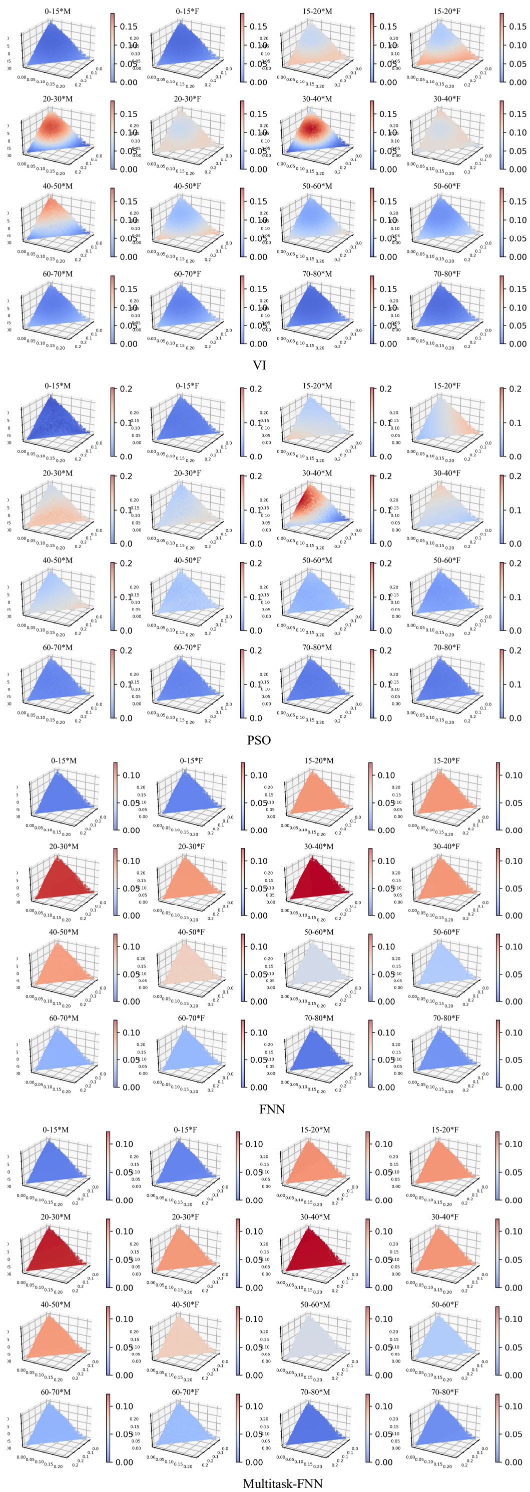


Fig. 6: Illustration of possibility distribution of different demographic groups in life-pattern space.

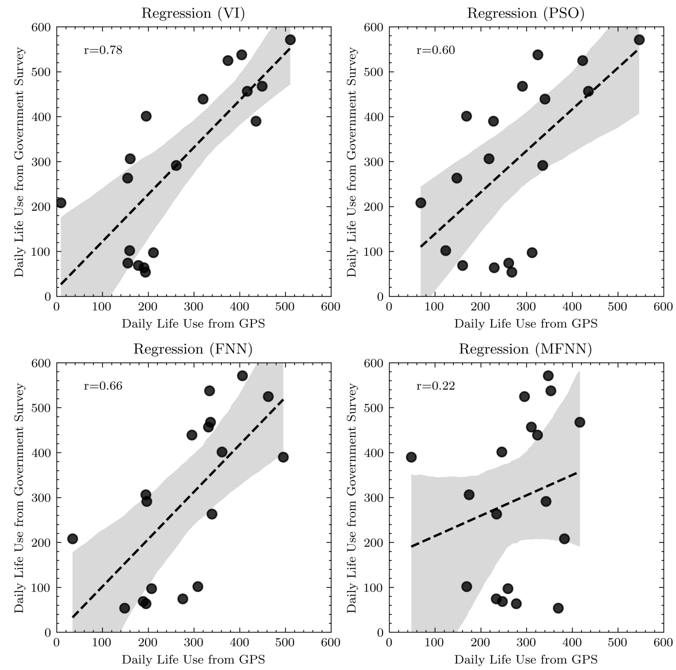


Fig. 7: Daily time use regression between government survey and GPS trajectory data

D. Overall Results

VI achieved the highest accuracy with comparable time cost against other baselines. As Figure 3 shows, VI achieved an MAE of 0.0123, while that of PSO, FCN, and MultiTask-FCN are 0.013, 0.0126, and 0.0126, respectively. On the other hand, VI converged within 3000 epochs, which was close to the epoch of the deep learning approach (FCN and MultiTask-FCN), and was much faster than the heuristic algorithm (PSO) which took more than 8000 epochs.

Further, we compared the relative error distribution of different demographic groups. From this aspect, VI also achieved better results compared with other baselines. Figure 4 shows that all of the methods performed relatively better when inferring targets aged from 15$\sim$70, but relatively worse when the age was under 15 (the younger group) and above 70 (the elder group). The discrepancy resulted from the lower penetration of mobile phones among these two groups; as a result, their GPS trajectory data covered only a small portion of our dataset, leading to the difficulty in performing a fine fitting. Despite this obstacle, we still found the superiority of VI compared with heuristic algorithm and deep learning model – VI inferred the younger and the elder groups far better than the other two methods even though their relative errors were similar when estimating demographic groups between 15 $\sim$ 70-year-old.

E. Evaluation by Time Use Survey Data

Although results of training loss and relative errors indicated the superior performance of VI, it is yet to be demonstrated whether our method evidently reveals the underlying life-pattern of different demographic groups. This section is dedicated to investigating and evaluating the life-pattern features extracted from the inference of demographic groups.

The life-pattern tendencies of different demographic groups. As is mentioned above, we employed a metagraph-based data structure to illustrate users' life-pattern, with the high-dimensional data being reduced to a 3-dimensional space, i.e., the life-pattern space. Additionally, our previous

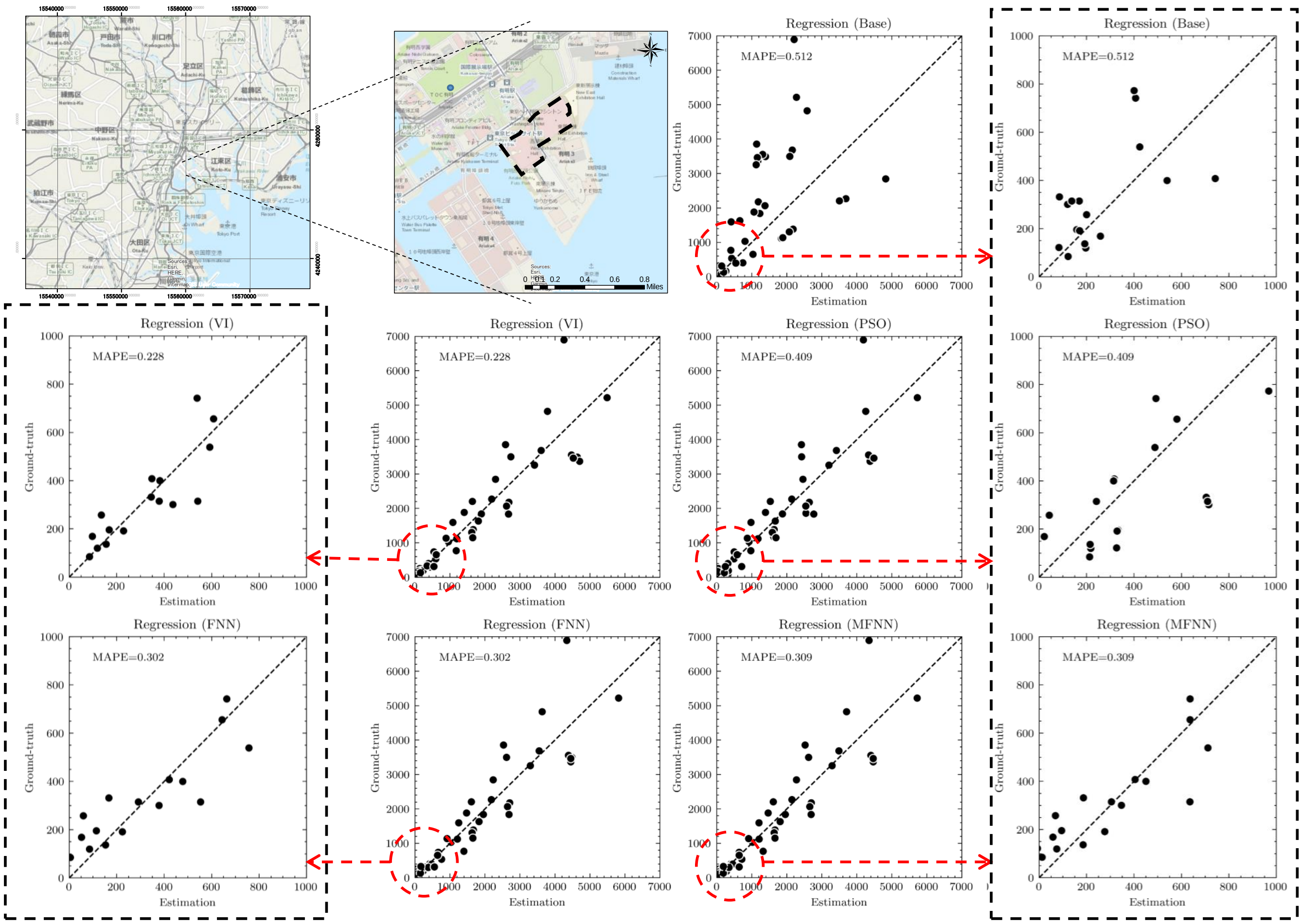


Fig. 8: Scatter graph between estimated demographics and ground-truth demographics.

work has informed that the three axes of the life-pattern space represented three different life-pattern tendencies: work-preferred, home-preferred, and other-preferred [17], as shown in Figure 5). To reveal the life-pattern tendencies of different demographic groups, we mapped the normalized possibility distribution of different demographic groups aggregating all users into the life-pattern space (Figure 6).

In general, VI and PSO methods successfully distinguished different life-pattern tendencies of different demographic groups which could be reflected by the probability densities in the life-pattern space. On the other hand, FCN and MultiTask-FCN method failed to discern the tendencies as the probability density were almost uniform for all demographic groups.

If we also focus on the major demographic groups (i.e., males and females of 15∼50-year-old), better rationality could be reached from the VI method than from the PSO method. According to the result obtained by the VI method, a 15∼20-year-old person (regardless of the gender) tends to stay at home or other places as opposed to the workplace; a 20∼50-year-old male prefers spending more time at his workplace, while a 20∼50-year-old female is more likely to stay at home or other places. By contrast, results obtained by the PSO method indicate that a 20∼30-year-old male tends to stay at home, while a 30∼40-year-old female would prefer the workplace. Nevertheless, neither of these two methods can reveal the tendencies for the younger and the elder group (users whose ages are under 15 or above 70-year-old).

Quantitative validation of extracted life-pattern features. To quantitatively evaluate whether the estimations are consistent with the practical data or not, we compared the daily time use data derived from the inferred users' demographic characteristics with the statistical data '*Survey on Time Use and Leisure Activities*' (conducted by Statistics Bureau of Japan).

As Figure 7 shows, the results of VI matched the government survey data better than other baselines, reaching a 0.78 r-square while Multi-task FCN shows the worst matching with a 0.22 r-square. The results show that the inference result from VI well mined the life-pattern difference between different demographic groups.

F. *Evaluation by Built Environment Demographics*

In terms of the estimation results of built environment demographics, VI also shows superior performance than baselines. The MAPE of different methods are 0.23 for VI, 0.51 for Base, 0.40 for PSO, 0.30 for FCN, and 0.31 for multitask FCN, meaning the MAPE is improved by 0.07 ∼ 0.28. If we investigate the scatter graph between estimation and ground-truth, VI better performance could be observed especially when estimating the minor demographic groups, as Figure 8 shows.

V. Conclusion

This paper proposed a variation-inference-theory-based approach to perform the demographics inference sorely using GPS trajectory data and census data, which could serve as an effective way for trustworthy demographic dynamics tracking.

We demonstrated the feasibility to infer demographics sorely by using GPS trajectory data and census data. The overall MAE between the inferred demographics and census demographics reached 0.0123. The Pearson correlation coefficient between the estimated daily time use and the government survey data reached 0.78. Also, taking a region in Tokyo as a case study, we evaluated the estimation of built environment demographics: the MAPE obtained based on ground-truth data reached 0.23.

Further, we concluded the efficiency of our VI-based method in terms of the convergence rate - the VI method consumed less than half of the time to converge compared to the PSO method. Also, through different evaluation metrics, the VI method demonstrated a superior accuracy against other baselines - MAPE of built environment demographics estimation was improved by $0.07 \sim 0.28$.

Despite the performance of our proposed VI method, several uncertainties that could further complicate the inference problem remain untreated in our experiment. This study only considered the proportion of different demographic groups instead of using a set of scaling factors to calculate the absolute population of different demographic groups. In addition, the generality of our method may be limited as we only selected a region of 23 wards in Tokyo to conduct the case study.

References

[1] M. N. Alverti, K. Themistocleous, P. C. Kyriakidis, and D. G. Hadjimitsis, "A Study of the Interaction of Human Smart Characteristics with Demographic Dynamics and Built Environment: The Case of Limassol, Cyprus," *Smart Cities*, vol. 3, no. 1, pp. 48–73, feb 2020. [Online]. Available: https://www.mdpi.com/2624-6511/3/1/4/htmhttps://www.mdpi.com/2624-6511/3/1/4

[2] Q. Zhang, J. Wu, M. Zanella, W. Yang, A. K. Bashir, and W. Fornaciari, "Sema-iiovt: Emergent semantic-based trustworthy information-centric fog system and testbed for intelligent internet of vehicles," *IEEE Consumer Electronics Magazine*, 2021.

[3] W. Z. Khan, Y. Xiang, M. Y. Aalsalem, and Q. Arshad, "Mobile phone sensing systems: A survey," *IEEE Communications Surveys Tutorials*, vol. 15, no. 1, pp. 402–427, 2013.

[4] M. Thejaswini, P. Rajalakshmi, and U. B. Desai, "Novel sampling algorithm for human mobility-based mobile phone sensing," in *IEEE Internet of Things Journal*, vol. 2, no. 3. Institute of Electrical and Electronics Engineers Inc., jun 2015, pp. 210–220.

[5] A. Solomon, A. Bar, C. Yanai, B. Shapira, and L. Rokach, "Predict demographic information using Word2vec on spatial trajectories," in *UMAP 2018 - Proceedings of the 26th Conference on User Modeling, Adaptation and Personalization*, vol. 18. Association for Computing Machinery, Inc, jul 2018, pp. 331–339. [Online]. Available: https://doi.org/10.1145/3209219.3209249

[6] Z. Yu, E. Xu, H. Du, B. Guo, and L. Yao, "Inferring user profile attributes from multidimensional mobile phone sensory data," *IEEE Internet of Things Journal*, vol. 6, no. 3, pp. 5152–5162, jun 2019.

[7] F. Xu, Z. Lin, T. Xia, D. Guo, and Y. Li, "SUME: Semantic-enhanced Urban Mobility Network Embedding for User Demographic Inference," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 4, no. 3, sep 2020. [Online]. Available: https://doi.org/10.1145/3411807

[8] X. Lu and E. I. Pas, "Socio-demographics, activity participation and travel behavior," *Transportation Research Part A: Policy and Practice*, vol. 33A, no. 1, pp. 1–18, jan 1999.

[9] A. Almaatouq, F. Prieto-Castrillo, and A. Pentland, "Mobile communication signatures of unemployment," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 10046 LNCS. Springer Verlag, nov 2016, pp. 407–418. [Online]. Available: https://link.springer.com/chapter/10.1007/978-3-319-47880-7_25

[10] F. Luo, G. Cao, K. Mulligan, and X. Li, "Explore spatiotemporal and demographic characteristics of human mobility via Twitter: A case study of Chicago," *Applied Geography*, vol. 70, pp. 11–25, may 2016.

[11] P. Wang, F. Sun, D. Wang, J. Tao, X. Guan, and A. Bifet, "Inferring demographics and social networks of mobile device users on campus from ap-trajectories," in *26th International World Wide Web Conference 2017, WWW 2017 Companion*, 2017, pp. 139–147. [Online]. Available: http://dx.doi.org/10.1145/3041021.3054140

[12] L. Wu, L. Yang, Z. Huang, Y. Wang, Y. Chai, X. Peng, and Y. Liu, "Inferring demographics from human trajectories and geographical context," *Computers, Environment and Urban Systems*, vol. 77, p. 101368, sep 2019.

[13] A. Roy and E. Pebesma, "A machine learning approach to demographic prediction using geohashes," in *Proceedings - 2017 2nd International Workshop on Social Sensing, SocialSens 2017 (part of CPS Week)*, 2017, pp. 15–20.

[14] J. Chen, J. Wu, H. Liang, S. Mumtaz, J. Li, K. Konstantin, A. K. Bashir, and R. Nawaz, "Collaborative trust blockchain based unbiased control transfer mechanism for industrial automation," *IEEE Transactions on Industry Applications*, vol. 56, no. 4, pp. 4478–4488, 2019.

[15] F. Qiao, J. Wu, J. Li, A. K. Bashir, S. Mumtaz, and U. Tariq, "Trustworthy edge storage orchestration in intelligent transportation systems using reinforcement learning," *IEEE Transactions on Intelligent Transportation Systems*, 2020.

[16] C. Kang, S. Gao, X. Lin, Y. Xiao, Y. Yuan, Y. Liu, and X. Ma, "Analyzing and geo-visualizing individual human mobility patterns using mobile call records," in *2010 18th International Conference on Geoinformatics, Geoinformatics 2010*, 2010.

[17] W. Li, H. Zhang, J. Chen, P. Li, Y. Yao, M. Shibasaki, X. Song, and R. Shibasaki, "Effective Metagraph-based Life Pattern Clustering with Big Human Mobility Data," apr 2021. [Online]. Available: http://arxiv.org/abs/2104.11968

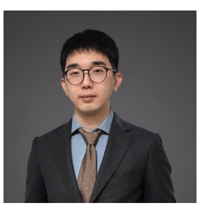
[18] D. M. Blei, A. Kucukelbir, and J. D. Mcauliffe, "Variational Inference: A Review for Statisticians," Tech. Rep., 2018.

[19] A. Y. Zhang and H. H. Zhou, "Theoretical and computational guarantees of mean field variational inference for community detection," *Annals of Statistics*, vol. 48, no. 5, pp. 2575–2598, oct 2020. [Online]. Available: https://doi.org/10.1214/19-AOS1898

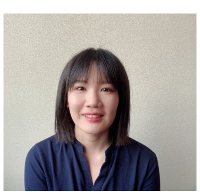
[20] Statistics Burea of Japan, "Survey on Time Use and Leisure Activities," 2011. [Online]. Available: http://www.stat.go.jp/english/data/shakai/2011/gaiyo.html

[21] M. Clerc, "Particle Swarm Optimization," *Particle Swarm Optimization*, pp. 1942–1948, 2010.

Peiran Li Peiran Li is now a Ph.D. candidate of the Center for Spatial Information Science, at The University of Tokyo. He has experience in the field of remote sensing, and currently his work focuses on urban computing and energy issues based on big data and AI tools.



Haoran Zhang Haoran Zhang is a researcher in the Center for Spatial Information Science at the University of Tokyo, and also works as a researcher at the Future Energy Center at Mälardalen University in Sweden, and Senior Scientist at Locationmind Inc. in Japan. His research includes smart supply chain technologies, GPS data in shared transportation, urban sustainable performance, GIS technologies in renewable energy systems, and smart cities.

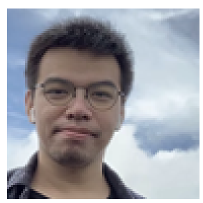


Wenjing Li Wenjing Li obtained an M.S. degree in the Graduate school of Frontier Sciences, the University of Tokyo. She obtained a B.E' from the Department of Landscape Architecture, Harbin Institute of Technology. She participates in several research works of spatial analysis and urban computing as a member of the Center for Spatial Information Science, the University of Tokyo. Her research focuses on human mobility, spatial science and data-driven urban studies.

Keping Yu Keping Yu received the M.E. and Ph.D. degrees from the Graduate School of Global Information and Telecommunication Studies, Waseda University, Tokyo, Japan, in 2012 and 2016, respectively. He was a Research Associate and a Junior Researcher with the Global Information and Telecommunication Institute, Waseda University, from 2015 to 2019 and 2019 to 2020, respectively, where he is currently a Researcher.

Ali Kashif Bashir ALI KASHIF BASHIR is a Senior Lecturer at the Department of Computing and Mathematics, Manchester Metropolitan University, United Kingdom. He is a senior member of IEEE and Distinguished Speaker of ACM. His past assignments include Associate Professor of Information and Communication Technologies, Faculty of Science and Technology, University of the Faroe Islands, Denmark; Osaka University, Japan (71 in QS Ranking 2020); Nara National College of Technology, Japan; the National Fusion Research Institute, South Korea; Southern Power Company Ltd., South Korea, and the Seoul Metropolitan Government, South Korea.

Ahmad Alzubi Ahmad Alzubi received the Ph.D. degree in computer networks engineering from the National Technical University of Ukraine (Ukraine), in 1999. He is currently a Professor with King Saud University (KSU). His current research interests include computer networks, grid computing, cloud computing, big data, and data extracting. He also served for three years as a consultant and a member of the Saudi National Team for measuring e-Government in Saudi Arabia.



Jinyu Chen Jinyu Chen is currently a PhD candidate in Center for Spatial Information Science, the University of Tokyo. He has published over 12 academic papers in peer-reviewed journals and conferences of both SCI and EI including Applied Energy and Advances in Applied Energy. He covers various fields including GPS trajectory data mining, deep learning and big data analysis.

Xuan Song Xuan Song received the Ph.D. degree in signal and information processing from Peking University in 2010. In 2017, he was selected as Excellent Young Researcher of Japan MEXT. He served as Associate Editor, Guest Editor, Program Chair, Area Chair, Program Committee Member or reviewer for many famous journals and top-tier conferences, such as IMWUT, IEEE Transactions on Multimedia, WWW Journal, Big Data Journal, ISTC, MIPR, ACM TIST, IEEE TKDE, UbiComp, ICCV, CVPR, ICRA and etc.

Ryosuke Shibasaki Ryosuke Shibasaki received the MS and PhD degrees in civil engineering from the University of Tokyo in 1982 and 1987, respectively. From 1982 to 1988, he was with the Public Works Research Institute, Ministry of Construction. From 1988 to 1991, he was an associate professor in the Civil Engineering Department, University of Tokyo. In 1991, he joined the Institute of Industrial Science, University of Tokyo. In 1998, he was promoted to a professor in the Center for Spatial Information Science, University of Tokyo. His research interest covers three-dimensional data acquisition for GIS, conceptual modeling for spatial objects, and agent-based microsimulation in a GIS environment.

APPENDIX A
BASELINE1: PSO BASED METHOD

Particle swarm optimization (PSO) is a powerful optimization algorithm family, it defines the set of candidate solutions as a swarm of particles that may flow through the parameter space, driven by their own and neighbors' best performances [21].

Basic Idea. To solve our problem by using the PSO algorithm, the definitions of 'particle' – the input parameters to be optimized, and 'aim' – the cost function to evaluate the input parameters should be given. In terms of input parameters, as mentioned above, there are so numerous users that we could not take all user's demographic characteristics as the input parameters. To construct a solvable input parameter form, we assume that each demographic group is Gaussian distributed in the life-pattern space (Appendix Figure A.1), thus, we only need four parameters to describe each demographic group (64 parameters in total for 16 demographic groups). Once the parameters of Gaussian distribution have been given, all users' demographic characteristics could be calculated. Through a series of iterations, we can achieve the best parameters (Appendix Figure A.1).

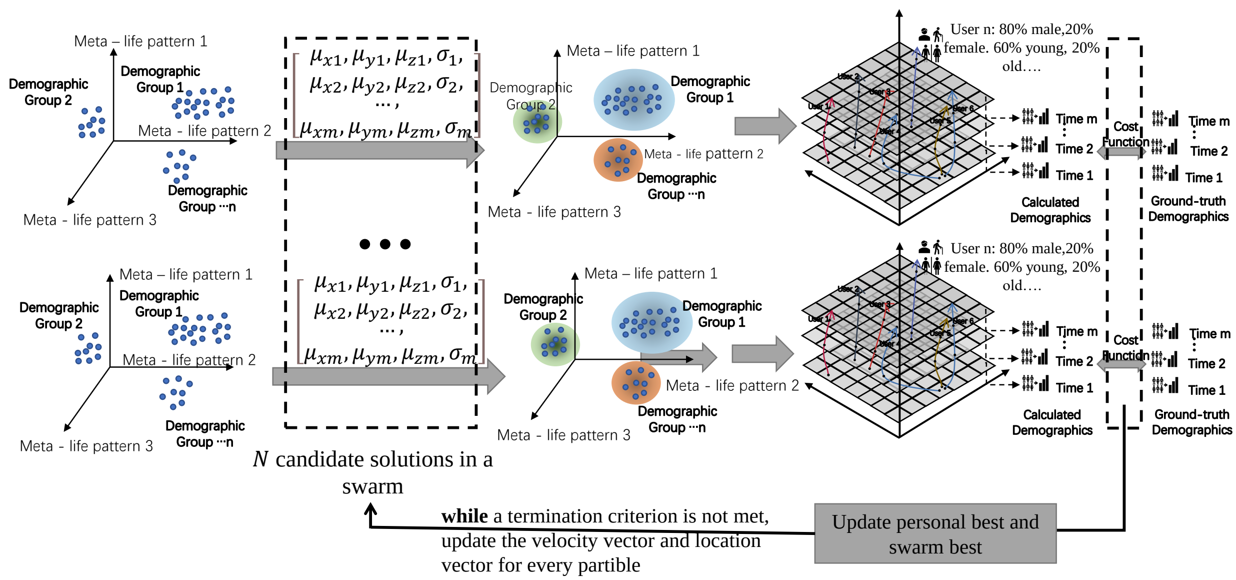


Fig. A.1: Illustration of PSO-based method.

Mathematical Description. According to the PSO theory, we define our problem as follows:

Suppose there are m demographic groups, the particles could be defined as:

$$\boldsymbol{x}_i = \left[\begin{array}{r} \mu_{x1}, \mu_{y1}, \mu_{z1}, \sigma_1, \mu_{x2}, \mu_{y2}, \mu_{z2}, \sigma_2, \ldots, \\ \mu_{xm}, \mu_{ym}, \mu_{zm}, \sigma_m \end{array} \right] \tag{A.1}$$

where μ_x, μ_y, μ_z stands for the center coordinate of a demographic group's Gaussian distribution in the life-pattern space, σ stands for this Gaussian distribution's standard error.

And N candidate solutions constitute the swarm:

$$\boldsymbol{X} = \{\boldsymbol{x_1}, \boldsymbol{x_2}, \ldots, \boldsymbol{x_N}\} \tag{A.2}$$

Then, for each user, the demographic characteristics could be calculated as:

$$\vec{p_i} = \left[\begin{array}{r} \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-\left((x-\mu_{x1})^2+(y-\mu_{y1})^2+(z-\mu_{z1})^2\right)}{2\delta^2}}, \ldots, \\ \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-\left((x-\mu_{xm})^2+(y-\mu_{ym})^2+(z-\mu_{zm})^2\right)}{2\delta^2}} \end{array} \right] \tag{A.3}$$

and the age/gender state matrix of all the users $\mathbb{U}$ should be:

$$\boldsymbol{A} = \{\vec{p_0}, \ldots, \vec{p_n}\} \tag{A.4}$$

Considering equation (3), we define the cost function f as:

$$f(\boldsymbol{x}_i) = \mathrm{MAE}\left(\overline{\boldsymbol{SA}}, \boldsymbol{S} \cdot \boldsymbol{A}^T \mid \boldsymbol{A}\right) \tag{A.5}$$

So far, we have completed the problem construction, and the fake code is shown as follow:

---

for each particle $i = 1, \ldots, S$ **do**
  Initialize the Gaussian distributions' parameters of demographic groups: $\mathbf{x_i} \sim U(\mathbf{b_{lo}}, \mathbf{b_{up}})$
  Initialize the particle's best-known parameters to its initial parameters: $\mathbf{p}_i \leftarrow \mathbf{x}_i$
  if $f(\mathbf{p_i}) < f(\mathbf{g})$ **then**
    update the swarm's best-known parameters: $\mathbf{g} \leftarrow \mathbf{p_i}$
  Initialize the particle's velocity:
  $\mathbf{v}_\mathrm{i} \sim U(-|\mathbf{b_{up}} - \mathbf{b_{lo}}|, |\mathbf{b_{up}} - \mathbf{b_{lo}}|)$
while a termination criterion is not met **do**:
  for each particle $i = 1, \ldots, S$ **do**
    for each dimension $d = 1, \ldots, n$ **do**
      Pick random numbers: $r_p, r_g \sim U(0,1)$
      Update the particle's velocity: $\mathbf{v}_{\mathrm{i,d}} \leftarrow \omega \mathbf{v}_{\mathrm{i,d}} + \varphi_\mathrm{p} r_\mathrm{p} (\mathbf{p}_{\mathrm{i,d}} - \mathbf{x}_{\mathrm{i,d}}) + \varphi_\mathrm{g} r_\mathrm{g} (\mathbf{g}_\mathrm{d} - \mathbf{x}_{\mathrm{i,d}})$
    Update the particle's parameters: $\mathbf{x}_\mathrm{i} \leftarrow \mathbf{x}_\mathrm{i} + \mathbf{lr} \mathbf{v}_\mathrm{i}$
    if $f(\mathbf{x}_\mathrm{i}) < f(\mathbf{p}_\mathrm{i})$ **then**

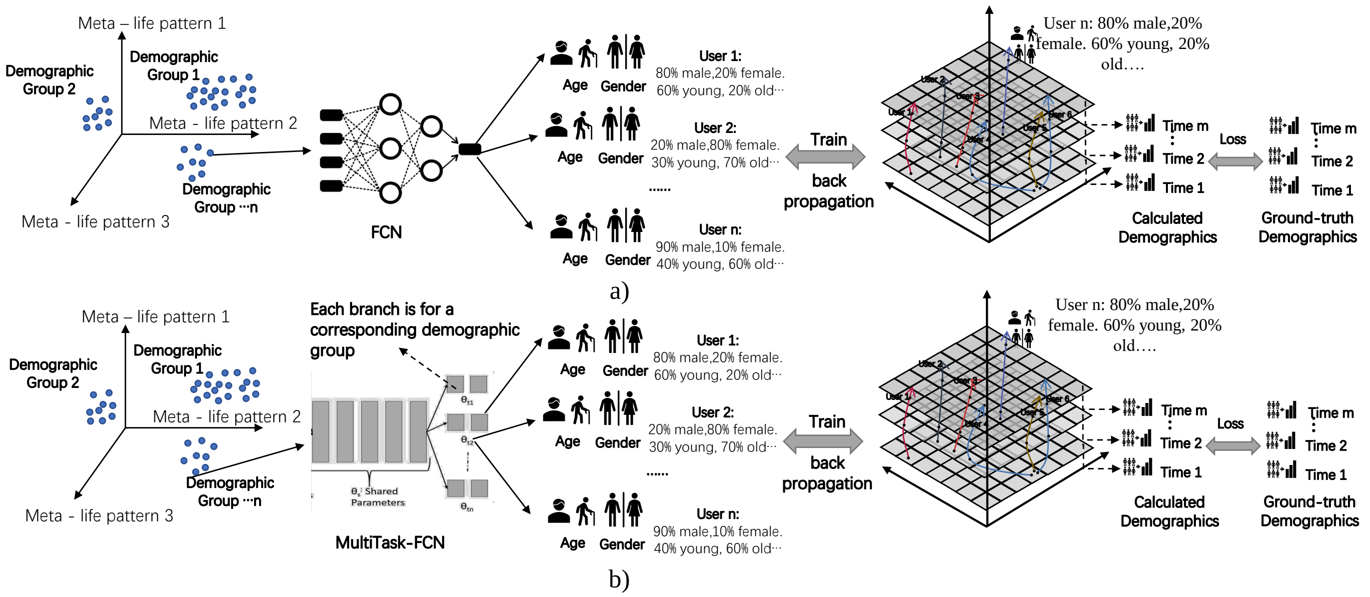


Fig. A.2: Flow of Deep Learning based method.

Update the particle's best-known parameters:
$\mathbf{p}_i \leftarrow \mathbf{x}_i$
if $f(\mathbf{p_i}) < f(\mathbf{g})$ **then**
  Update the swarm's best-known parameters:
  $\mathbf{g} \longleftarrow \mathbf{p_i}$

Finally, once the optimal parameters for Gaussian distributions of all demographic groups are determined, we can calculate the demographic characteristics of all users by equation (A.3).

APPENDIX B
BASELINE2: DEEP LEARNING BASED METHOD

Recent years have witnessed the magnificent booming of deep learning technology. Here, we employed an FCN and Multi-Task FCN to model the relationship between life-pattern features and users' demographic characteristics.

For FCN, we trained a model by taking life-pattern features (e.g. the x, y, z coordinates) as input and output a vector of possibilities that a user belongs to each demographic group. For Multi-Task FCN, the input is the same with FCN, but we used different output branches for different demographic groups. Each branch generates the possibility for the user of belonging to the corresponding demographic group, as A.2 shows.