Discovery of Regular and Irregular Spatio-Temporal Patterns from Location-Based SNS by Diffusion-Type Estimation

Yoshitatsu MATSUDA\textsuperscript{(a)}, Nonmember, Kazunori YAMAGUCHI\textsuperscript{(b)}, Member, and Ken-ichiro NISHIOKA\textsuperscript{(c)}, Nonmember

SUMMARY In this paper, a new approach is proposed for extracting the spatio-temporal patterns from a location-based social networking system (SNS) such as Foursquare. The proposed approach consists of the following procedures. First, the spatio-temporal behaviors of users in SNS are approximated as a probabilistic distribution by using a diffusion-type formula. Since the SNS datasets generally consist of sparse check-in’s of users at some time points and locations, it is difficult to investigate the spatio-temporal patterns on a wide range of time and space scales. The proposed method can estimate such wide range patterns by smoothing the sparse datasets by a diffusion-type formula. It is crucial in this method to estimate robustly the scale parameter by giving a prior generative model on check-in’s of users. The robust estimation enables the method to extract appropriate patterns even in small local areas. Next, the covariance matrix among the time points is calculated from the estimated distribution. Then, the principal eigenfunctions are approximately extracted as the spatio-temporal patterns from a location-based SNS dataset. It consists of the following procedures. First, the sparse SNS dataset is smoothed by employing a two-dimensional diffusion-type formula \((2\pi t)^{-1} \exp \left(-|x - y|^2/2\right)\) and the spatio-temporal distribution of users is estimated. The diffusion-type formula is often used in quantum mechanics \cite{6} and seismology \cite{7}, and it gives a natural model for estimating the movement of users when the location information between the check-in’s is not available. While the scale parameter is a physical constant in quantum mechanics, it must be estimated statistically from the check-in’s in the SNS dataset. It is crucial in the proposed method to estimate the scale parameter robustly by a Bayesian method with a prior model. Second, the covariance matrix among the time points is calculated from the estimated distribution. Then, the various spatio-temporal patterns are extracted from the covariance matrix by using principal component analysis (PCA), which is a widely used method in signal processing and feature extraction.

This paper is organized as follows. In Sect. 2, the proposed approach is described. Section 2.1 describes the estimation of the spatio-temporal distribution by a diffusion-type formula. Section 2.2 describes the estimation method of the scale parameter. In Sect. 2.3, PCA is applied to the distribution in order to extract the spatio-temporal patterns. In Sect. 3, the proposed approach is applied to an actual SNS dataset from Foursquare in order to verify its effectiveness. Section 4 concludes this paper. This paper is an extended and elaborated version of our previous publication \cite{8}.

1. Introduction

Recently, many location-based social networking systems (SNS) such as Foursquare\textsuperscript{*}, Facebook Places\textsuperscript{**}, and Waze\textsuperscript{***} have emerged. From such datasets, various information such as points of interests \cite{1}, traffic patterns \cite{2}, patterns across different temporal scales \cite{3}, and personal behavior patterns \cite{4} have been extracted. In addition, the extracted information has been used in some applications, for example, for improving road safety \cite{5}. Though location-based SNS datasets are useful, one of its weak points is its sparseness. The SNS records on many users are many sequences of check-in’s at some time points and locations, where the locations of the users are unknown in the intervals between the check-in’s. In addition, the intervals are non-periodic and uncontrollable. Therefore, it is difficult to estimate the spatio-temporal distribution in SNS on a wide-range of time and space scales. Furthermore, even if a wide-range spatio-temporal distribution is given, it is difficult to investigate the distribution, which is a complicated mixture of various unknown spatio-temporal patterns.

In this paper, a new approach for analyzing a location-based SNS dataset is proposed. It consists of the following procedures. First, the sparse SNS dataset is smoothed by employing a two-dimensional diffusion-type formula \((2\pi t)^{-1} \exp \left(-|x - y|^2/2\right)\) and the spatio-temporal distribution of users is estimated. The diffusion-type formula is often used in quantum mechanics \cite{6} and seismology \cite{7}, and it gives a natural model for estimating the movement of users when the location information between the check-in’s is not available. While the scale parameter is a physical constant in quantum mechanics, it must be estimated statistically from the check-in’s in the SNS dataset. It is crucial in the proposed method to estimate the scale parameter robustly by a Bayesian method with a prior model. Second, the covariance matrix among the time points is calculated from the estimated distribution. Then, the various spatio-temporal patterns are extracted from the covariance matrix by using principal component analysis (PCA), which is a widely used method in signal processing and feature extraction.

This paper is organized as follows. In Sect. 2, the proposed approach is described. Section 2.1 describes the estimation of the spatio-temporal distribution by a diffusion-type formula. Section 2.2 describes the estimation method of the scale parameter. In Sect. 2.3, PCA is applied to the distribution in order to extract the spatio-temporal patterns. In Sect. 3, the proposed approach is applied to an actual SNS dataset from Foursquare in order to verify its effectiveness. Section 4 concludes this paper. This paper is an extended and elaborated version of our previous publication \cite{8}.

2. Method

2.1 Estimation of Spatio-Temporal Distribution

Here, the method for estimating the spatio-temporal distribution from a location-based SNS dataset is described. Each check-in of a user in a location-based SNS indicates a specific location and a time point (often with some

\textsuperscript{*}https://foursquare.com/
\textsuperscript{**}https://www.facebook.com/about/location
\textsuperscript{***}http://www.waze.com/
function defined as \( x \) which moves from a check-in to a two-dimensional flat surface for simplicity. This assumption is suitable if the range of the check-in’s is no larger than a continent. As there is no information between the check-in’s, it is assumed that each user moves as a free particle, which moves from a check-in \( x_t \) to a check-in \( y_u \) with the probability \( P_{\text{diff}}(y|xt, u) \) in the following equation (a two-dimensional diffusion-type formula):

\[
P_{\text{diff}}(y|xt, u) = \frac{\lambda}{2\pi(u-t)} \exp\left(-\frac{\lambda |x - y|^2}{2(u-t)}\right) \tag{1}
\]

where \( \lambda \) is the scale parameter on the time and \( |x - y|^2 \) is the square of the Euclidean distance between \( x \) and \( y \). The estimation of \( \lambda \) is discussed in Sect.2.2. Here, it is assumed to be given as an appropriate constant. Note that the time point \( u \) is regarded as a given variable in the conditional distribution. \( P_{\text{diff}} \) in Eq. (1) has the following preferable property. Let \( z \) be a candidate check-in on a path from \( xt \) to \( yu \). It is also assumed that \( P(y|zt, u) \) and \( P(z|xt, v) \) are given by \( P_{\text{diff}} \). Then, \( P(y|xt, u) \) is given as the following marginalized form w.r.t. \( z \) and \( v \):

\[
P(y|xt, u) = \int P(y|zt, u) P_{\text{diff}}(z|xt, v) P(v) \, dz \, dv
\]

\[
= \int P_{\text{diff}}(y|zt, u) P(v) \, dv = P_{\text{diff}}(y|xt, u), \tag{2}
\]

where \( P(v) \) is any distribution of the candidate time point \( v \) which does not affect the final equation. This shows that Eq. (1) gives a consistent framework.

Now, \( P(x|x_{t_i}, x_{t_{i+1}}, t_i) \) is estimated for \( t_i \leq t < t_{i+1} \). We assume the 1-Markov property of check-in’s by \( P(x_i|xt, x_{t_{i+1}}, t_i) = P_{\text{diff}}(x_{t_{i+1}}|xt, t_i) \). Then, \( P(x|x_{t_i}, x_{t_{i+1}}, t_i) \) is given as

\[
P(x|x_{t_i}, x_{t_{i+1}}, t_i) = \frac{P(x, x_{t_{i+1}}|xt, t_{i+1}, t_i)}{P(x_{t_{i+1}}|xt, t_{i+1})} = \frac{P_{\text{diff}}(x_{t_{i+1}}|xt, t_{i+1}) P_{\text{diff}}(x|x_{t_i}, t_i)}{P_{\text{diff}}(x_{t_{i+1}}|xt, t_{i+1})} = N_x(\mu_{\text{dist}}, \sigma_{\text{dist}}^2) \tag{3}
\]

where \( N_x(\mu, \sigma^2) \) is a two-dimensional circular Gaussian function defined as

\[
N_x(\mu, \sigma^2) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{|x - \mu|^2}{2\sigma^2}\right). \tag{4}
\]

\( \mu_{\text{dist}} \) and \( \sigma_{\text{dist}}^2 \) are given as

\[
\mu_{\text{dist}} = \frac{(t - t_i) x_{t_{i+1}} + (t_{i+1} - t) x_i}{t_{i+1} - t_i} \tag{5}
\]

and

\[
\sigma_{\text{dist}}^2 = \frac{(t - t_i)(t_{i+1} - t)}{\lambda(t_{i+1} - t_i)} \tag{6}
\]

respectively. The mean \( \mu_{\text{dist}} \) is the location linearly interpolating \( x_i \) and \( x_{t_{i+1}} \) with the ratio of \( t \) in \([t_i, t_{i+1}]\). The variance \( \sigma_{\text{dist}}^2 \) takes the maximum at the middle of \( t_i \) and \( t_{i+1} \), and it takes the minimum 0 at \( t = t_i \) or \( t = t_{i+1} \). Therefore, \( P(x|x_{t_i}, x_{t_{i+1}}, t) \) distributes widely at the middle of \( t_i \) and \( t_{i+1} \), and it sharply (as the Dirac delta function) converges to \( x_i \) at \( t_i \) and \( x_{t_{i+1}} \) at \( t_{i+1} \). Regarding the multiple users, the check-in sequence of a user \( \omega \) is denoted as \( (x_{t_{1,\omega}}, x_{t_{2,\omega}}, \ldots, x_{t_{n_{\omega}}}) \). If the dataset consists of \( N \) users, the average of Eq. (3) over all users is utilized. Thus, \( P(x|t) \) for all users is given as

\[
P(x|t) = \frac{1}{N} \sum_{\omega} P(x|x_{t_{1,\omega}}, x_{t_{2,\omega}}, x_{t_{1,\omega}}, t) \tag{7}
\]

where \( i \) is such that \( t_i \leq t \leq t_{i+1} \). Then, the covariance function between any time points \( t \) and \( t' \) over the locations (denoted as \( \Sigma(t, t') \)) is given as

\[
\Sigma(t, t') = \int P(x|t) P(x|t') \, dx = \frac{1}{N^2} \sum_{\omega, \omega'} \frac{1}{2\pi\sigma_{\text{cov}}^2} \exp\left(-\frac{|\mu_{\text{cov}}(\omega, \omega')|^2}{2\sigma_{\text{cov}}^2}\right) \tag{8}
\]

where \( \mu_{\text{cov}} \) and \( \sigma_{\text{cov}}^2 \) are given by

\[
\mu_{\text{cov}}(\omega, \omega') = \frac{(t - t_i') x_{t_{i+1}} + (t_{i+1} - t') x_{t_i}}{t_{i+1} - t_i},
\]

\[
- \frac{(t' - t_i')(t_{i+1} - t) - (t_i' - t')(t_{i+1} - t_i)}{t_{i+1} - t_i - t_{i+1}}, \tag{9}
\]

and

\[
\sigma_{\text{cov}}^2(\omega, \omega') = \frac{(t - t_i')(t_{i+1} - t) + (t_i' - t')(t_{i+1} - t_i)}{\lambda(t_{i+1} - t_i)}. \tag{10}
\]

\( \Sigma(t, t') \) plays an important role in the extraction of patterns from the distribution (see Sect.2.3).

2.2 Estimation of Scale Parameter

The scale parameter \( \lambda \) is the only model parameter and it is quite important to estimate it robustly and accurately. Though it is originally given as a physical constant in quantum mechanics, it needs to be estimated statistically from a given dataset in the proposed approach. \( P_{\text{diff}} \) in Eq. (1) is used as the basic probabilistic model. In a check-in sequence \((x_1, x_2, \ldots, x_n)\), each pair \((x_t, x_{t+1})\) is used as a sample from a given dataset. As \( P_{\text{diff}} \) depends on only the differences \( \delta_t = x_{t+1} - x_t \) and \( \delta_t = t_{i+1} - t_i \), \( P_{\text{diff}}(y|xt, u) \) can be rewritten as
\[ P_{\text{diff}}(\bar{x}|\bar{x}, \lambda) = \frac{\lambda}{2\pi t} \exp\left(-\frac{\lambda|\bar{x}|^2}{2t}\right). \] (11)

In order to simplify the form of Eq. (11), \( \bar{x} \) is replaced with \( \sqrt{D} \cos \theta, \sqrt{D} \sin \theta \) in a polar coordinate system where \( D \) is the square of the Euclidean norm of \( \bar{x} \) (namely, the square Euclidean distance between \( x_{i+1} \) and \( x_i \)) and \( \theta \) is an angle \((0 \leq \theta < 2\pi)\). Then, the distribution of \( D \) and \( \theta \) is given as

\[ P(D, \theta|\bar{x}, \lambda) = \frac{\lambda}{2\pi} \exp\left(-\frac{\lambda D}{2}\right)\left|J\right| \]

where \( \left|J\right| \) is the determinant of the Jacobian matrix of the coordinate transformation, which is given as

\[ \left|J\right| = \left|\begin{array}{cc} \frac{\partial \sqrt{D} \cos \theta}{\partial \bar{x}} & \frac{\partial \sqrt{D} \sin \theta}{\partial \bar{x}} \\ \frac{\partial \sqrt{D} \sin \theta}{\partial \bar{x}} & \frac{\partial \sqrt{D} \cos \theta}{\partial \bar{x}} \end{array}\right| = \frac{1}{2} \] (12)

As \( P(D, \theta|\bar{x}, \lambda) \) does not depend on \( \theta \), \( P(D|\bar{x}, \lambda) \) is given by

\[ P(D|\bar{x}, \lambda) = \int P(D, \theta|\bar{x}, \lambda) d\theta = \frac{\lambda}{2\pi} \exp\left(-\frac{\lambda D}{2}\right). \] (13)

Consequently, the pair of the square of the Euclidean distance and the time interval between two sequential check-in’s (denoted as \( (D, \bar{t}) \)) replaces \( (x_i t_i, x_{i+1} t_{i+1}) \). Then, the set of samples consists of the pairs \( (D_k, \bar{t}_k) \) where \( k \) is the index of samples.

Now, the maximum likelihood estimation (MLE) of \( \lambda \) is given from the set of samples \( (D_k, \bar{t}_k) \) by maximizing \( \sum \log P(D_k, \bar{t}_k) \) with respect to \( \lambda \) where \( P(D_k, \bar{t}_k|\lambda) = P(D|\bar{t}, \lambda) P(\bar{t}) \). If \( P(\bar{t}) \) does not depend on \( \lambda \), the simple MLE estimator \( \hat{\lambda}_{\text{MLE}} \) is given as

\[ \hat{\lambda}_{\text{MLE}} = \arg\max_{\lambda} \sum_k \log P(D_k, \bar{t}_k|\lambda) \]

\[ = \arg\max_{\lambda} \sum_k \log P(D_k|\bar{t}_k, \lambda) = \frac{K}{\sum_k (D_k/2\bar{t}_k)} \] (14)

where \( K \) is the sample size. However, the simple estimation is not appropriate in the actual datasets (see Sect. 3.2). Since the term \( D_k/2\bar{t}_k \) diverges to the infinity if \( \bar{t}_k \) is near to 0, the estimation often lacks both accuracy and robustness.

In order to achieve a more accurate and robust estimation, the Bayesian approach is employed. First, an appropriate prior model for \( P(\bar{t}) \) is assumed and estimated by the samples \( (\bar{t}_k) \). Next, the marginalized distribution \( P(D|\lambda) = \int P(D|\bar{t}, \lambda) P(\bar{t}) d\bar{t} \) is calculated and the MLE of \( \lambda \) is estimated by \( (D_k) \) under \( P(D|\lambda) \). By observing the actual datasets, the following characteristics of the time interval \( \bar{t} \) were discovered:

1. \( \bar{t} > 0 \). In other words, the time interval is always positive and there is no conflict of multiple check-in’s of a user at a time point.
2. There is a peak near to 0 in the distribution of \( \bar{t} \). It means that the most of check-in’s occur immediately after the previous check-in’s.
3. The distribution of \( \bar{t} \) is quite long-tailed. In other words, some check-in’s occur after a quite long period.

Thus, the inverse-gamma distribution is employed as \( P(\bar{t}) \) in this paper. The inverse-gamma distribution satisfies all the above three characteristics and it gives an analytical form of the marginalized distribution \( P(D|\lambda) \). By using the inverse \( \tau = 1/\bar{t} \), the distribution of \( \tau \) is given as the following gamma distribution:

\[ P(\tau|\alpha, \beta) = \frac{\beta^\tau}{\Gamma(\alpha)} \tau^{\alpha-1} \exp(-\beta \tau) \] (15)

where \( \Gamma(\alpha) \) is the gamma function defined as \( \int_0^\infty x^{\alpha-1} \exp(-x) dx \). The MLE of \( (\alpha, \beta) \) is given as

\[ (\alpha, \beta) = \arg\max_{\alpha, \beta} \sum_k \log P(\tau_k|\alpha, \beta) \] (16)

where \( \tau_k = 1/\bar{t}_k \). Though it has no analytical solution, the optimum can be calculated easily, for example, by a MATLAB function “gamfit.” The optimal estimators are denoted as \( \hat{\alpha} \) and \( \hat{\beta} \). Thus, the marginalized distribution \( P(D|\lambda) \) is given as

\[ P(D|\lambda) = \int_0^\infty P(D|\tau, \lambda) P(\tau|\hat{\alpha}, \hat{\beta}) d\tau \]

\[ = \int_0^\infty \frac{\lambda \tau}{2} \exp\left(-\frac{\lambda D}{2}\right) \frac{\hat{\beta}\tau^{\hat{\alpha}-1} \exp(-\beta \tau)}{\Gamma(\hat{\alpha})} d\tau \]

\[ = \left(\frac{\lambda \hat{\alpha}}{2\hat{\beta}}\right) \frac{\lambda D}{2\hat{\beta} + 1}^{(\hat{\alpha}-1)}. \] (17)

This type of probability density function is known as the Lomax distribution [9]. The MLE of \( \lambda \) under the prior model is given as

\[ \hat{\lambda}_{\text{prior}} = \arg\max_{\lambda} \sum_k \log P(D_k|\lambda). \] (18)

The derivative of Eq. (18) w.r.t. \( \lambda \) is given as

\[ \frac{d}{d\lambda} \sum_k \log P(D_k|\lambda) = \sum_k \left(1 - \frac{1}{\hat{\alpha}} + \frac{D_k}{\lambda D_k + 2\hat{\beta}}\right) \] (19)

By letting the derivative be 0, \( \hat{\lambda}_{\text{prior}} \) has to satisfy the following equation:

\[ \frac{1}{K} \sum_k D_k / (2\hat{\beta} + \hat{\lambda}_{\text{prior}}) = 1 / \hat{\alpha} + 1. \] (20)

Note that both \( \hat{\alpha} \) and \( \hat{\beta} \) are positive. Then, it is easily shown that the left side of Eq. (20) is a monotonically increasing function in the range \((0, 1)\) with respect to \( \hat{\lambda}_{\text{prior}} \) and the right one of Eq. (20) is a constant depending on \( \hat{\alpha} \) in the range \((0, 1)\). Therefore, there exists a unique solution \( \hat{\lambda}_{\text{prior}} \) satisfying Eq. (20), which can be obtained easily, for example, by a MATLAB function “fzero.” The experimental results on the estimation of the scale parameter are shown in Sect. 3.2.
2.3 Extraction of Patterns by PCA

Here, we describe the PCA-based extraction of the spatio-temporal patterns from the estimated distribution. The estimated distribution from a location-based SNS dataset is given as $P(x|t)$ in Eq. (7). The infinite-dimensional singular value decomposition (SVD) is employed in order to extract spatio-temporal patterns. SVD of $P(x|t)$ is given as follows:

$$P(x|t) \approx \sum_{m=1}^{M} \nu_m U_m(x) V_m(t)$$ (21)

where $\nu_m$ is a constant depending on the $m$-th functions. The $m$-th spatial pattern $U_m(x)$ and the $m$-th temporal pattern $V_m(t)$ are functions depending on only $x$ and only $t$, respectively. In addition, $\int V_m(t) V_n(t) \, dt$ is constrained to 1 for $m = n$ or 0 otherwise ($U_m(x)$ is done in the same way). In other words, $V_m(t)$ and $U_m(x)$ satisfy the orthonormality condition. $M$ is the number of the principal functions and is set to a small number in advance. Then, it is easily shown that $V_m(t)$ is the $m$-th eigenfunction of the covariance function among the time points given as $\Sigma(t,t')$ in Eq. (8). In other words, $V_m(t)$ satisfies the following equation:

$$\nu_m^2 V_m(t) = \int \Sigma(t,t') V_m(t') \, dt'$$ (22)

where $\nu_m^2$ corresponds to the $m$-th eigenvalue. Unfortunately, there is no analytical solution of $V_m(t)$ in this equation. Therefore, the finite-dimensional matrix decomposition is utilized by discretizing $t$. By letting $t_1, \ldots, t_L$ be a sequence of discretized time points, an $L \times L$ matrix $\Sigma$ is defined whose $(p,q)$-th element is given as $\Sigma(t_p,t_q)$. Then, $V_m(t)$ and $\nu_m^2$ are approximated as the $m$-th eigenvector and the $m$-th eigenvalue of $\Sigma$, respectively. The matrix consisting of the $M$ eigenvectors is denoted as $V = (v_{np})$. In addition, the dominance ratio of the $m$-th eigenvector $\rho_m$ is calculated by $\rho_m = \nu_m^2 / \sum_{m=1}^{L} \nu_m^2$. Regarding $U_m(x)$, the following relation holds because of the orthonormality condition:

$$\int P(x|t) V_m(t) \, dt = \nu_m U_m(x).$$ (23)

Therefore, $U_m(x)$ is approximated by

$$U_m(x) = \sum_{p} P(x|t_p)v_{np}$$ (24)

where the normalization factor is omitted. This discretized estimation is almost equivalent to the usual PCA. Note that $V$ and $U_m(x)$ can be estimated only from $\hat{\Sigma}$ and $P(x|t_p)$ for $p = 1, \ldots, L$.

3. Results

3.1 Data Collection and Implementation

The sequences of check-in’s in Foursquare were collected by the following method. Foursquare provides an API\(^1\) to get the check-in information. However, the number of API accesses per day is limited and not enough for our analysis. Therefore, the check-in information from Twitter\(^2\) was collected in a similar way as used in [10]. Foursquare enables the users to share their check-in’s if they choose to share them. Each shared check-in is available as a tweet for anyone if its twitter feed is public. Each tweet contains the unique URL from which anyone can get the detailed check-in information such as time, location (name, latitude, and longitude), and tags. Only the Japanese area is focused on in this paper. Rigorously, the longitude does not correspond to the horizontal coordinate in the sphere. However, the pair of the latitude and longitude was used as the two-dimensional coordinates because the Japanese area is far from the Pole. Regarding the implementation, the diffusion-type estimation was implemented in C++ with multi-threading on 60 cores. Equation (3) diverges to the infinity at $t = t_i$ or $t = t_{i+1}$ because of $\sigma^2_{\text{dist}} \to 0$. In order to avoid this divergence, a small lower bound (set to 8.64 seconds in this paper) on the $\sigma^2_{\text{dist}}$ is used. If $\sigma^2_{\text{dist}}$ is smaller than the bound, the bound replaces it. The spatial pattern on the map is displayed by Google Maps API.

3.2 Experiments on Estimation of Scale Parameter

Here, the scale parameters $\lambda_{\text{MLE}}$ and $\lambda_{\text{prior}}$ (described in Sect. 2.2) are compared. In order to avoid the interval $\bar{\lambda}_k = 0$, such intervals were replaced by the minimum of the positive $\bar{\lambda}_k > 0$ over all the samples. By using all the check-in’s in the Foursquare dataset in Sect. 3.1, $\lambda_{\text{MLE}}$ and $\lambda_{\text{prior}}$ were estimated as 0.0184 and 460, respectively. $\sqrt{\lambda_{\text{MLE}}}$ approximately gives the average speed of the users. By setting the unit to the kilometers per hour, $\lambda_{\text{MLE}}$ and $\lambda_{\text{prior}}$ roughly corresponds to 150 kph and 1 kph, respectively. $\lambda_{\text{MLE}}$ gave a too high average velocity. On the other hand, $\lambda_{\text{prior}}$ gave a plausible estimation of the average velocity of users. In order to investigate the robustness, $\hat{\lambda}_{\text{MLE}}$ and $\hat{\lambda}_{\text{prior}}$ were estimated over 100 runs. In each run, only 10000 samples were selected randomly from all the samples. The mean (with the standard deviation) was given as 0.0448 (0.0465) for $\lambda_{\text{MLE}}$ or 454 (25.7) for $\lambda_{\text{prior}}$. The mean of $\hat{\lambda}_{\text{MLE}}$ is quite different from $\lambda_{\text{MLE}} = 0.0184$ over all the samples. Moreover, the standard deviation is comparable to the mean. On the other hand, the mean of $\hat{\lambda}_{\text{prior}}$ is quite similar to the optimal $\lambda_{\text{prior}} = 454$ over all the samples. Moreover, the standard deviation is relatively small. In addition to the above investigation, Fig. 1 shows the estimation of the log-likelihood of $P(x|t)$ for the various scale parameter settings by the five-fold cross vali-\footnotetext{\(^1\)https://developer.foursquare.com/\(^2\)https://twitter.com/}
3.3 Extraction of Patterns from Foursquare

Here, all the check-in’s from the Foursquare dataset (described in Sect. 3.1) were utilized for extracting the spatio-temporal patterns by PCA. The scale parameter $\lambda$ was set to $\hat{\lambda}_{\text{prior}} = 460$ (see Sect. 3.2). The time is discretized by setting the constant time interval to 1 hour (the number of the time points $L$ was 721). Then, the $721 \times 721$ matrix $\tilde{\Sigma}$ was calculated by Eq. (8) and $V$ was given as the $M$ principal components of $\tilde{\Sigma}$ by the usual PCA. The number of principal components $M$ was set to 6. The dominance ratio of the $m$-th principal components is given by $\rho_m$. The total dominance ratio of the six components was 90.3%.

Figure 2 shows the six extracted spatial patterns (SP1–SP6) ($U_m(x)$). Though the original dataset is distributed over Japan, only the central area of Japan including the three largest cities (Tokyo, Osaka, and Nagoya) is shown. Figure 3 shows the corresponding temporal patterns (TP1–TP6) of $V$. The temporal patterns are displayed in the time and frequency domains. Table 1 shows the list of salient days in each extracted temporal pattern, which are statistically verified by the two-sample t-test. The extracted temporal patterns were roughly classified into the three types: the regular patterns with no salient day (TP1 and TP3), and the irregular ones with several salient days caused by some transient events (TP2, TP5, and TP6), and the long-term pe-
Fig. 3  Extracted temporal patterns (TP1–TP6) in the time and frequency domains: Each temporal pattern from TP1 to TP6 corresponds to the spatial pattern from SP1 to SP6 in Fig. 2. The upper and lower correspond to the time and frequency domains, respectively.

Fig. 4  Extracted spatial patterns in a local area (LSP1–LSP3) with the dominance ratios: The map is limited to the central area of Tokyo including Akihabara Station at the top center and Tokyo International Exhibition Center "Kokusaitenjijo" at the bottom right (which is the venue for Comiket). The rightmost (d) is the enlargement of (b) around Tokyo International Exhibition Center.

Table 1  Salient days with strong significance in temporal patterns TP1–TP6: Each temporal pattern shown in Fig. 3 was divided into one day patterns (24 hourly time points in 0:00–23:00) and the two-sample t-test was applied to comparing the pattern of each day with those of the other days. The significance level was set to 0.0001. The right tailed test was employed, which means that only the “positive” salient days are tested. In addition, the equal variance was assumed except for TP4 (assuming the unequal variance as well).

<table>
<thead>
<tr>
<th>Pattern</th>
<th>Significance Days</th>
</tr>
</thead>
<tbody>
<tr>
<td>TP2</td>
<td></td>
</tr>
<tr>
<td>TP3</td>
<td></td>
</tr>
<tr>
<td>TP4 (equal)</td>
<td>Jan. 1–5, 12.</td>
</tr>
<tr>
<td>TP5</td>
<td>Dec. 28–30.</td>
</tr>
<tr>
<td>TP6</td>
<td>Dec. 31, Jan. 1–3.</td>
</tr>
</tbody>
</table>

The first regular spatio-temporal pattern (SP1 in Fig. 2 (a) and TP1 in Fig. 3 (a)), the spatial pattern is distributed around the three largest cities of Japan and the temporal pattern is stationary. Thus, this pattern can be regarded as the stationary distribution of users in Japan. Regarding the second regular pattern (SP3 in Fig. 2 (c) and TP3 in Fig. 3 (c)), the dominant cycle length is one day. In addition, a dispersion on the centers of the largest cities is observed in the spatial pattern. Thus, this pattern corresponds to the concentration and dispersion of the population on the hearts of urban areas in each day. Regarding the irregular spatio-temporal patterns (the pairs (SP2, TP2), (SP5, TP5), and (SP6, TP6)), there were a few peaks in their temporal patterns. First, (SP2, TP2) suggests that the users tend to stay at home in the end of the year, where the population at Tokyo is clearly dispersed in SP2. Second, (SP5, TP5) shows that there was a sudden crowd of the users at the area on Dec. 28, 29, and 30 and there is a peak at the east part of the center in Tokyo. By inspecting closely the comments of the users at the peak, it was suggested that this pattern is strongly related to Comiket at the east of Tokyo, which is one of the largest festivals of comics in the world. Though the actual dates for Comiket are Dec. 29–31, the observed peaks preceded by one day. The further investigation about Comiket in the small local area is given in Sect. 3.4 (see Figs. 4 (b) and (d)). Third, (SP6, TP6) seems to correspond to many users returning to their hometown in the New Year’s holidays, where they are spread widely in the map. Regarding the long-term peri-
In Sect. 3.3, we used only the check-in's within latitudes in a small local area. Though the total dataset is the same to verify the usefulness of the proposed method even in the small area. Here, additional experimental results are given in order to verify the effectiveness of the proposed method. Consequently, the scale parameter in the diffusion-type formula is estimated appropriately and robustly by a Bayesian approach with a prior model. Moreover, the proposed method could extract some plausible and interesting patterns from the actual dataset. We are now planning to analyze many other Foursquare datasets by the proposed method and compare it with other methods. Moreover, we are planning to extend the proposed method by selecting the dataset based on the other factors such as the area, the users, and so on. Such selection is promising for discovering interesting patterns. For example, the clustering of users is expected to extract specific patterns of a group of users. In addition, we are planning to improve the setting process, especially for the number of the principal components $M$ and to employ the other extraction methods such as independent component analysis. Furthermore, we are planning to incorporate the location effects (such as roads, rails, rivers, and so on) into the proposed probabilistic model which currently employs the simple free-particle approximation.

### Acknowledgments

This work is partially supported by Grant-in-Aid for Young Scientists (KAKENHI) 26730013.
References


Yoshitatsu Matsuda received his Ph.D. from the University of Tokyo, Japan in 2002. He is currently a research fellow in the University of Tokyo. His research interests include independent component analysis, massive data analysis, and self-organizing neural networks.

Kazunori Yamaguchi received the B.S., M.S. and Doctor of Science degrees in information science from the University of Tokyo, in 1979, 1981, and 1985, respectively. Currently, he is a professor of the University of Tokyo. His research interests are in data models and data analysis.

Ken-ichiro Nishioka received the B.S., M.S. of Arts and Science degree from the University of Tokyo, in 2009 and 2011, respectively. He is currently a Ph.D. candidate in the University of Tokyo. His research interest is recommender system.