Detecting human activity profiles with Dirichlet enhanced inhomogeneous Poisson processes

Masamichi Shimosaka, Takahito Ishino, Hiroshi Noguchi, Tomomasa Sato and Taketoshi Mori
The University of Tokyo
{*simosaka,ishino,noguchi,tmori*}*@ics.t.u-tokyo.ac.jp*

Abstract

This paper describes an activity pattern mining method via inhomogeneous Poisson point processes (IPPPs) from time-series of count data generated in behavior detection by pyroelectric sensors. IPPP reflects the idea that typical human activity is rhythmic and periodic. We also focus on the idea that activity patterns are affected by exogenous phenomena, such as the day of the week, and weather condition. Because single IPPP could not tackle this idea, Dirichlet process mixtures (DPM) are leveraged in order to discriminate and discover different activity patterns caused by such factors. The use of DPM leads us to discover the appropriate number of the typical daily patterns automatically. Experimental result using long-term count data shows that our model successfully and efficiently discovers typical daily patterns.

1. Introduction

As sensor and storage technologies continue to improve in terms of both cost and performance, increasingly rich data sets are becoming available that characterize the rhythms of human activity over time. Example includes logs of freeway traffic data, crime statistics, e-mail and Web access log, and surveillance scenario. Since these data measure the aggregated behavior of individual human beings, it typically exhibits a periodicity in time that reflects the rhythms of human activity. Hence, these data can be used to support a variety of different applications such as classification of human activities and detection of anomaly events. Especially, monitoring behavior of solitary elderly residence with cheap pyroelectric sensors becomes important as a practical approach to avoid unexpected fatal accidents and to provide safe and secure for their life, because rapid aging of the population resulting from the decline in the birthrate in most advanced nations will increase solitary elderly people. Analysis for typical rhythms promotes support application for predicting demented state and for making urgent messages.

To make analysis practical, we must take consideration for exogenous phenomena such as weekdays / weekends, rainy / sunny from long term sensor logs as well as to model underlying intensity of the data, where intensity here refers to the rate at which events occur and to detect. Hence our paper focuses not only on how inferring rhythms linked to human activity but also on finding multiple rhythms varying across multiple day recordings. Specifically, we tackle a data mining problem of finding activity profiles specific to categories of human activities and determining the number of the categories. In this paper, we propose a flexible nonparametric Bayesian approach with inhomogeneous Poisson processes to provide scalable and systematic manner of both estimating and finding multiple rhythms of human activity.

1.1. Related works

There has been a significant amount of prior work in statistics on estimating the underlying activity rhythms (time-intensity profiles) of the sensor log. The traditional approach is using inhomogeneous Poisson processes [12]. It contains both a rate constant and time intensity function. Traditionally, Gaussian processes [11], a nonparametric kernel method, are often used for modeling time-intensity function because of their flexibility. The model is called log Gauss-Cox processes [12] in this case. Unfortunately, no analytical posterior sampling techniques are available in log Gauss-Cox processes until relatively-recent years. The first exact sampling technique proposed by Adams et al. [1] is infeasible for long term dataset handled in our paper. Instead of leveraging Gaussian processes, recent advances focuses on repetition or circular property of the rhythms with parametric density models. For example, Gill and Hangartner [6] proposed a framework using von-Mises distribution and hierarchical Bayesian approach. Because no conjugate priors are available in the model, the model needs sampling methods with auxiliary variables [3] or rejection sampling [6]. In our empirical study, these sampling techniques are inefficient to draw the posterior samples of concentration parameter [11]. Kottas et al. [10] proposed mixture of the beta distribution for time-intensity function. Similar to the case of von-Mises distribution, the posterior sampling of the parameters of the beta distribution is inefficient. Recent advances from pattern recognition community focus on simplifying the complicated inhomogeneous Poisson processes with Gaussian mixture model or Dirichlet distribution [8, 9]. This leads us to have capability of handling a large scale dataset. We also use such simple approaches in this paper.

Although most of them in the field of modeling activity focus on how making feasible time-intensity function, it is noteworthy that the underlying rhythms of human activity varies across days as well as hours due to exogenous phenomena. We focus on these phenomena and extend inhomogeneous Poisson processes to detect multiple activity patterns. Most simple approach to categorize multiple activity patterns across days leverages parametric mixture model [8], however, the number of categories must be given a priori. Indeed additive model [9] promotes multiple patterns of human activity, but the model also needs a number of clusters a priori. Furthermore it does not find categories of daily behaviors but decomposes time-intensity into several factors similar to latent topic models. This discourages readability from the result. Our model relaxes the predetermined assumption of groups by using nonparametric Bayesian approach to provide systematic manner for mining human activity profiles.

2. A baseline model representing rhythms

2.1. Modeling periodicity by Poisson processes

As mentioned above, we assume that human activity is inferred by the counting data generated and aggregated by pyroelectric sensors. Perhaps most common and simple probability model representing counting data is Poisson distribution. Let $m \in \mathbb{N}_0$ depict the observed counts in a fixed time interval, then the discrete Poisson distribution [11] is given by $p(m|\lambda) = \frac{1}{m!}\lambda^m \exp(-\lambda)$ with a rate parameter λ, or averaged number of counts in a fixed time interval. Because the rate parameter is constant with respect to time, it is not proper to use Poisson distribution to analyze rhythm linked to human activity. Hence we leverage Poisson process [12] as a baseline model, a generalization of Poisson distribution. In contrast to the Poisson distribution, the Poisson processes vary the rate function with time. Specifically, the process is characterized by a time varying rate function $\lambda(t)$ instead of using fixed number λ. When the number of the events aggregated from time t from the starting point is M_t, and the events occur $m = M_{t_2} - M_{t_1}$ times within t_1 and $t_2 > t_1$, then the process in this interval is equivalent to Poisson distribution. The likelihood of m can be written as

$$p(m|\lambda(\cdot)) = \frac{1}{m!}\left[\int_{t_1}^{t_2} \lambda(t)\mathrm{d}t\right]^m \exp\left\{-\int_{t_1}^{t_2} \lambda(t)\mathrm{d}t\right\}.$$

Let us suppose that we have a collection of event times $\tau_1, \tau_2, \ldots, \tau_N$ from the beginning to the end (i.e. 0 hours $\leq \tau_n \leq$ 24 hours), the likelihood of the data $\{\tau_n\}_{n=1}^N$ is given by $p(\{\tau_n\}_{n=1}^N|\lambda(\cdot)) = \exp\left\{-\int_0^T \lambda(t)\mathrm{d}t\right\}\prod_{n=1}^N \lambda(\tau_n)$. We introduce a fixed positive number λ_0 for the normalization constant of $\lambda(t)$. Specifically let λ_0 be $\lambda_0 = \int_0^T \lambda(t)\mathrm{d}t$, then we decompose $\lambda(t)$ with $f(t)$ and λ_0 as $f(t) = \lambda(t)/\lambda_0$. In this setting, $f(t)$ can be viewed as a time-intensity function linked to the rated function $\lambda(t)$, because $f(\cdot)$ satisfies $1 = \int f(t)\mathrm{d}t$. The likelihood is rewritten as

$$p(\{\tau_n\}_{n=1}^N|\lambda_0, f(\cdot)) = \exp(-\lambda_0)\lambda_0^N \prod_{n=1}^N f(\tau_n) \quad (1)$$

The term involving $f(\cdot)$ has the same form as the likelihood of the τ_n as i.i.d. samples from the probability density function of continuous scalar data if we omit the term $\exp(-\lambda_0)\lambda_0^N$ from (1).

2.2. Histogram modeling by count data

Under the assumption that we have collection of event times, Gaussian mixture models are available for modeling time-intensity function [9]. In this paper, we think of another case where time-series data is discrete and the data are available only in grouped format. Here we refer grouped format to histogram data of counts by pyroelectric sensors within a fixed time interval. This situation often occurs when we want to make wireless sensor network systems run over a year with only batteries. In this section, adjustment from time stamp data to discrete count data will be given.

Let $\Delta = T/S$ be the unit interval for aggregating the counting data, where T represents 24 hours and S depicts the number of segmentation within a day. Here we refer positive number h_s as the number of counts from time $s\Delta$ to time $(s+1)\Delta$. Let us suppose that we have a collection of counting grouped data $\{h_s\}_{s=1}^S$ and the time-intensity function $f(\cdot)$ in (1) is constant within a fixed length interval $t \in [s\Delta, (s+1)\Delta)$, $f(t) = r_s$, then the likelihood of the data is simplified as

$$p(h_1, \ldots, h_S|\lambda_0, \boldsymbol{r}) = \exp(-\lambda_0)\lambda_0^N \prod_{s=1}^S r_s^{h_s}, \quad (2)$$

where N depicts the sum of counts as $N = \sum_s h_s$. Note that r_s satisfies $\sum_s r_s = 1, r_s \geq 0$, simplex over

$\mathbb{R}^S$. Let $\boldsymbol{r}$ be a column vector of $r_1, \dots, r_S$ and $\mathbb{R}^S_{\mathsf{SIM}}$ be a collection of simplex over $\mathbb{R}^S$.

2.3. Inference with hierarchical Bayes

In our model, learning and inference is accomplished through estimating parameter λ_0 and $\boldsymbol{r}$ defined in (2). In this paper, inference is accomplished by Bayesian techniques instead of maximum likelihood estimation to avoid overfitting and to be used data mining framework mentioned later. To provide scalable Bayesian inference, conjugate prior [11] must be given. Fortunately, the model mentioned above has conjugate prior.

Designing prior We use Gamma distribution as a prior for λ_0 as $p(\lambda_0) = \mathcal{G}(\lambda_0|a, b) \propto \lambda_0^{a-1}\exp(-b\lambda_0)$with hyperparameter $a > 0$, $b > 0$. Dirichlet distribution is used as a prior for $\boldsymbol{r} \in \mathbb{R}^S_{\mathsf{SIM}}$. The density is defined as $p(\boldsymbol{r}) = \mathcal{D}(\boldsymbol{r}|\boldsymbol{\alpha}) \propto \prod_s r_s^{\alpha_s - 1}$ with positive hyperparameter $\alpha_s > 0$.

Posterior distribution Suppose that we have collection of multiple days of counting data. Specifically, the dataset contains D day data. Here we refer $h_s^{(d)}$ to the number of counts as s-th segment in d-th day. To make it convenient, we concatenate activity data in d-th day $h_s^{(d)}$ into a vector $\boldsymbol{h}^{(d)} \in \mathbb{N}_0^S$. Let $\boldsymbol{H} = \{\boldsymbol{h}^{(d)}\}_{d=1}^D$ be a collection of the counting data used for inference and $N_d = \sum_{s=1}^S h_s^{(d)}$ be a total number of counting in d-th day. Under the assumption mentioned above, the posterior for the parameter λ_0 and $\boldsymbol{r}$ are defined as $p(\lambda_0|\boldsymbol{H}) = \mathcal{G}\left(\lambda_0|\tilde{a}, \tilde{b}\right)$, $p(\boldsymbol{r}|\boldsymbol{H}) = \mathcal{D}\left(\boldsymbol{r}|\tilde{\boldsymbol{\alpha}}\right)$, where $\tilde{a} = a + \sum_{d=1}^D N_d$, $\tilde{b} = b + D$ and $\tilde{\boldsymbol{\alpha}} = \boldsymbol{\alpha} + \sum_{d=1}^D \boldsymbol{h}^{(d)}$ denotes updated hyperparameter. Since the form of the posterior is conjugate to the prior, these parameters could be learned efficiently. The marginal likelihood could be obtained in closed-form similar to [7].

3. Dirichlet enhanced Poisson processes for detecting activity profiles

To consider the time-intensity functions varying across multiple day recordings, we must extend our baseline model. The key idea here is to separate observations into groups, and yet concatenate day recordings in each group to make statistical strength. One simple idea is to use simple parametric mixture model. Another approach is to use day-specific method proposed by Ihler et al. [8]. These approaches are too restrictive for human activity modeling. In the Bayesian methodology, such sharing is provided naturally through hierarchical modeling, especially Dirichlet process mixtures.

3.1. Dirichlet process mixtures

Dirichlet process (DP) mixtures [5, 4] are gaining popularity in machine learning community in recent years because of the flexibility for clustering applications where the number of clusters is unknown a priori. In this paper we formulate a DP mixture model in the stick-breaking representation [2] to accomplish the idea mentioned above. Let ϕ denote a prior on some space Θ. In our model, Θ corresponds to $\lambda_0, \boldsymbol{r}$, and Dirichlet distribution for $\boldsymbol{r}$ and Gamma distribution for λ_0 is linked to prior ϕ. A *Dirichlet process* (DP) with concentration parameter β, denoted by $\mathrm{DP}(\beta, \phi)$ defines a prior over infinite mixtures:

$$\pi_k = v_k \prod_{l=1}^{k-1} (1 - v_l), \quad v_l \sim \mathcal{B}(v_k|1, \beta) \tag{3}$$

$$p(\boldsymbol{h}|\boldsymbol{\pi}, \theta_1, \dots,) = \sum_{k=1}^{\infty} \pi_k g(\boldsymbol{h}|\theta_k), \tag{4}$$

where $\mathcal{B}$ denotes Beta distribution as $\mathcal{B}(x|b_1, b_2) \propto x^{b_1-1}(1-x)^{b_2-1}$ and $g(\cdot|\theta_k)$ denotes our parametric inhomogeneous Poisson processes to represent the k-th rhythm of human daily activity. Component parameters are independently drawn as $\theta_k \sim \phi$. In this setting, we assume that $\boldsymbol{h^{(d)}}$ is generated by the topic selection with multinomial distribution $\mathcal{M}(z|\boldsymbol{\pi}) = \pi_z$ as $z_d \sim \mathcal{M}(z_d|\boldsymbol{\pi})$, and $\boldsymbol{h}^{(d)} \sim g(\boldsymbol{h}^{(d)}|\boldsymbol{\theta}_{z_d})$.

3.2. Truncation and truncated Gibbs Sampling

To implement DP efficiently, we use truncation of DP mixtures [11]. Under this assumption, the likelihood of DP mixtures defines a finite K mixture models: $p(\boldsymbol{h}|\boldsymbol{\pi}, \theta_1, \dots,) = \sum_{k=1}^K \pi_k g(\boldsymbol{h}|\theta_k)$. Note that $v_K = 1$. It is important to note that the truncation level K is not taken to be the number of mixture components observed in the data, but rather a loose upper bound on that number. In this experiment, we set truncation level $K = 30$. Given truncation level K, our truncated Gibbs sampler alternates between blocked resampling of pattern assignment indices z and the parameter λ_0, $\boldsymbol{r}$ in each activity pattern.

4. Experimental result

In this section, we consider the application of our model to the activity log of solitary elderly people inside residence and provide a reasonable result of our model.

Data Human activity dataset used in the experiment contains a counting data over a year where five pyroelectric sensors are attached in the residence. The sensors are attached in kitchen, living room, bed room, corridor and entrance. The counting refers to a number of detection of people movement per one minute. This means $S = 1440$. Although it is very severe to grasp residence behavior with one minute counting data, such cheap cost systems lead us to gather activity log over

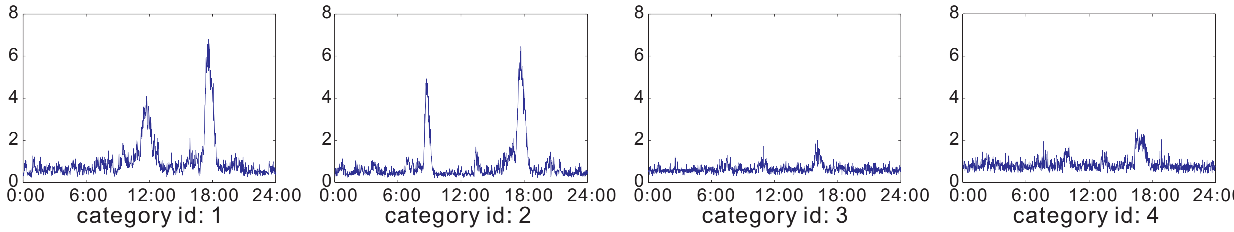

Figure 1. Extracted activity patterns $\lambda(t)$ at kitchen from 3 monthes dataset

months. We select a sensor attached in kitchen from the five sensors because the log of this sensor is assumed to be highly linked to the rhythm of human activity. We extract a part of dataset because it has a deficit due to the error of wireless communication between sensor logger and sensors. Specifically, we choose a consecutive 90 days sensor log from the dataset (Jan. 2007 – Mar. 2007). This consists of $123,085$ counts in total: $1,300$ times event detection per one day, 0.95 times event detection per minute. Figure 2 shows the result of averaged aggregated count data over 90 days.

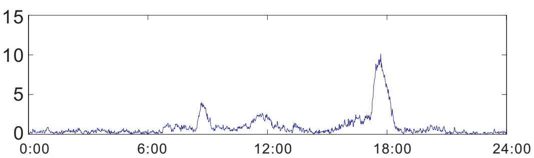

Figure 2. Mean activity pattern at kitchen

Result Figure 1 shows the activity profiles detected by our method under the condition that we set hyperparameter $\alpha_s = 50$, then set a, b from the mean and the variance of the counts per day. The number of detected patterns is 4. Each pattern contains 40, 40, 5, and 5 days, respectively. The inference is accomplished within 10 sec. From Figure 2, the residence seems to have meals three times per day, however, the residence goes out in daylight when breakfast is taken in early hours. Instead, the time residence takes breafkfast is not clear when he / she does not go out at noon. It becomes apparent that it contains a day with low activity. Note that the extracted pattern at late-evening is active than the actual, because we set α_s large number.

5. Conclusion

The increasing availability of logs of human activity provides interesting opportunities for the application of statistical learning techniques. In this paper, we proposed a nonparametric Bayesian framework with parametric inhomogeneous Poisson processes from the counting data to learning and detecting time-intensity profiles. The key idea is to separate observations into groups, and concatenate the recordings in each group to make statistical strength. Experimental result shows that our approach successfully splits and groups the dataset to categorize the activity patterns based on the underlying rhythms of human activity. Note that our approach can be available by introducing Gaussian mixture models when the dataset contains time-stamped data instead of counting data by using *nested* Dirichlet processes [13]. Directions for further work in this area include richer models that allow for incorporation of observed covariates such as weather and other exogenous phenomena, as well as multi-sensor problem.

References

[1] R. P. Adams, I. Murray, and D. J. MacKay. Tractable nonparametric Bayesian inference in Poisson processes with Gaussian process intensities. In *Proc. of ICML 2009*.

[2] D. Blei and M. Jordan. Variational methods for the Dirichlet process. In *Proc. of ICML2004*.

[3] P. Damien and S. Walker. A full Bayesian analysis of circular data using the von Mises distribution. *Canadian J. of Stats.*, 27(2), 1999.

[4] M. Escobar and M. West. Bayesian density estimation and inference using mixtures. *J. of the American Stats. Assoc.*, 90(430), 1995.

[5] T. Ferguson. A Bayesian analysis of some nonparametric problems. *Annals of Stats.*, 1(2), 1973.

[6] J. Gill and D. Hangartner. Bayesian circular-linear regression. Seminar on Bayesian Inference in Econometrics and Statistics, 2009.

[7] K. A. Heller and Z. Ghahramani. Bayesian Hierarchical Clustering. In *Proc. of ICML 2005*.

[8] A. Ihler, J. Hutchins, and P. Smyth. Adaptive event detection with time-varying Poisson processes. In *Proc. of KDD 2006*.

[9] A. T. Ihler and P. Smyth. Learning time-intensity profiles of human activity using non-parametric Bayesian models. In *Advances in NIPS 19*, 2007.

[10] A. Kottas. Dirichlet process mixtures of Beta distributions, with applications to density and intensity estimation. In *Proc. of Workshop on Learning with Nonparametric Bayesian Methods (conjunction with ICML 2006)*.

[11] D. J. MacKay. *Information Theory, Inference, and Learning Algorithms*. Cambride University Press, 2003.

[12] J. Moller, A. Syversveen, and R. Waagepetersen. Log Gaussian Cox processes. *Scandanavian J. of Stats.*, 25(3), 1998.

[13] A. Rodriguez, D. B. Dunson, and A. E. Gelfand. The nested Dirichlet process. *J. of the American Stats. Assoc.*, 103(483):1131–1144, 2008.