# A Unified Framework for Modeling and Predicting Going-out Behavior

Shoji Tominaga, Masamichi Shimosaka, Rui Fukui, and Tomomasa Sato

Intelligent Cooperative Systems Laboratory, the University of Tokyo,
7-3-1 Hongo, Bunkyo-ku, Tokyo, 113-0033, Japan
{tominaga,simosaka,fukui,tsato}@ics.t.u-tokyo.ac.jp

**Abstract.** Living in society, to go out is almost inevitable for healthy life. There is increasing attention to it in many fields, including pervasive computing, medical science, etc. There are various factors affecting the daily going-out behavior such as the day of the week, the condition of one's health, and weather. We assume that a person has one's own rhythm or patterns of going out as a result of the factors. In this paper, we propose a non-parametric clustering method to extract one's rhythm of the daily going-out behavior and a prediction method of one's future presence using the extracted models. We collect time histories of going out/coming home (6 subjects, total 827 days). Experimental results show that our method copes with the complexity of patterns and flexibly adapts to unknown observation.

**Keywords:** Methodology, Activity recognition, Location representation

## 1 Introduction

Living in society and interacting with others, going outside one's house is almost inevitable for healthy life. There is increasing attention to this behavior, in many fields of study. For example, Kono et al. [10] show that the frequency of going out during one's period of life preceding old age affects activities of daily living (ADL) of several years later, and Gupta et al. [7] show that grasping going-out behavior of households saves their heating bill. However, Kono et al. used questionnaires about one's going out to collect data in their work. As Krumm and Brush [11] show modeling the rhythm of the behavior outperforms self-reported schedules, we believe that introducing sensor data processing to it enables deeper analyses in many fields such as pervasive computing, medical science, and life log, and the more precise prediction reduces the unnecessary cost of energy.

To predict someone's future behavior elementally needs estimation of someone's state at current time. Most of these researches in pervasive computing can be categorized into two approaches. One employs wearable sensors such as GPS loggers [7, 11] and the other installs sensors in the environment [14, 16]. When it comes to estimating the state of going out, it can be binarized as "home (inside)" or "away (outside)". Assuming most people go out through the individually same place (i.e. most people go out through the front entrance), sensing the passage of such places should be enough to estimate the state.

Modeling and predicting going out, can be thought as one of the problems to model and predict one's presence or occupancy. Recently, there are increasing studies with probabilistic approaches on this topic in the field of pervasive computing. In these approaches, methods predict the probability of future presence, and so they have some flexibility to unknown observation, even if there are not many training data. Krumm and Brush [11] introduced probabilistic models for going out. The method classifies the going-out data by the day of the week and also predicts one's presence using the pattern of the day. Scott et al. [16] introduced an occupancy prediction algorithm for controlling heating, ventilation, and cooling (HVAC). The algorithm predicts the future presence probability via matching the occupancy data by the current time to the past observation.

There are a lot of factors affecting one's daily going out, such as the day of the week, the condition of one's health, and weather. The behavior is influenced not by only one factor of them, but by the combination with one another. In addition, the number of factors and how they affect it are individually different. However, if we look at a person, there should be one's rhythm or patterns of it. For example, the rhythm of some people definitely differs between working days and holidays. To extract the one's own patterns and model the individually different behavior, prior knowledge cannot be put directory (e.g. the number of patterns). From a statistical point of view, intuitively deciding the number of patterns sometimes causes low performance. Too many patterns cause overfitting, and too few patterns cannot represent the complexity of the data. The more accurate modeling is feasible by estimating the appropriate number of patterns and correctly classifying the data.

There has been a significant amount of prior work in the field of statistics on extracting patterns or rhythm from the sensor data. For example, Farrahi and Gatica-Perez [5] used latent Dirichlet allocation and author-topic model, Ihler et al. [8] used time-varying Poisson processes, and Gill and Hangertner [6] used von-Mises distribution to extract the underlying rhythm. Actually, most of these studies of probabilistic approaches need the number of rhythms or categories a priori. There also exists work using event-mining approaches (e.g. the system of Rashidi and Cook [14]), however, these approaches need large-scale data. The nearest concept to that of ours is done by Shimosaka, et al. [18] with Dirichlet process mixtures (DPM) [2]. We develop this approach from raw data of a sensor as they used, to this abstracted behavior, going out.

In this work, we give three assumptions about daily going-out behavior. 1) It is done in a 24-hour cycle. 2) Each behavior of a day belongs to a certain category. 3) The number of categories is different by each person. These assumptions can also be used to predict the future observation: the target day itself also belongs to one of the categories of the person. You may think some patterns can be shared with other persons, and of course, it is often that going-out patterns of a person are quite similar to those of another. However, thanks to the data-driven approaches, our method eliminates such prior knowledge without performance drawbacks.

Our contribution is summarized as follows: we develop a unified framework for modeling and predicting going-out behavior, coping with the individual tendency. There are two key points in our method. One is that the method mathematically represents this complicated behavior affected by many factors and classifies the data with estimating the number of underlying categories simultaneously. The other is that the method predicts one's future presence from current observation of one day, by estimating to which category the day belongs. Our framework only needs time histories of going out/coming home, and so it is adaptable to many studies or systems in pervasive computing. We collect time histories of total 827 days of 6 subjects to evaluate our method. Experimental results show that our method flexibly copes with the complexity of going out and predicts the future observation with robust performance.

## 2 Collecting Time Histories of Going out/Coming Home

We employ two systems to accumulate time histories of going out and coming home: a tracking system and trail cameras. In this section, we show how these systems accumulate the data. As a result, we collect the time history data of 350 days (subject 1) and 239 days (subject 2) with the tracking system, and 31 days (subject 3) and 69 days (subject 4 – 6) with the trail cameras.

### 2.1 Collecting Time Histories via Tracking System

In the first time histories collection, we employed a human location tracking system [13] with range sensors and installed it into a one-bedroom type apartment for living alone. The size of the apartment is 4.9 meters by 9.5 meters. We installed five laser range finder (LRF) modules (Fig. 1-A), which are combinations of URG-LX04 (Hokuyo Automatic Co., Ltd.) and Armadillo-220 (Atmark Techno Inc.). The LRF modules are arranged at hip-height and located so that most area of the house is covered (the locations of the modules are shown in Fig. 1-B). The tracking system integrates the scan data, detects the resident by background subtraction, and tracks one's position by particle filter. In our work, the system automatically estimates the resident is out, if the system stops tracking at the entrance in more than 10 minutes. An example of trajectories just before the resident go out is shown in Fig. 1-B. Two subjects, both of whom were graduate students in their twenties from our laboratory lived in the house in a different period. Subject 1 lived from Apr. 1, 2009 to Mar. 16, 2010 and subject 2 from Apr. 14, 2010 to Dec. 23, 2010. Total number of days of each subject is 350 and 239, respectively. The original number of days of subject 2 is 255, however, we eliminate 16 days of subject 2 due to unexpected lack of trajectories.

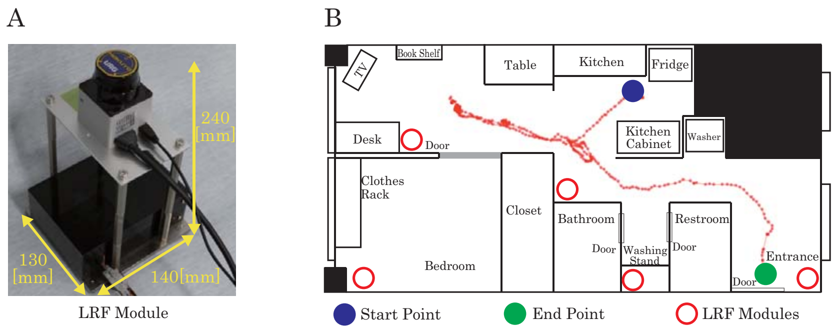


**Fig. 1.** Figure A shows the picture of a LRF module. Figure B shows the layout of the experimental house to calculate trajectories. The red circles in this figure are the position of the modules. The trajectory when the resident was just going out is written as a red line (the blue circle is the start point of it and the green circle is the end point).

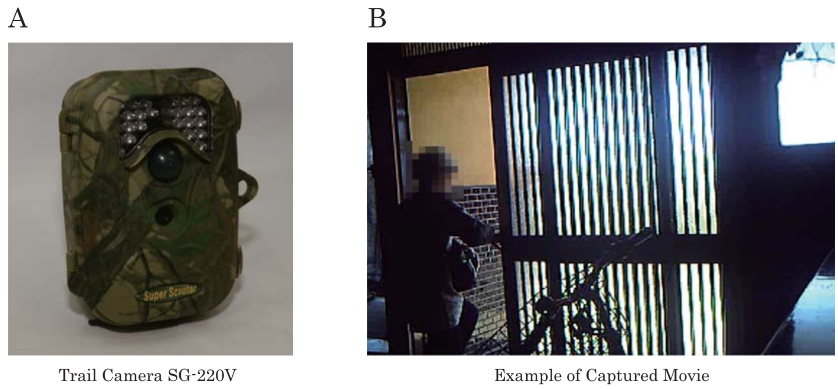


**Fig. 2.** Appearance of Trail Camera (Figure A) and Example of Captured Movie (Figure B): The movie was captured just before one of the subjects went out.

## 2.2 Collecting Time Histories via Trail Cameras

In the second time histories collection, we installed trail cameras SG-220V from Shenzhen Siyuan Digital Technology Co., Ltd. (see Fig. 2-A) in the entrance of houses. The camera has passive infrared ray sensors to detect people and records movies when someone passes in front of the camera. Though it is originally used to record behavior of wild animals, it can be also used as a simple security camera. In our work, we recruited 4 volunteers in 3 households, all of whom are elders, with some rewards. We met them, looked at the layout of their houses, and carefully decided where to set up the cameras with simple questionnaires. We also took care about their privacies, so that the camera recorded only the behavior of their entrance passage. A frame example of the captured movies is shown in Fig. 2-B. Since one's going out and coming home were obvious with the recorded movies, we collected time histories manually. As Krumm and Brush [11] did, GPS loggers can substitute for this work. Recording started in Aug. 21, 2011 at the earliest (subject 4 − 6: subject 4 and 5 lived together), not later than Sep. 28, 2011 (subject 3) and we used the data by Oct. 28, 2011.

## 3 Modeling Going-out Behavior

### 3.1 Outline

As described in Section 1, a state of one's going out at some time can be binarized (whether someone is away from home or not). The states is described as random variable $x = \{0, 1\}$ $(0 : \textit{home},\ 1 : \textit{away})$ following Bernoulli distribution $p(x|\mu) = \mu^x(1-\mu)^{1-x}$, where, $\mu = \mu(t)$ $(0 < \mu(t) < 1)$ is the time-varying parameter (e.g. if $\mu(t) = 0.9$, the person is out with a probability of 0.9 at the time $t$). For simplicity, the method discretizes $\mu(t)$, as a sequence $\{\mu_1, ..., \mu_T\}$ with length $T$. Each content of the sequence corresponds to the probability of going out at the corresponding time.

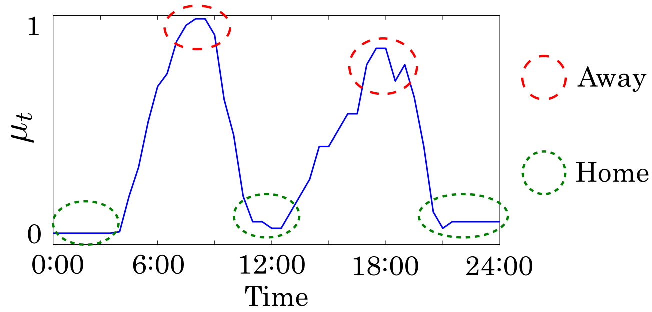


**Fig. 3.** Description Example of Going out

An example of parameter sequences is shown in Fig. 3. This example indicates that the person is likely to go out around 8 o'clock and 18 o'clock, and stay at home during night and around noon. Each category has its own pattern of parameter sequence.

### 3.2 Mathematical Description

Let $t = 1, ..., T$ be the time of a day, $n = 1, ..., N$ be the day of each observation, $k = 1, ..., K$ be the ID of the category, and $\boldsymbol{\mu}_k = (\mu_{k,1}\ ...\ \mu_{k,T})^\mathsf{T}$ be the parameter sequence of category $k$. We set the observation data of $n$-th day $\boldsymbol{x}_n = (x_{n,1}\ ...\ x_{n,T})^\mathsf{T}$ $(x_{n,t} = 0 : \textit{home},\ x_{n,t} = 1 : \textit{away})$ by majority decision of each time span of a day (if $T = 24$, $t = 1$ represents the time span from 0

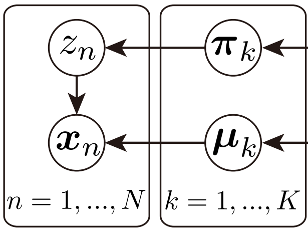


**Fig. 4.** Graphical Model of Going out

o'clock to 1 o'clock) The likelihood of the parameters for the data is written as the likelihood product of each time:

$$p(\boldsymbol{x}_n|\boldsymbol{\mu}_k) = \prod_{t=1}^{T} \mu_t^{x_{n,t}} (1 - \mu_{k,t})^{1-x_{n,t}}. \tag{1}$$

In addition, suppose $\{\mathbf{Z}\}_{n,k} = \{z_{n,k}|z_{n,k} = \{0,1\}, \sum_k z_{n,k} = 1\}$ be the parameters indicating to which category $k$ the data of $n$-th day belong, $z_n$ is a random variable following multinomial distribution $\mathcal{M}(z_n|\boldsymbol{\pi})$. Then, let $\mathbf{M} = (\boldsymbol{\mu}_1 \ ... \ \boldsymbol{\mu}_K)$ be parameter sequences of the categories, and the likelihood of the whole parameters are like the equation below:

$$p(\boldsymbol{x}_n|\mathbf{Z}, \mathbf{M}) = \prod_{k=1}^{K} \prod_{t=1}^{T} \left[ \mu_{k,t}^{x_{n,t}} (1 - \mu_{k,t})^{1-x_{n,t}} \right]^{z_{n,k}}. \tag{2}$$

Thanks to the simplicity of the model, it is possible to introduce hierarchical Bayesian representation for avoiding overfitting issues. This representation utilizes prior distributions for flexibility to unknown observations. Specifically, the method uses conjugate prior for each parameter, Dirichlet distribution for $\boldsymbol{\pi}$ and beta distribution for $\mu_{k,t}$:

$$\boldsymbol{\pi} \sim \mathcal{D}(\boldsymbol{\pi}|\alpha) \propto \prod_k \pi_k^{\alpha-1}, \tag{3}$$

$$\mu_{k,t} \sim \mathcal{B}(\mu_{k,t}|\beta_1, \beta_2) \propto \mu_{k,t}^{\beta_1-1} (1 - \mu_{k,t})^{\beta_2-1}. \tag{4}$$

Fig. 4 shows the graphical model of the representation. Each arrow represents the dependency between the parameters, for example, parameter $\boldsymbol{\mu}_k$ is generated using hyperparameters $\boldsymbol{\beta}$ (i.e. $\beta_1$, and $\beta_2$). The method estimates the model, by the posterior distribution of the parameters. For more details about this graphical representation, see Ref. [3].

### 3.3 Simultaneous Estimation of Parameters and Cluster Number

As we mentioned in Section 1, we cannot give the number of categories a priori. In this work, the method mentioned in Section 3.2 is extended to the simultaneous estimation of the model parameters and the cluster number. To tackle this problem, we employ Dirichlet process mixture (DPM) [2], a framework for Bayesian nonparametrics. DPM describes infinite Dirichlet distribution as a prior of categories. Compared with other methods to estimate the category size such as those introducing information criteria [1, 15], DPM parameterizes the category size distribution itself and provides flexible manners to estimate the number of clusters and parameters of each category simultaneously with less calculation cost. To implement DPM, the upper bound of category number $K$ is set as an enough bigger number than that expected (we set $K = 50$). The method estimates the number of them as $k \ll K$.

We use blocked Gibbs sampler [9], variational Bayes (VB) [4], collapsed Gibbs sampler [12] to implement DPM. The VB has an advantage of calculating cost since it provides obvious convergence and reduces the number of cycles, however, the solution of VB is completely affected by the initial state of learning so there exist local optima. In theory, blocked/collapsed sampler never causes the problems if the sampling cycles converge. Compared with each other, the blocked sampler has more chances to reach global optima than the other, and also more risks to diverge. As our prior experiment to evaluate the three, the model by the blocked sampler has the highest likelihood to the observation, and so we actually utilize this method.

The blocked Gibbs sampler for DPM utilizes truncated stick-breaking process [17] to approximate the infinite-dimensional Dirichlet distribution. The process represents beta distribution as a prior of each coefficient of multinomial distribution. In this case, coefficients $\pi_k$ are represented as:

$$\pi_k = v_k \prod_{i=1}^{k-1} (1 - v_i), \quad v_k \sim \mathcal{B}(v_k | 1, \alpha), \tag{5}$$

where, $v_K = 1$ for the maximum number of categories, $K$. Learning by the blocked sampler consists of two steps. Firstly, the method initializes the parameters. The method samples $\mathbf{Z}$, so that all data are allocated randomly and equally to each category. Lastly, the method iteratively alternate resampling from the posterior distribution (6), and (7):

$$\mathbf{V}, \mathbf{M} \sim p(\mathbf{V}, \mathbf{M} | \mathbf{X}, \mathbf{Z}), \tag{6}$$

$$\mathbf{Z} \sim p(\mathbf{Z} | \mathbf{X}, \mathbf{V}, \mathbf{M}). \tag{7}$$

In our model, the posterior distributions are expanded as:

$$v_k \sim \mathcal{B}(v_k | 1 + \sum_n z_{n,k}, \alpha + \sum_{i=k+1}^{K} \sum_n z_{n,i}), \tag{8}$$

$$\mu_{k,t} \sim \mathcal{B}(\mu_{k,t} | \beta_1 + \sum_n x_{n,t} z_{n,k}, \beta_2 + \sum_n (1 - x_{n,t}) z_{n,k}), \tag{9}$$

$$z_n \sim \mathcal{M}(z_n | \boldsymbol{\pi}^*), \quad \pi_k^* := \frac{\pi_k p(\boldsymbol{x}_n | \boldsymbol{\mu}_k)}{\sum_k \pi_k p(\boldsymbol{x}_n | \boldsymbol{\mu}_k)}. \tag{10}$$

During the iteration, most of categories have no data allocated, and the data aggregate into some classes. Since there cannot be complete convergence in Gibbs sampling, the number of iteration is set to 100 in our experiment. Finally, parameters of the posterior distribution $v_k \sim \mathcal{B}(v_k|\alpha^\star_{k,1}, \alpha^\star_{k,2})$, $\mu_{k,t} \sim \mathcal{B}(\mu_{k,t}|\beta^\star_{k,t,1}, \beta^\star_{k,t,2})$ are:

$$\begin{cases} \alpha^\star_{k,1} = 1 + \sum_n z_{n,k}. \\ \alpha^\star_{k,2} = \alpha + \sum_{i=k+1}^{K} \sum_n z_{n,i} \end{cases}, \tag{11}$$

$$\begin{cases} \beta^\star_{k,t,1} = \beta_1 + \sum_n x_{n,t} z_{n,k}. \\ \beta^\star_{k,t,2} = \beta_2 + \sum_n (1 - x_{n,t}) z_{n,k} \end{cases}. \tag{12}$$

# 4 Predicting Going-out Behavior

## 4.1 Problem Setting and Outline

As we mentioned in Section 1, many practical applications need prediction of future presence (e.g. controlling HVAC). In this section, we show our method of predicting one's future presence utilizing our going-out model (Section 3). We assume that one's observation of a new day should also belong to a category of one's own. The method represents the probabilistic distribution of future observation as a linear sum of a pattern of each category. The coefficient of patterns, sum of which is 1, is proportional to likelihood of each category given the observation by the current time of the target day.

## 4.2 Calculating Probability of New Observations

Given new observations $\boldsymbol{x}^\star_{1:T} := (x^\star_1 \ ... \ x^\star_T)^\intercal$ of one day, the approximate parameter likelihood of the posterior distribution (Section 3) is represented as:

$$p(\boldsymbol{x}^\star_{1:T}|\boldsymbol{\alpha}^\star, \boldsymbol{\beta}^\star) = \sum_k \mathbb{E}_{\mathbf{V}}[\pi_k] p(\boldsymbol{x}^\star_{1:T}|\beta^\star_{k,1:T,1}, \beta^\star_{k,1:T,2}). \tag{13}$$

There, each term on the right can be expanded into the equation below.

$$\mathbb{E}_{\mathbf{V}}[\pi_k] = \mathbb{E}[v_k] \prod_{i=1}^{k-1} (1 - \mathbb{E}[v_i]), \ \mathbb{E}[v_k] = \begin{cases} 1 & (k = K) \\ \frac{\alpha^\star_{k,1}}{\alpha^\star_{k,1} + \alpha^\star_{k,2}} & \text{(otherwise)} \end{cases}, \tag{14}$$

$$\begin{aligned} p(\boldsymbol{x}^\star_{1:T}|\beta^\star_{k,1:T,1}, \beta^\star_{k,1:T,2}) &= \prod_{t=1}^{T} \int p(x^\star_t|\mu) \mathcal{B}(\mu|\beta^\star_{k,t,1}, \beta^\star_{k,t,2}) d\mu \\ &= \prod_{t=1}^{T} \frac{Beta(x^\star_t + \beta^\star_{k,t,1}, 1 - x^\star_t + \beta^\star_{k,t,2})}{Beta(\beta^\star_{k,t,1}, \beta^\star_{k,t,2})}, \end{aligned} \tag{15}$$

where, $Beta(\cdot)$ is beta function $Beta(x, y) = \int_0^1 t^{x-1}(1-t)^{y-1} dt$. Each term of product in (15) follows beta-binomial distribution (the number of trials is fixed to 1 in this case).

### 4.3 Prediction Algorithm

Given observation data from time $t = 1$ to $t = t_p$ ($< T$), the probability that each category pattern generates the data can be described using the representation of Section 4.2 as $p(\boldsymbol{x}^\star_{1:t_p}, z^\star_k = 1|\boldsymbol{\alpha}^\star, \boldsymbol{\beta}^\star) = \mathbb{E}_{\mathbf{V}}[\pi_k]p(\boldsymbol{x}^\star_{1:t_p}|\beta^\star_{k,1:t_p,1}, \beta^\star_{k,1:t_p,2})$. Let $\gamma_{k,1:t_p}$ be the probability that the target day belongs to category $k$, and this can be described as:

$$\gamma_{k,1:t_p} := p(z^\star_k = 1|\boldsymbol{x}^\star_{1:t_p}) = \frac{p(\boldsymbol{x}^\star_{1:t_p}, z^\star_k = 1|\boldsymbol{\alpha}^\star, \boldsymbol{\beta}^\star)}{\sum_c p(\boldsymbol{x}^\star_{1:t_p}, z^\star_c = 1|\boldsymbol{\alpha}^\star, \boldsymbol{\beta}^\star)}. \tag{16}$$

The method predicts the presence probability of time $t_f$ ($> t_p$) by a linear sum of category patterns, with coefficient $\gamma_{k,1:t_p}$ as,

$$\begin{aligned} p(x^\star_{t_f} = 1|\boldsymbol{x}^\star_{1:t_p}, \boldsymbol{\alpha}^\star, \boldsymbol{\beta}^\star) &= \sum_k \gamma_{k,1:t_p} p(x^\star_{t_f} = 1|\beta^\star_{k,t_f,1}, \beta^\star_{k,t_f,2}) \\ &= \sum_k \gamma_{k,1:t_p} \frac{Beta(1 + \beta^\star_{k,t_f,1}, \beta^\star_{k,t_f,2})}{Beta(\beta^\star_{k,t_f,1}, \beta^\star_{k,t_f,2})} \\ &= \sum_k \gamma_{k,1:t_p} \frac{\beta^\star_{k,t_f,1}}{\beta^\star_{k,t_f,1} + \beta^\star_{k,t_f,2}}. \end{aligned} \tag{17}$$

For practical applications, the method needs to decide one's future state with the probability. We set the parameter $\tau$ ($0 \le \tau \le 1$), as a threshold to output prediction. If probability of one's going out exceeds $\tau$, the method predicts the one will be out, otherwise it predicts the one will be at home.

## 5 Results

### 5.1 Experiment Setting

The modeling and prediction performance of our method is evaluated with the time histories data collections described in Section 2. To evaluate the modeling method, we use log-likelihood as criteria, and to evaluate the predicting method, we use the classification accuracy of future observation from the past observation. The technical details are in following paragraphs.

For the former method, we see the convergence by the log-likelihood of parameters with all the data of each subject and the flexibility performance by expectation value of log-likelihood leave-one-out cross validation (LOOCV). The log-likelihood of training data shows the adaptation to the complexity of training data. The log-likelihood of LOOCV represents the flexibility to the unknown observation.

For the latter method, we also use LOOCV to predict unknown observations from the other data of the same subject. As more concrete setting, the method predicts the future observation of a day, given the 6, 9, and 12 hours data of the day. The method also predicts the presence by updating each time the method is given the presence of time $t$. Since the threshold of probability highly depends on applications, we set the threshold $\tau$ from 0 to 1 by 0.01 and evaluate the performance by receiver operating characteristic (ROC) curves and the area under the curves (AUC). Each predictive output is evaluated by the state of every minute. For example, suppose that a subject is out in 20 minutes of a time duration $t$ and at home in the rest 10 minutes and the method predicts the one is out, then true positive frames will be 20 and false positive frames will be 10. Note that the performance does not necessarily improve as the observation time of the day progresses, since the applied time of prediction varies from each experiment (e.g. if 6 hours given, the method predicts the rest 18 hours, and if given 12 hours, the method predicts shorter time, 12 hours).

Throughout the experiment, we set the start time of days to 0 o'clock, each time span of a day to 30 minutes (i.e. $T = 48$), the modeling hyper parameters $\alpha = 0.1$ empirically, and $\beta_1 = \beta_2 = 1$ so that prior knowledge about one's home/away presence is eliminated. In this condition, prior of $\mu \sim \mathcal{B}(\mu|\beta_1, \beta_2)$ will be the uniform distribution from 0 to 1.

For comparison to our methods, we implement the presence model of Krumm and Brush [11] (we call this, KB model) for the experiments of both modeling and predicting, and the algorithm for predicting presence by Scott et al. [16] (we call this, Scott's algorithm) for experiment of predicting. In the modeling experiment, we also implement KB mixture model, the mixture model of each pattern of the day of week, the coefficient of which are the ratio of number of days of each pattern to the total number, usually one seventh). Since KB model does not use the current observation of the target day and this might be unfair, we modify KB mixture so as to utilize the current observation. The future distribution is the linear sum of each pattern, and coefficients are proportional to the likelihood of patterns for the given observation. Actually Krumm and Brush also utilize the drive time prediction [7] to improve their performance. In this experiment we do not implement this method since we collect the presence data not by GPS loggers. The Scott algorithm could be affected by how many nearest neighbors are used to estimate future observation, so we set the number $k$ to not only 5 (as Scott et al. used), but also $3, 7$ for prediction.

### 5.2 Modeling Going out

In this section, we evaluate our modeling method by likelihood to the unknown data. With our implementation, learning the model converges in less than 1 second for each subject.

Table 1 shows log-likelihood expectation values of training data by using all data of each subject and Table 2 shows those during LOOCV of each subject. We examine the performance by separating the subjects into two groups. One is the group of subjects who has long-term data (subject 1 and 2) and the other is that of subjects who has short-term data (subject $3 - 6$). With the group of long-term data, our method has high performance in both training data and test data. This suggests that KB model and its mixture, which classify the data by the day of the week, cannot represent the long-term data since seven is not necessarily

**Table 1.** Log-likelihood Expectation Value for Training Data

| Subj. | 1 | 2 | 3 |
|---|---|---|---|
| Proposed | $-16.2 \pm 5.5$ | $-9.5 \pm 5.2$ | $-8.8 \pm 4.9$ |
| KB | $-31.7 \pm 5.2$ | $-22.2 \pm 12.7$ | $-8.4 \pm 5.1$ |
| KB mixture | $-30.4 \pm 4.1$ | $-20.0 \pm 9.0$ | $-7.9 \pm 5.2$ |
| Subj. | 4 | 5 | 6 |
| Proposed | $-9.0 \pm 5.2$ | $-5.5 \pm 4.3$ | $-10.9 \pm 5.3$ |
| KB | $-7.9 \pm 5.1$ | $-4.8 \pm 4.4$ | $-13.5 \pm 7.7$ |
| KB mixture | $-8.4 \pm 5.5$ | $-5.1 \pm 4.7$ | $-13.1 \pm 7.1$ |

**Table 2.** Log-likelihood Expectation Value of LOOCV

| Subj. | 1 | 2 | 3 |
|---|---|---|---|
| Proposed | $\mathbf{-18.7 \pm 7.4}$ | $\mathbf{-11.1 \pm 7.3}$ | $-9.8 \pm 5.6$ |
| KB | $-32.7 \pm 5.5$ | $-24.5 \pm 18.5$ | $-52.7 \pm 76.6$ |
| KB mixture | $-30.6 \pm 4.1$ | $-20.2 \pm 9.2$ | $-9.1 \pm 6.6$ |
| Subj. | 4 | 5 | 6 |
| Proposed | $\mathbf{-9.7 \pm 6.5}$ | $-6.3 \pm 5.6$ | $\mathbf{-12.7 \pm 7.6}$ |
| KB | $-22.8 \pm 46.3$ | $-15.9 \pm 26.6$ | $-32.4 \pm 59.7$ |
| KB mixture | $-11.0 \pm 12.5$ | $-5.5 \pm 5.2$ | $-20.4 \pm 47.8$ |

enough number of the category. With the group of short-term data, especially in subject 4 and 6, the KB model and its mixture overfit the training data (low likelihood for test data in spite of high likelihood for training data). Even though the compared methods have great gaps of likelihood between training and test data of these subjects, our method flexibly copes with the unknown observation with low gaps. In contrast, with subject 3 and 5, our method has slightly lower LOOCV performance than KB mixture model. The main cause of this result is related with hierarchical Bayes. We set parameter $\beta_1 = \beta_2 = 1$, and this means all category patterns have additional information about one's presence, one home observation and one away observation (i.e. the probability of one's home cannot be 0, even if the person is at home during the all data). In this regard, KB mixture gets higher score in subject 3 and 5. However, the disadvantage of additional information in prior distribution seen in these subjects is trivial, compared with the advantage of flexibility with the data size and appropriateness about categorizing data seen in subject 1, 2, 4, and 6.

Clustered patterns of subject 2 and 6 by our method are shown in Fig. 5. Each graph in the figure represents the going-out pattern of the corresponding category and each value in the graph is the expectation value of posterior distribution (i.e. the probability of being away from home). Patterns of each subject are sorted into descending order by the number of allocated days and patterns with less than 3 days are eliminated. For example, type 8 represents the subject goes out around 10 o'clock and comes home around 20 o'clock. The figure shows that our method can extract peculiar patterns such as type 3 (the subject are out almost all day) and type 6 (the subject are at home almost all day). The patterns

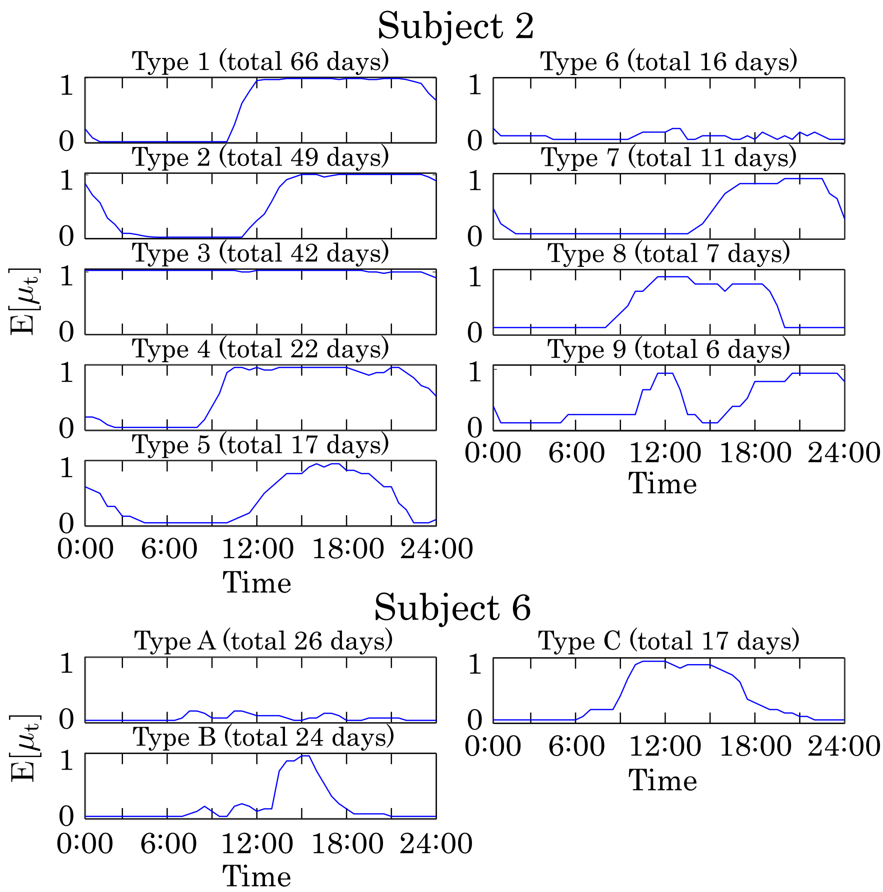


**Fig. 5.** Examples of Extracted Typical Patterns (Upper: Subj. 2, Lower: Subj. 6): To emphasize the patterns of the subjects are extracted separately, we use different representation about categories between each subject.

of subject 2 sum up to be 98 percent of total days (236/239), so it can be said to the subject that the number of the patterns is 9. The 3 outliers are allocated into two groups (two days and one day), both of which are the patterns of going-out twice in different periods of a day. Similarly, the number of patterns of subject 6 is 3. For other subjects 1, 3, 4, and 5, the number is 22, 2, 2, 2, respectively. The number of patterns has to do with the size of data. Visualization examples of continuous 140 days of subject 2 and all days of subject 6 are shown in Fig. 6. Categories in the figure corresponds to those in Fig. 5. Not only our method automatically classifies the data into appropriate groups, but also it enables us to analyze long-term going-out data by visualizing each pattern and the sequence of classification result, for example, from the end of the 2nd week to the start of the 3rd week, subject 2 might be traveling.

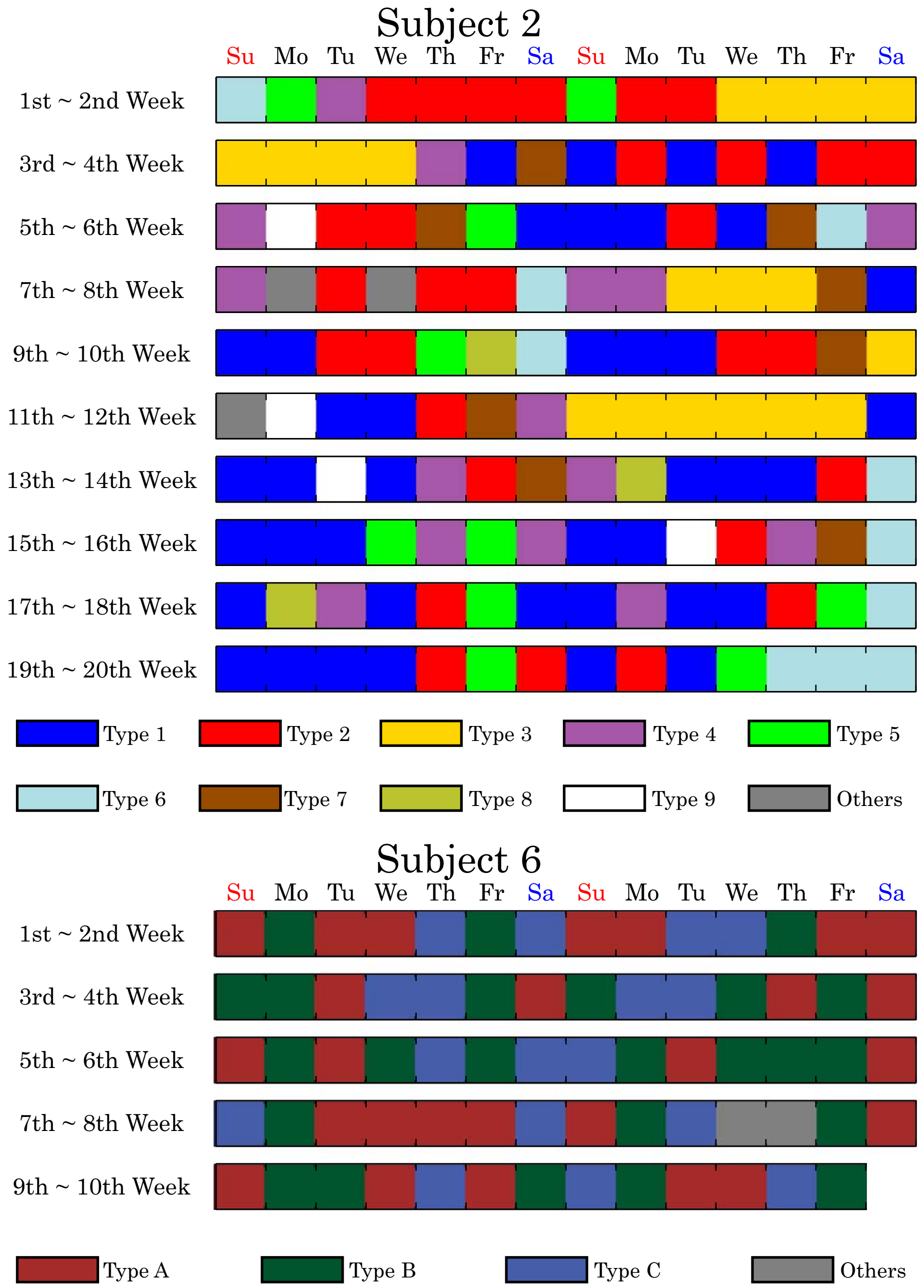


**Fig. 6.** Examples of Classification Result (Upper: Continuous 140 Days of Subj. 2, Lower: All 69 Days of Subj. 6)

**Table 3.** Area under the ROC Curves (upper-left: using first 6 hours, upper-right: using first 9 hours, lower-left: using first 12 hours, lower-right: by updating from observation of each time)

| Subj. | 1 | 2 | 3 | 4 | 5 | 6 | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Proposed | 0.68 | 0.87 | 0.79 | 0.78 | 0.77 | 0.79 | 0.63 | 0.79 | 0.73 | 0.76 | 0.75 | 0.80 |
| KB | 0.60 | 0.83 | 0.68 | 0.73 | 0.69 | 0.79 | 0.55 | 0.78 | 0.66 | 0.73 | 0.69 | 0.79 |
| KB mixture | 0.62 | 0.82 | 0.80 | 0.78 | 0.76 | 0.78 | 0.56 | 0.75 | 0.77 | 0.76 | 0.76 | 0.79 |
| Scott($k = 3$) | 0.65 | 0.81 | 0.77 | 0.57 | 0.52 | 0.67 | 0.62 | 0.76 | 0.67 | 0.57 | 0.52 | 0.74 |
| Scott($k = 5$) | 0.66 | 0.86 | 0.77 | 0.56 | 0.54 | 0.76 | 0.63 | 0.77 | 0.74 | 0.57 | 0.55 | 0.76 |
| Scott($k = 7$) | 0.67 | 0.87 | 0.79 | 0.54 | 0.54 | 0.78 | 0.63 | 0.78 | 0.75 | 0.59 | 0.57 | 0.79 |

| Subj. | 1 | 2 | 3 | 4 | 5 | 6 | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Proposed | 0.62 | 0.74 | 0.84 | 0.78 | 0.80 | 0.85 | 0.90 | 0.97 | 0.93 | 0.84 | 0.85 | 0.91 |
| KB | 0.50 | 0.71 | 0.71 | 0.74 | 0.65 | 0.82 | 0.60 | 0.82 | 0.72 | 0.78 | 0.73 | 0.83 |
| KB mixture | 0.54 | 0.74 | 0.86 | 0.78 | 0.86 | 0.84 | 0.70 | 0.90 | 0.94 | 0.86 | 0.90 | 0.88 |
| Scott($k = 3$) | 0.60 | 0.71 | 0.70 | 0.55 | 0.68 | 0.81 | 0.91 | 0.94 | 0.91 | 0.79 | 0.72 | 0.90 |
| Scott($k = 5$) | 0.60 | 0.72 | 0.72 | 0.57 | 0.69 | 0.82 | 0.92 | 0.95 | 0.92 | 0.81 | 0.72 | 0.91 |
| Scott($k = 7$) | 0.61 | 0.74 | 0.79 | 0.58 | 0.72 | 0.84 | 0.92 | 0.95 | 0.92 | 0.82 | 0.76 | 0.91 |

### 5.3 Presence Prediction

In this section, we evaluate our method of predicting future presence of new data. Fig. 7 shows the ROC curves with different conditions and subjects by changing the threshold $\tau$ from 0 to 1. The frames of true positive, true negative, false positive, and false negative are summed up with two groups by the term of data. One group contains subject 1 and 2 (subjects of long-term data, graphs on the left), and the other contains the rest subjects (subjects of short-term data, graphs on the right). The AUC values with individual subject are shown in Table 3. From the results, our method has stable high performance by all subjects and all conditions. KB or KB mixture is not good at incremental prediction of subjects of long-term data. The main cause of this result is the same as that of the modeling experiment in Section 5.2. That is, KB model cannot handle the complexity of long-term data. Scott algorithm is not good at subject of short-term data, except the incremental prediction. The main cause is that this algorithm relatively needs plenty of past observation for stable prediction since the core of the algorithm is $k$-nearest neighbor. Compared with KB or KB mixture, our method has an advantage of handling long-term data. Compared with Scott algorithm, our method has an advantage of handling short-term data. It can also be said that our predictive output is the real number (more finer than that of Scott algorithm, which is the rational number), so our method have more flexibility in the trade-off between the true positive and the false positive. It is found that our method does not have the best performance in all subjects and conditions, however, our method achieves the best or nearly the best of all, in all subjects and conditions.

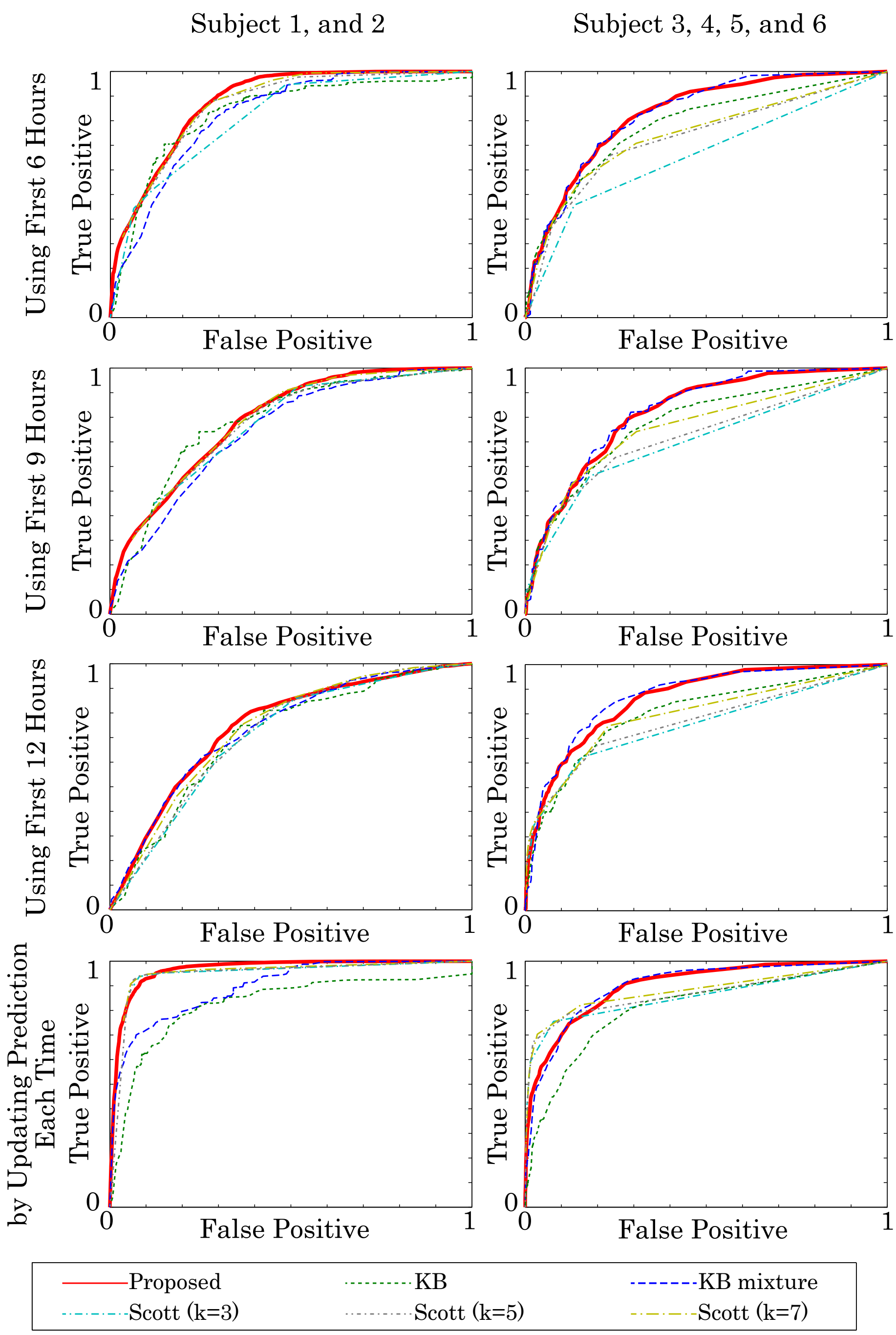


**Fig. 7.** Receiver Operating Characteristic Curves of Prediction

## 6 Discussion

We give as little prior knowledge as possible for one's going out to our method. Indeed, there are few parameters in our modeling and predicting method. In our method, there are four parameters $\alpha$, $\beta_1$, $\beta_2$, and $\tau$. However, under the precondition that there are no information about one's going out (uniform distribution from 0 to 1 as we did in the whole experiment), the parameters are in effect only $\alpha$ and $\tau$. $\alpha$ is the parameter indicating the degree of category separation. As this parameter grows, our method tends to make more categories. $\tau$ is the threshold of prediction, and the proper value differs according to the kinds of applications. Thus, our methods can avoid the troublesome parameter estimation. Experimental results comparing our methods with existing methods suggest modeling going out definitely by the day of the week may cause low performance. There should have relationship between the day of the week and going-out behavior, however, there should also be other unignorable factors affecting the rhythm of the behavior. We believe that even if there is a person whose going-out patterns are entirely decided by the day of the week, our method can classify the date as so. However, we still need the larger data sets of more people to find correspondence between the behavior rhythm and factors such as weather, the day of the week, and social events.

There are two additional advantages of our method in practical use: easiness of updating the model and possibility of making another category. When new data are given, the method updates the parameters by only two steps: assigning the data to the categories with probabilities that each category generates the data (i.e. $\gamma_{k,1:T}$ in Section 4.3), and updating the parameters of the posterior distributions. That is, the method does not need to cluster all the data again. In addition, if the data are too different from the patterns of the categories, our method can make another category and assign the data to it. This can be used to detect the anomaly of one's life.

As we mentioned in Section 1, we regard the problem of modeling and predicting going out, as one of the problems of modeling and predicting one's presence or occupancy (someone is there or not there at that time). Our method can also be used to predict one's occupancy at certain places such as rooms in a house, an office, and a laboratory. In addition, it can be possible that our method represents other behavior of people that can be binarized.

## 7 Conclusion

In this work, we propose a unified framework for modeling and predicting one's going-out behavior. We assume that it is cyclic and observation of a single day independently belongs to a certain category. Our modeling method estimates the number of categories and assigns the data into the corresponding category simultaneously. Our predicting method outputs the future state of one's going out of a day by estimating the category of the day. We collected the going-out data of 6 subjects and total 827 days by employing a tracking system and trail cameras. Experimental results comparing our methods with existing methods and improved existing methods show that our method achieves the best or nearly the best of all, in all subjects and conditions stably. In addition, since patterns extracted by our framework are easy to understand, the results also show the possibility of helping people analyze the long-term data of one's going out.

As our future work, we will collect the larger dataset of more people, not only for more accurate evaluation of our framework, but also for finding correspondence between going-out behavior rhythm and underlying factors.

## Acknowledgement

We desire to express our thanks to Taketoshi Mori and Hiroshi Noguchi for helping us to collect data by a laser range finder tracking system, Masanori Ogihara and Shinya Masuda for negotiation and execution of setting up trail cameras with us, local government officials of Hokuto-shi, Yamanashi prefecture, Japan for permitting us to set up trail cameras into houses, all subjects for giving us data of going out, and all anonymous reviewers for fair review and constructive suggestions.

## References

1. Akaike, H.: Information theory and an extension of the maximum likelihood principle. In: Proceedings of the Second International Symposium on Information Theory. pp. 267 – 281 (1973)
2. Antoniak, C.E.: Mixtures of Dirichlet processes with applications to Bayesian nonparametric problems. The Annals of Statistics 2(6), 1152 – 1174 (1974)
3. Bishop, C.M.: Pattern Recognition and Machine Learning. Springer-Verlag (2006)
4. Blei, D.M., Jordan, M.I.: Variational inference for Dirichlet process mixtures. International Society for Bayesian Analysis 1, 121 – 144 (2006)
5. Farrahi, K., Gatica-Perez, D.: Discovering routines from large-scale human locations using probabilistic topic models. ACM Transactions on Intelligent Systems and Technology 2, 3:1 – 27 (2011)
6. Gill, J., Hangartner, D.: Circular data in political science and how to handle it. Political Analysis 18(3) (2010)
7. Gupta, M., Intille, S.S., Larson, K.: Adding GPS-control to traditional thermostats: An exploration of potential energy savings and design challenges. In: Proceedings of International Conference on Pervasive Computing. pp. 95 – 114 (2009)
8. Ihler, A., Hutchins, J., Smyth, P.: Adaptive event detection with time-varying Poisson processes. In: Proceedings of Knowledge Discovery and Data Mining. pp. 207 – 216 (2006)
9. Ishwaran, H., James, L.F.: Gibbs sampling methods for stick-breaking priors. Journal of the American Statistical Association 96, 161 – 174 (2001)
10. Kono, A., Kai, I., Sakato, C., Rubenstein, L.Z.: Frequency of going outdoors predicts long-range functional change among ambulatory frail elders living at home. Archives of Gerontology and Geriatrics 45(3), 233 – 242 (2007)
11. Krumm, J., Brush, A.J.B.: Learning time-based presence probabilities. In: Proceedings of International Conference on Pervasive Computing. pp. 79 – 96 (2011)

12. Maceachern, S.N.: Estimating normal means with a conjugate style Dirichlet process prior. Communications in Statistics B 23, 727 – 741 (1994)
13. Noguchi, H., Urushibata, R., Sato, T., Mori, T., Sato, T.: System for tracking human position by multiple laser range finders deployed in existing home environment. Aging Friendly Technology for Health and Independence pp. 226 – 229 (2010)
14. Rashidi, P., Cook, D.J.: Mining sensor streams for discovering human activity patterns over time. In: Proceedings of International Conference on Data Mining. pp. 431 – 440 (2010)
15. Schwarz, G.: Estimating the dimension of a model. The Annals of Statistics 6(2), 461 – 464 (1978)
16. Scott, J., Brush, A.J.B., Krumm, J., Meyers, B., Hazas, M., Hodges, S., Villar, N.: PreHeat: controlling home heating using occupancy prediction. In: Proceedings of the International Conference on Ubiquitous Computing. pp. 281 – 290 (2011)
17. Sethuraman, J.: A constructive definition of Dirichlet priors. Statistica Sinica 4, 639 – 650 (1994)
18. Shimosaka, M., Ishino, T., Noguchi, H., Sato, T., Mori, T.: Detecting human activity profiles with Dirichlet enhanced inhomogeneous Poisson processes. In: Proceedings of International Conference on Pattern Recognition. pp. 4384 – 4387 (2010)