Discovering Spatio-Temporal Latent Influence in Geographical Attention Dynamics

Minoru Higuchi[1], Kanji Matsutani[2], Masahito Kumano[1], and Masahiro Kimura[1]

[1] Department of Electronics and Informatics, Ryukoku University, Otsu, Japan
kimura@rins.ryukoku.ac.jp
[2] Tokai Regional Headquarter, NTT West Corporation, Nagoya, Japan

Abstract. We address the problem of modeling the occurrence process of events for visiting attractive places, called points-of-interest (POIs), in a sightseeing city in the setting of a continuous time-axis and a continuous spatial domain, which is referred to as modeling geographical attention dynamics. By combining a Hawkes process with a time-varying Gaussian mixture model in a novel way and incorporating the influence structure depending on time slots as well, we propose a probabilistic model for discovering the spatio-temporal influence structure among major sightseeing areas from the viewpoint of geographical attention dynamics, and aim to accurately predict POI visit events in the near future. We develop an efficient method of inferring the parameters in the proposed model from the observed sequence of POI visit events, and present an analysis method for the geographical attention dynamics. Using real data of POI visit events in a Japanese sightseeing city, we demonstrate that the proposed model outperforms conventional models in terms of predictive accuracy, and uncover the spatio-temporal influence structure among major sightseeing areas in the city from the perspective of geographical attention dynamics.

Keywords: Geographical attention dynamics · Point process model · Spatio-temporal influence structure.

1 Introduction

With the development of smart mobile devices, location acquisition technologies and social media, a large amount of event data with spatio-temporal information has become available and offers an opportunity to better understand people's location preferences and mobility patterns in a sightseeing city [3]. In location-based social networking services (LBSNs) such as Foursquare and Facebook Places, check-in sequences of users to *points-of-interest (POIs)* are observed, where a finite number of venues are listed as POIs in advance. Clearly, there exist infinitely many attractive places in the city, including various geographical points giving beautiful views and street spots with artistic atmosphere. In photo-sharing services such as Flickr, observations of where and when people took photos are obtained. In order to take into account infinitely many attractive places on a continuous spatial domain, we consider extending the definition of POI. In particular, we also refer to the geographical locations of such photos that were taken on sightseeing tours and uploaded to a photo-sharing site as POIs, and aim at precisely

investigating people's experiences in visiting attractive places in the city. Namely, in our definition, it is supposed that any POI offers an attractive place in a sense. Note that a complete list of all POIs cannot be obtained in advance.

Recently, researchers [2, 8, 21] have examined the next POI recommendation problem, that is, the problem of predicting which POI a user is most likely to visit at the next discrete time-step given the current check-in POI, where it is assumed that a finite set of POIs is specified in advance and historical check-in sequences of users in an LBSN are provided. However, these studies were unable to fully capture the continuous structure of space-time, and thus Liu et al [14] extended them to the case of a continuous time-axis by integrating temporal interval assessment. On the other hand, since online items posted on social media sites such as Facebook and Twitter gain their popularity by the amount of attention received (e.g., the number of Facebook shares and the number of retweets), several studies have been made on modeling the attention dynamics of online items in a continuous time-axis [19, 16, 11, 22]. Zhou et al [24] presented a point process model in a discretized time-axis and a continuous spatial domain by fusing a time-varying Gaussian mixture model with a non-homogeneous Poisson process, and successfully estimated the spatial distribution of Toronto's ambulance demand at a specified discrete time-step (i.e., each two-hour interval), where each Gaussian component shows a representative geographical area. In the case of dealing with events of visiting POIs in a sightseeing city, such a component may correspond to a major sightseeing area. However, this study is unable to extract the influence structure among components from the viewpoint of visiting POIs, while such knowledge can become important for tourism marketing.

For a given sightseeing city, we consider the problem of modeling the occurrence process of events for visiting POIs in a continuous time-axis and a continuous spatial domain, which is referred to as that of modeling the *geographical attention dynamics*, and aim to provide deep insights into the properties of people's location preferences and mobility patterns on sightseeing tours in the city. What we observe is both a time-sequence of events (see Fig. 1a) and their locations (see Fig. 1b). Given a season, the sightseeing city should have a finite number of major sightseeing areas $C_1, \ldots, C_K$ (see Fig. 1c), where these represent major tourism topics, and are allowed to geographically intersect each other. In the same way as the attention dynamics of online items in social media, we first assume that the occurrences of previous events increase the possibility of future events. In particular, POI-visit events should exhibit a geographically self-exciting nature, where an event that happened in an area C_k may cause its subsequent events in the same area C_k. Also, it is natural to suppose that the geographical attention dynamics has a geographically mutually-exciting nature, where an event that happened in an area C_k can trigger the subsequent events in any other area C_ℓ (see Fig. 1d). Moreover, the temporal decay rate of such effect should vary according to area C_k. Thus, based on the data of many people's POI visit events in the season, it is desirable to identify major sightseeing areas $C_1, \ldots, C_K$ and find the spatio-temporal influence relations among $C_1, \ldots, C_K$ in terms of geographical mutual-excitation (see Fig. 1e).

In this paper, we propose a probabilistic model for discovering the spatio-temporal influence structure among major sightseeing areas from the perspective of geographical attention dynamics in a continuous space-time, and aim at accurately predicting the

(a) A time sequence of events. (b) Geographical locations of events. (c) Major sightseeing areas C_1, C_2 and C_3.

(d) Geographical self-excitations and mutual-excitations. (e) Main spatio-temporal influence relations among C_1, C_2 and C_3.

Fig. 1: An illustration of geographical attention dynamics. Down-pointing triangles indicate the time points or geographical locations for POI visit events. Arrows in (d) indicates triggering relations between events. For example, (d) illustrates that the event in C_1 at time $t_{1,2}$ triggered the event in C_2 at time $t_{2,1}$. Arrows in (e) represent the main influence relations among latent components (major sightseeing areas).

future POI visit events. To this end, we combine a Hawkes process [13], which is a counting process [1] frequently utilized to capture mutual excitations between events, with a time-varying Gaussian mixture model in a novel way. Also, we incorporate the influence structure depending on time slots into our model since it is known that users' activities in LBSNs are often influenced by time [10, 20] and such temporal properties may rely on sightseeing cities as well as seasons. We develop an efficient method of inferring the parameters in the proposed model from the observed sequence of POI visit events, and provide an analysis method for the geographical attention dynamics in terms of spatio-temporal influence relations among major sightseeing areas. Using real data of POI visit events in Japanese sightseeing city "Kyoto" obtained from a photo-sharing site, we evaluate the proposed method. First, for predicting the future POI visit events, we show the effectiveness of the proposed model compared to a baseline model and such a conventional spatio-temporal point process model that simply integrates a Hawkes process with a time-varying Gaussian mixture model (see [24]), which is hereafter referred to as *HG model*. Next, by applying the proposed model, we uncover the spatio-temporal influence relation structure among major sightseeing areas in Kyoto in view of geographical attention dynamics.

The rest of the paper is organized as follows: In Section 2, we briefly summarize the related work. In Section 3, we introduce the HG model, and present the proposed model as its extension. In Section 4, we develop a probabilistic inference method for the proposed model, and present an analysis method for geographical attention dynamics.

In Section 5, we report the experimental results. We conclude the paper by summarizing the main results in Section 6.

2   Related Work

Several studies have been made on predicting POI visit events in the near future. As described in Section 1, Chen et al [2], Feng et al [8] and Zhang et al [21] investigated the next POI recommendation problem for a finite number of given POIs in a discretized time-axis. To address the problem in a continuous time-axis, Liu et al [14] presented a method of exploiting temporal interval assessment. To construct an accurate predictive model of event data with mark information such as POI in a continuous time-axis for a finite number of given marks, Du et al [5] extended a marked Hawkes process, and proposed such a marked temporal point process that incorporates a recurrent neural network. However, unlike our current approach, these works have a limitation in treating a continuous spatial domain, that is, it is difficult to handle the situation where there may be infinitely many POIs and a complete list of all POIs is unavailable. To model ambulance demand in a discretized time-axis and a continuous spatial domain, Zhou et al [24] integrated a time-varying Gaussian mixture model with a non-homogeneous Poisson process. Note that this model can simply be extended to the HG model, which is a model for a continuous space-time. However, as already mentioned, the HG model also has a limitation in analyzing the spatio-temporal influence structure among latent components. In this paper, by properly extending the HG model, we propose a probabilistic model for the geographical attention dynamics in a continuous space-time, and aim at discovering the spatio-temporal influence structure among major sightseeing areas in a given season.

There have been many investigations related to modeling continuous-time events generated by users in social media. As described in Section 1, Wang et al [19], Shen et al [16], Gao et al [11] and Zhao et al [22] considered individually modeling the attention dynamics of online items posted on social media sites in order to predict their future popularity. Thus, unlike our current approach, these works are unable to properly analyze the relations among all the online items involved. On the other hand, Gomez-Rodriguez et al [12] and Daneshmand et al [4] examined the problem of extracting the social influence network structure among users from the observed information cascades (i.e., the observed sequences of events for sharing the same online items). Multivariate Hawkes processes are often leveraged to model event sequences forming information cascades in social networks [23, 6]. They are also extended to model coevolution dynamics of information diffusion and social network growth [7]. However, these studies exploited who shared which online item, and assume that a finite set of online items and users is given in advance. In this paper, we also focus on a Hawkes process, but unlike the studies mentioned above, we try to infer the spatio-temporal influence relation structure among POIs and predict the future POI visit events without knowing who visited which POI from the viewpoint of privacy protection. We also note that multivariate Hawkes processes cannot simply be applied to our problem since a complete list of all POIs and users is unavailable in advance.

3 Model

For predictive modeling of geographical attention dynamics, we consider modeling the occurrence process of events for visiting POIs in a sightseeing city during a time period $[0, T')$ corresponding to one of its tourist seasons in the setting of a continuous space-time, where T' (> 0) is assumed to be a few months, and the corresponding continuous spatial domain is denoted by $\Omega \subset \mathbb{R}^2$.

3.1 Preliminaries

For any $t \in (0, T')$, let N_t be the total number of events during time period $[0, t)$, and for each $n = 1, \dots, N_t$, we represent the nth event as tuple $(t_n, \boldsymbol{x}_n)$, meaning that location $\boldsymbol{x}_n = (x_{n,1}, x_{n,2}) \in \Omega$ was visited and registered as a POI at time $t_n \in [0, T')$ on a sightseeing tour. We also denote the sequence of events (i.e., the history) up to but not including time t as

$$\mathcal{H}_t = \{(t_n, \boldsymbol{x}_n); n = 1, \dots, N_t\}.$$

Based on the previous work [24], we focus on modeling the event occurrence process as a spatio-temporal point process with intensity function $\lambda(t) f(\boldsymbol{x}|t)$ for $\forall t \in (0, T')$ and $\forall \boldsymbol{x} = (x_1, x_2) \in \Omega$, where $\lambda(t)$ is the intensity function of a temporal point process and $f(\boldsymbol{x}|t)$ is a time-varying Gaussian mixture for the spatial distribution. Namely, $\lambda(t) f(\boldsymbol{x}|t) \, dt \, d\boldsymbol{x}$ is the conditional probability of observing an event within a small domain $[t, t + dt) \times \{[x_1, x_1 + dx_1) \times [x_2, x_2 + dx_2)\}$ given the history $\mathcal{H}_t$ (see [1, 5, 7]). Note that for $0 < \forall T \leq T'$, the probability density of $\mathcal{H}_T$ is given by

$$p(\mathcal{H}_T) = \exp\left\{-\int_0^T \lambda(t) \, dt\right\} \prod_{n=1}^{N_T} \{\lambda(t_n) f(\boldsymbol{x}_n | t_n)\}. \tag{1}$$

3.2 Spatial Distribution

We begin with defining $f(\boldsymbol{x}|t)$ for any $(t, \boldsymbol{x}) \in (0, T') \times \Omega$. Since people's travel behaviors should vary by time of day, several studies [10, 20] separated each day into different time slots to improve POI recommendations in LBSNs. We also adopt this idea, and decompose each day into such M time slots $TS_1, \dots, TS_M$ that are appropriate for the city and the season to be considered[1]. Let $h : [0, T') \to \{1, \dots, M\}$ be the *time slot function*, meaning that each time $t \in [0, T')$ belongs to time slot $TS_{h(t)}$.

Previous work [24] fixed the mixture component distributions across all time slots to overcome data sparsity issues, and tried to capture an accurate spatial structure. In the same way as [24], we define the spatial distribution $f(\boldsymbol{x}|t)$ by

$$f(\boldsymbol{x}|t, \Theta) = \sum_{k=1}^K \phi_{h(t),k} \, g(\boldsymbol{x}|\boldsymbol{\mu}_k, \Sigma_k), \quad \forall \boldsymbol{x} \in \Omega, \tag{2}$$

---

[1] Although it is desirable to automatically detect such a decomposition from data, we here assume that $TS_1, \dots, TS_M$ are specified in advance. Our future work will involve developing this kind of method.

where K is the number of components, and $g(\boldsymbol{x} \,|\, \boldsymbol{\mu}_k, \Sigma_k)$ is the 2-dimensional Gaussian density with mean vector $\boldsymbol{\mu}_k$ and covariance matrix Σ_k for $k = 1, \ldots, K$. The mixing coefficients $\{\phi_{m,k}\}$ for time slot TS_m satisfy $0 < \phi_{m,k} < 1$ together with $\sum_{k=1}^{K} \phi_{m,k} = 1$ for $m = 1, \ldots, M$. Also, the parameters for $f(\boldsymbol{x} \,|\, t)$ are aggregated into the parameter set $\Theta = \{\phi_{m,k}, \boldsymbol{\mu}_k, \Sigma_k; m = 1, \ldots, M, k = 1, \ldots, K\}$. We can consider that each Gaussian component C_k essentially represents a geographical area corresponding to a major tourism topic, and is identified with a major sightseeing area. Thus, we leverage this identification and try to analyze the influence relations among those major sightseeing areas.

3.3 Spatio-Temporal Point Process

Next, we consider modeling $\lambda(t)$ for any $t \in (0, T')$.

Baseline Model One of the simplest models for a temporal point process is a Poisson process, where $\lambda(t)$ is assumed to be independent of history $\mathcal{H}_t$ and given by

$$\lambda(t \,|\, \alpha) = \alpha, \quad \forall t \in (0, T'). \tag{3}$$

Here, α is a positive constant. Thus, the spatio-temporal point process model defined by intensity function $\lambda(t \,|\, \alpha) f(\boldsymbol{x} \,|\, t, \Theta)$ (see Eqs. (2) and (3)) is regarded as a baseline.

Conventional Model As described in the previous sections, a Hawkes process is frequently used to model continuous-time events and capture mutually-exciting interactions between events, and has been investigated for various applications (see [13, 22]). Thus, $\lambda(t)$ can be modeled as a Hawkes process,

$$\lambda(t \,|\, \alpha, \beta, \gamma) = \alpha + \beta \sum_{(t_n, \boldsymbol{X}_n) \in \mathcal{H}_t} \exp\{-\gamma (t - t_n)\}, \quad \forall t \in (0, T'). \tag{4}$$

where α, β and γ are positive constants. Here, the spatio-temporal point process defined by intensity function $\lambda(t \,|\, \alpha, \beta, \gamma) f(\boldsymbol{x} \,|\, t, \Theta)$ (see Eqs. (2) and (4)) is referred to as *HG model*. Note that the HG model can be regarded as a conventional model presented in the previous work [24].

Proposed Model By incorporating both component dependent temporal influence decay (see Section 1) and time-slot varying influence degree (see Section 3.2), we extend the HG model, and aim to discover the spatio-temporal influence structure among components from the viewpoint of geographical attention dynamics and to more accurately predict POI visit events in the near future. The proposed model is defined as the spatio-temporal point process with intensity function $\lambda(t \,|\, Z_t, \alpha, \boldsymbol{\beta}, \boldsymbol{\gamma}) f(\boldsymbol{x} \,|\, t, \Theta)$ (see Eqs. (2) and (5)). Here, $\lambda(t)$ is modeled as

$$\lambda(t \,|\, Z_t, \alpha, \boldsymbol{\beta}, \boldsymbol{\gamma}) = \alpha + \sum_{(t_n, \boldsymbol{X}_n) \in \mathcal{H}_t} \beta_{h(t_n)} \exp\left\{-\gamma_{z(\boldsymbol{X}_n \,|\, t_n)} (t - t_n)\right\}, \quad \forall t \in (0, T'), \tag{5}$$

where $z(\boldsymbol{x}_n | t_n)$ denotes the component ID of location $\boldsymbol{x}_n$ drawn from Gaussian mixture $f(\boldsymbol{x} | t_n, \Theta)$ at time t_n, i.e., $z(\boldsymbol{x}_n | t_n) = k$ if and only if $\boldsymbol{x}_n \in C_k$ at time t_n, for $n = 1, \ldots, N_t$. Z_t is defined as

$$Z_t = \{z(\boldsymbol{x}_n | t_n); \ n = 1, \ldots, N_t\}.$$

Also, for the city during the current season, $\alpha > 0$ expresses its underlying attractiveness, $\beta_m > 0$ represents the influence degree of time slot TS_m for $m = 1, \ldots, M$, and $\gamma_k > 0$ indicates the temporal influence decay rate of component C_k for $k = 1, \ldots, K$. Parameters $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ are defined as $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_M)$ and $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_K)$, respectively. Here, based on the additivity for independent Poisson processes (see [13, 15, 6]), for any $t \in (0, T')$, we introduce a set of latent variables,

$$Y_t = \{y_n; \ n = 1, \ldots, N_t\},$$

such that the nth event $(t_n, \boldsymbol{x}_n)$ was triggered by the y_nth event $(t_{y_n}, \boldsymbol{x}_{y_n})$, where $y_n = 0, 1, \ldots, n-1$, and $y_n = 0$ means that the nth event was triggered by the underlying attractiveness, i.e., the background intensity α. Namely, it is known that the point process with intensity function $\lambda(t_n | Z_{t_n}, \alpha, \boldsymbol{\beta}, \boldsymbol{\gamma})$ at time t_n is the superposition of the Poisson processes with intensity functions $\lambda(t_n; y_n | Z_{t_n}, \alpha, \boldsymbol{\beta}, \boldsymbol{\gamma}), (y_n = 0, 1, \ldots, n-1)$, where

$$\lambda(t_n; y_n | Z_{t_n}, \alpha, \boldsymbol{\beta}, \boldsymbol{\gamma}) = \begin{cases} \alpha & \text{if} \quad y_n = 0 \\ \beta_{h(t_{y_n})} \exp\left\{-\gamma_{z(\boldsymbol{x}_{y_n} | t_{y_n})} (t_n - t_{y_n})\right\} & \text{if} \quad 1 \le y_n < n \end{cases} \quad (6)$$

for $n = 1, \ldots, N_t$. We consider extracting the influence relation $R_{k,\ell}$ from component C_ℓ to component C_k for $k, \ell = 1, \ldots, K$ by leveraging Z_T and Y_T.

4 Learning Method

For the observed data $\mathcal{H}_T$ with $0 < T < T'$, we develop a method of inferring the parameters Θ, Z_T, α, $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$ and Y_T in the proposed model, and provide a method for prediction and analysis of the geographical attention dynamics.

4.1 Inference

We present an inference method of the proposed model from $\mathcal{H}_T$. We consider estimating the parameters in the proposed model by maximizing the likelihood function of $\mathcal{H}_T$, $p(\mathcal{H}_T | \Theta, Z_T, \alpha, \boldsymbol{\beta}, \boldsymbol{\gamma})$ (see Eq. (1)).

First, we estimate Θ by maximizing function $\mathcal{L}(\Theta) = \prod_{(t_n, \boldsymbol{x}_n) \in \mathcal{H}_T} f(\boldsymbol{x}_n | t_n, \Theta)$. We employ an EM algorithm. Note that the number K of components is assumed to be fixed in this paper although it can also be estimated from the observed data by exploiting some techniques such as affinity propagation [9] and birth-and-death Markov chain Monte Carlo [18, 24]. Let $\bar{\Theta}$ be the current estimate of Θ. Then, the update rule "$\hat{\Theta} =$

$\left\{ \{\hat{\phi}_{m,k}\}, \{\hat{\mu}_k\}, \{\hat{\Sigma}_k\} \right\} \leftarrow \bar{\Theta} = \left\{ \{\bar{\phi}_{m,k}\}, \{\bar{\mu}_k\}, \{\bar{\Sigma}_k\} \right\}$" is obtained as follows[2]:

$$\hat{\phi}_{m,k} = \frac{1}{|\mathcal{H}_T^m|} \sum_{(t_n, \boldsymbol{x}_n) \in \mathcal{H}_T^m} \frac{\bar{\phi}_{m,k} \, g(\boldsymbol{x}_n \mid \bar{\mu}_k, \bar{\Sigma}_k)}{f(\boldsymbol{x}_n \mid t_n, \bar{\Theta})},$$

$$\hat{\mu}_k = \frac{1}{\sum_{n=1}^{N_T} \bar{a}_{n,k}} \sum_{n=1}^{N_T} \bar{a}_{n,k} \, \boldsymbol{x}_n, \quad \hat{\Sigma}_k = \frac{1}{\sum_{n=1}^{N_T} \bar{a}_{n,k}} \sum_{n=1}^{N_T} \bar{a}_{n,k} \, (\boldsymbol{x}_n - \hat{\mu}_k) \, (\boldsymbol{x}_n - \hat{\mu}_k)^{\mathrm{T}}$$

for $m = 1, \ldots, M$ and $k = 1, \ldots, K$, where the superscript T stands for a matrix transpose, each 2-vector is treated as a 2×1 matrix, and

$$\mathcal{H}_T^m = \{(t, \boldsymbol{x}) \in \mathcal{H}_T; \, h(t) = m\}, \quad \bar{a}_{n,k} = \frac{\bar{\phi}_{h(t_n),k} \, g(\boldsymbol{x}_n \mid \bar{\mu}_k, \bar{\Sigma}_k)}{f(\boldsymbol{x}_n \mid t_n, \bar{\Theta})}.$$

Also, $|S|$ denotes the number of elements in a set S. With this method, we get the estimate Θ^* of Θ. Then, for each $(t_n, \boldsymbol{x}_n) \in \mathcal{H}_T$ and $k = 1, \ldots, K$, the posterior probability $\psi_k(\boldsymbol{x}_n \mid t_n)$ of location $\boldsymbol{x}_n$ at time t_n is given by

$$\psi_k(\boldsymbol{x}_n \mid t_n) = P(z(\boldsymbol{x}_n \mid t_n) = k \mid t_n, \boldsymbol{x}_n, \Theta^*) \; = \; \frac{\phi^*_{h(t_n),k} \, g(\boldsymbol{x}_n \mid \mu_k^*, \Sigma_k^*)}{f(\boldsymbol{x}_n \mid t_n, \Theta^*)}, \tag{7}$$

and thus the estimate Z_T^* of Z_T can be obtained by

$$z^*(\boldsymbol{x}_n \mid t_n) = \operatorname*{argmax}_{1 \leq k \leq K} \psi_k(\boldsymbol{x}_n \mid t_n).$$

Next, we estimate α, $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ by maximizing $p(\mathcal{H}_T \mid \Theta^*, Z_T^*, \alpha, \boldsymbol{\beta}, \boldsymbol{\gamma})$ (see Eq. (1)). To this end, we introduce the latent variables Y_T and try to infer Y_T as well (see Section 3.3). We consider leveraging the joint likelihood $p(\mathcal{H}_T, Y_T \mid \Theta^*, Z_T^*, \alpha, \boldsymbol{\beta}, \boldsymbol{\gamma})$,

$$p(\mathcal{H}_T, Y_T \mid \Theta^*, Z_T^*, \alpha, \boldsymbol{\beta}, \boldsymbol{\gamma})$$
$$\propto \exp\left\{ -T\alpha - \sum_{m=1}^{M} \beta_m \, G_m(\boldsymbol{\gamma} \mid Z_T^*) \right\} \prod_{(t_n, \boldsymbol{x}_n) \in \mathcal{H}_T} \lambda(t_n; \, y_n \mid Z_T^*, \alpha, \boldsymbol{\beta}, \boldsymbol{\gamma}), \tag{8}$$

where

$$G_m(\boldsymbol{\gamma} \mid Z_T^*) \; = \; \sum_{n=1}^{N_T} \frac{1}{\gamma_{z(\boldsymbol{x}_n \mid t_n)}} \left(1 - \exp\left\{ -\gamma_{z(\boldsymbol{x}_n \mid t_n)} (T - t_n) \right\} \right) I(h(t_n) = m).$$

Here, $I(v)$ is an indicator function such that $I(v) = 1$ if v is true, $I(v) = 0$ otherwise. Suppose that α, $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ are independently generated from the following priors (i.e., gamma distributions):

$$\alpha \sim \mathrm{Gamma}(\nu_\alpha, \eta_\alpha), \quad \beta_m \sim \mathrm{Gamma}(\nu_\beta, \eta_\beta), \quad \gamma_k \sim \mathrm{Gamma}(\nu_\gamma, \eta_\gamma), \tag{9}$$

---

[2] For simplicity, no priors are here assumed for Θ. Note that it is clearly possible to give some natural priors.

for $m = 1, \ldots, M$ and $k = 1, \ldots, K$, where $\nu_\alpha, \eta_\alpha, \nu_\beta, \eta_\beta, \nu_\gamma, \eta_\gamma > 0$ are hyper-parameters. Then, $p(\mathcal{H}_T, Y_T \mid \Theta^*, Z_T^*, \alpha, \beta, \gamma)$ can be analytically marginalized over α and β for priors (see Eqs. (8) and (9)), and we have

$$
\begin{aligned}
&p(\mathcal{H}_T, Y_T \mid \Theta^*, Z_T^*, \gamma, \nu_\alpha, \eta_\alpha, \nu_\beta, \eta_\beta) \\
&= \int_{\mathbb{R}_+ \times \mathbb{R}_+^M} p(\mathcal{H}_T, Y_T \mid \Theta^*, Z_T^*, \alpha, \beta, \gamma)\, p(\alpha \mid \nu_\alpha, \eta_\alpha)\, p(\beta \mid \nu_\beta, \eta_\beta)\, d\alpha\, d\beta \\
&\propto \exp\left\{ -\sum_{n=1}^{N_T} \gamma_{z(\boldsymbol{x}_{y_n} \mid t_n)}(t_n - t_{y_n}) \right\} \frac{\Gamma(L_0 + \nu_\alpha)}{(T + \eta_\alpha)^{L_0 + \nu_\alpha}} \frac{\eta_\alpha^{\nu_\alpha}}{\Gamma(\nu_\alpha)} \\
&\quad \times \prod_{m=1}^{M} \left\{ \frac{\Gamma(L_m + \nu_\beta)}{\{G_m(\gamma \mid Z_T^*) + \eta_\beta\}^{L_m + \nu_\beta}} \frac{\eta_\beta^{\nu_\beta}}{\Gamma(\nu_\beta)} \right\},
\end{aligned}
\tag{10}
$$

where $\mathbb{R}_+$ denotes the space of positive real numbers, $\Gamma(s)$ is the gamma function,

$$
L_0 = \sum_{n=1}^{N_T} I(y_n = 0)
$$

indicates the number of events triggered by the background intensity, and

$$
L_m = \sum_{n=2}^{N_T} I(h(t_{y_n}) = m)\, I(y_n \geq 1)
$$

indicates the number of events triggered by the preceding events within time slot TS_m. By iterating the following three steps, we obtain the estimates α^*, β^* and γ^* of α, β and γ, respectively: 1) Gibbs sampling for Y_T. 2) Metropolis-Hastings sampling for γ. 3) Sampling for α and β, and updating of hyper-parameters. Moreover, based on the superposition theorem of independent Poisson processes (see [15, 6]), we estimate the posterior probability $\xi_{n,i} = P(y_n = i \mid \mathcal{H}_T, \Theta^*, Z_T^*, \alpha^*, \beta^*, \gamma^*)$ as

$$
\xi_{n,i} = \frac{\lambda(t_n;\, y_n = i \mid Z_T^*, \alpha^*, \beta^*, \gamma^*)}{\sum_{j=0}^{n-1} \lambda(t_n;\, y_n = j \mid Z_T^*, \alpha^*, \beta^*, \gamma^*)}
\tag{11}
$$

for $n = 1, \ldots, N_T$, $i = 0, 1, \ldots, n-1$ (see Eq. (6)). Note that $\{\xi_{n,i}\}$ provide the posterior distribution of Y_T. Below, we will describe the above three steps 1), 2) and 3) in detail.

Gibbs sampling for Y_T: Given the current samples of Y_T, a new value of y_n for $n = 1, \ldots, N_T$ is sampled from $\{0, \ldots, n-1\}$ using the Gibbs sampler of the conditional probability (see Eq. (10)),

$$
P(y_n = i \mid \mathcal{H}_T, Y_T^{-n}, \Theta^*, Z_T^*, \gamma, \nu_\alpha, \eta_\alpha, \nu_\beta, \eta_\beta) \propto p(y_n = i, \mathcal{H}_T \mid Y_T^{-n}, \Theta^*, Z_T^*, \gamma, \nu_\alpha, \eta_\alpha, \nu_\beta, \eta_\beta)
$$

$$
\propto \begin{cases} \dfrac{L_0^{-n} + \nu_\alpha}{T + \eta_\alpha} & \text{if} \quad i = 0 \\[2ex] \dfrac{L_{h(t_i)}^{-n} + \nu_\beta}{G_{h(t_i)}(\gamma \mid Z_T^*) + \eta_\beta} \exp\left\{ -\gamma_{z(\boldsymbol{x}_i \mid t_i)}(t_n - t_i) \right\} & \text{if} \quad i = 1, \ldots, n, \end{cases}
$$

where the superscript $-n$ stands for the set or value excluding the nth event.

Metropolis-Hastings sampling for γ: Due to the nonconjugacy of γ, we consider leveraging a Metropolis-Hastings algorithm to obtain the invariant distribution of γ for current samples of Y_T. Here, we exploit a normal distribution $q(\gamma' | \gamma)$ as a proposal distribution for candidate γ', Using the symmetric property, $q(\gamma' | \gamma) = q(\gamma | \gamma')$, the acceptance probability of γ' is obtained by

$$Q(\gamma' | \gamma) = \min\left\{1, \frac{p(\gamma' | \mathcal{H}_T, Y_T, \Theta^*, Z_T^*, \nu_\alpha, \eta_\alpha, \nu_\beta, \eta_\beta)}{p(\gamma | \mathcal{H}_T, Y_T, \Theta^*, Z_T^*, \nu_\alpha, \eta_\alpha, \nu_\beta, \eta_\beta)}\right\}$$

Note that $Q(\gamma' | \gamma)$ is easily computed by using the relation,

$$p(\gamma | \mathcal{H}_T, Y_T, \Theta^*, Z_T^*, \nu_\alpha, \eta_\alpha, \nu_\beta, \eta_\beta) \propto p(\mathcal{H}_T, Y_T | \gamma, \Theta^*, Z_T^*, \nu_\alpha, \eta_\alpha, \nu_\beta, \eta_\beta) \, p(\gamma | \nu_\gamma, \eta_\gamma)$$

(see Eqs. (9) and (10)). We accept γ' according to $Q(\gamma' | \gamma)$. By iterating these operations, we obtain a sample for γ.

Sampling for α and β, and updating of hyper-parameters: Given the current samples for Y_T and γ, we sample α and β by the expected values of the posterior distributions $p(\alpha | \mathcal{H}_T, Y_T, \nu_\alpha, \eta_\alpha) = \text{Gamma}(L_0 + \nu_\alpha, T + \eta_\alpha)$ and $p(\beta_m | \mathcal{H}_T, Y_T, Z_T^*, \gamma, \nu_\beta, \eta_\beta)$ $= \text{Gamma}(L_m + \nu_\beta, G_m(\gamma | Z_T^*) + \eta_\beta)$ as follows:

$$\alpha = \frac{L_0 + \nu_\alpha}{T + \eta_\alpha}, \quad \beta_m = \frac{L_m + \nu_\beta}{G_m(\gamma | Z_T^*) + \eta_\beta}$$

for $m = 1, \ldots, M$ (see Eqs. (8) and (9)). Next, we update the hyperparameters $\nu_\alpha, \eta_\alpha, \nu_\beta$, η_β, ν_γ and η_γ through the maximum likelihood estimations (see Eqs. (8),(9) and (10)). Here, the objective function for ν_α and η_α is given by

$$\mathcal{L}_\alpha(\nu_\alpha, \eta_\alpha) = \ln \frac{\Gamma(L_0 + \nu_\alpha)}{(T + \eta_\alpha)^{L_0 + \nu_\alpha}} + \ln \frac{\eta_\alpha^{\nu_\alpha}}{\Gamma(\nu_\alpha)}.$$

Also, the objective functions for ν_β and η_β is given by

$$\mathcal{L}_\beta(\nu_\beta, \eta_\beta) = \sum_{m=1}^{M} \ln \frac{\Gamma(L_m + \nu_\beta)}{\left\{G_m(\gamma | Z_T^*) + \eta_\beta\right\}^{L_m + \eta_\beta}} + M \ln \frac{\eta_\beta^{\nu_\beta}}{\Gamma(\nu_\beta)},$$

and the objective function for ν_γ and η_γ is given by

$$\mathcal{L}_\gamma(\nu_\gamma, \eta_\gamma) = \sum_{k=1}^{K} \left(\ln \gamma_k^{\nu_\gamma - 1} - \eta_\gamma \gamma_k\right) + K \ln \frac{\eta_\gamma^{\nu_\gamma}}{\Gamma(\nu_\gamma)}.$$

Based on Newton's method, we obtain the update rules for these hyper-parameters.

4.2 Prediction and Analysis

Using the proposed model inferred from $\mathcal{H}_T$, we provide a framework for predicting the future events and analyzing the geographical attention dynamics.

We predict the events occurring in $[T, T')$ by simulating the proposed model under Ogata's thinning algorithm [15] based on the intensity function given by

$$\lambda^*(t) f^*(\boldsymbol{x} \,|\, t)$$

$$= \left(\alpha^* + \sum_{(t_i, \boldsymbol{x}_i) \in \mathcal{H}_t} \beta^*_{h(t_i)} \sum_{k=1}^{K} \psi_k(\boldsymbol{x}_i \,|\, t_i) \, \exp\left\{ -\gamma^*_k (t - t_i) \right\} \right) \sum_{k=1}^{K} \phi^*_{h(t),k} \, g(\boldsymbol{x} \,|\, \boldsymbol{\mu}^*_k, \Sigma^*_k) \quad (12)$$

for any $t \in [T, T')$ (see Eqs. (2), (5) and (7)).

We analyze the geographical attention dynamics in the following way. We first examine the estimated parameters α^*, $\boldsymbol{\beta}^*$, $\boldsymbol{\gamma}^*$ and Θ^* in detail. Next, we extract the influence relation $R_{k,\ell}$ from latent component C_ℓ to latent component C_k by

$$R_{k,\ell} = \sum_{n=2}^{N_T} \sum_{i=1}^{n-1} \xi_{n,i} \, \psi_k(\boldsymbol{x}_n \,|\, t_n) \, \psi_\ell(\boldsymbol{x}_i \,|\, t_i)$$

for $k, \ell = 1, \ldots, K$ (see Eq. (11)), and analyze it. Here, note that each component C_k is identified with a major sightseeing area, and each $\xi_{n,i}$ measures the spatio-temporal influence from the ith event $(t_i, \boldsymbol{x}_i)$ to the nth event $(t_n, \boldsymbol{x}_n)$.

5 Experiments

Using real data of POI visit events in "Kyoto", the ancient capital of Japan (a famous sightseeing city), we first evaluate the proposed model in terms of prediction performance. Next, by applying the proposed analysis method, we try to examine the properties of the geographical attention dynamics in Kyoto.

5.1 Datasets

We collected such photos that were taken within Kyoto city in 2014 and uploaded to photo-sharing site Flickr[3]. By regarding those photos as a set of photos taken on sight-

---

[3] `https://www.flickr.com/`

Fig. 3: Predictive accuracy for the spring data.

seeing tours, we constructed real data for POI visit events in Kyoto. The total number of those photos was $78,239$. By taking into account Kyoto's attractive seasons represented by cherry blossoms and autumn leaves, we focus on the spring data from March 1 to May 7 and the autumn data from October 1 to December 7. Also, from the perspective of Kyoto's sightseeing, we divide one day into $M = 4$ time slots, and set time slots TS_1, TS_2, TS_3 and TS_4 as 6 am to 11 am, 11 am to 4 pm, 4 pm to 9 pm and 9 pm to 6 am, respectively. Figure 2a indicates the number of events within each TS_m. Unsurprisingly, it is seen that many events occurred in daytime TS_2, and a relatively small number of events occurred in night-time and early-morning TS_4.

For each of the spring and autumn data, we constructed seven datasets $\mathcal{D}_1, \ldots, \mathcal{D}_7$ in the following way: We let the training period $[0, T)$ and the test period $[T, T')$ be two months and one day, respectively. In the case of the spring data, for example, for dataset $\mathcal{D}_1$, training period $[0, T)$ is March 1 to April 30 and test period $[T, T')$ is May 1, and for dataset $\mathcal{D}_2$, training period $[0, T)$ is March 2 to May 1 and test period $[T, T')$ is May 2. In the case of the autumn data, for example, for dataset $\mathcal{D}_1$, training period $[0, T)$ is October 1 to November 30 and test period $[T, T')$ is December 1, and for dataset $\mathcal{D}_2$, training period $[0, T)$ is October 2 to December 1 and test period $[T, T')$ is December 2.

5.2   Evaluation of Prediction Performance

For predicting future POI visit events, we compared the proposed model (see Eq. (5)) with the HG model (see Eq. (4)) and the baseline model (see Eq. (3)). Here, the parameters α, β and γ for the HG model were estimated by a commonly used method for learning a Hawkes process (i.e., a maximal likelihood method based on an EM algorithm (see [7])), and the parameter α for the baseline model was also estimated in the same way. For inferring α, $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$ and $\{\xi_{n,i}\}$ in the proposed model, we in particular implemented $1,000$ iterations with 200 burn-in. In view of Kyoto's sightseeing, the number of components was set as $K = 8$, and eight representative tourist spots were always used as the initial positions of parameters $\{\boldsymbol{\mu}_k; k = 1, \ldots, 8\}$ in parameter inference, for all three models.

(a) Results for seven datasets. (b) Results for four time-slots.

Fig. 4: Predictive accuracy for the autumn data.

By taking the issue of spatial resolution limitation into consideration, we decompose an appropriate rectangular region covering Kyoto's spatial domain Ω into a collection of 250×400 consecutive *tiles* $\{\Omega(b);\ b = 1, \ldots, 250 \times 400\}$ (see [17]), where each tile $\Omega(b)$ is a 100 meter square region, and we consider evaluating the predictive accuracy for future POI visit events in terms of these tiles. For each TS_m, we counted the number of events occurring within every tile $\Omega(b)$. Figure 2b shows the distribution of the number of events for each time slot TS_m in terms of the number of tiles. Interestingly, it can be seen that the distribution for TS_m, $(m = 1, 2, 3, 4)$, exhibits a power law with almost the same scaling exponent. We evaluate the predictive accuracy of an inferred spatio-temporal point process model with intensity function $\lambda^*(t) f^*(\boldsymbol{x} \mid t)$ by

$$
PA = \frac{1}{|\mathcal{H}(T, T')|} \left(\sum_{(t_n, \boldsymbol{x}_n) \in \mathcal{H}(T, T')} \ln \left\{ \lambda^*(t_n) \int_{\Omega(b(\boldsymbol{x}_n))} f(\boldsymbol{x} \mid t_n) \, d\boldsymbol{x} \right\} - \int_T^{T'} \lambda^*(t) \, dt \right)
$$

(13)

(see [24])[4], where $\mathcal{H}(T, T') = \mathcal{H}_{T'} \setminus \mathcal{H}_T$ stands for the set of events occurring in test period $[T, T']$, and $\Omega(b(\boldsymbol{x}_n))$ denotes the tile to which location $\boldsymbol{x}_n$ belongs. Here, note that PA measures the average prediction log-likelihood of $\mathcal{H}(T, T')$ (see Eq. (1)).

For all three models, the parameters other than $\{\boldsymbol{\mu}_k\}$ were randomly initialized in parameter inference. Figures 3 and 4 show the average prediction performance on five trials for the spring and autumn data, respectively. Here, the proposed model is evaluated in terms of metric PA (see Eq. (13)), compared with the HG and baseline models. Figures 3a and 4a indicate the value of PA for each dataset $\mathcal{D}_j$, and Figures 3b and 4b indicate the average value of PA restricted to each time slot TS_m. We see that the proposed model performs the best, the conventional HG model follows, and the baseline model is always worse than these two models. Unlike the other two models, the prediction performance of the proposed model was stable, and did not heavily depend on datasets and time slots. These results imply that it is significant to incorporate both

---

[4] Other performance metrics based on industry practices (see [24]) can also be used to evaluate it. We confirmed that the results for such a metric were similar to those for PA.
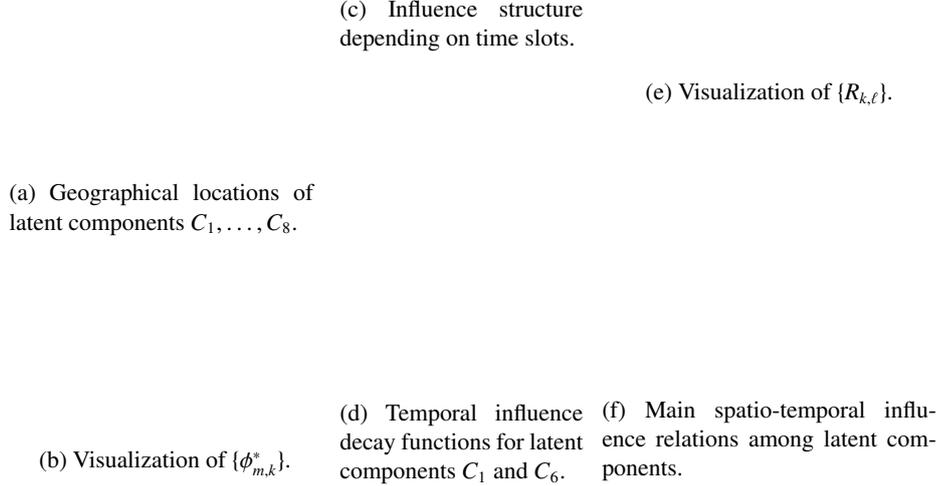
(c) Influence structure depending on time slots.

(e) Visualization of $\{R_{k,\ell}\}$.

(a) Geographical locations of latent components $C_1, \ldots, C_8$.

(d) Temporal influence decay functions for latent components C_1 and C_6.

(f) Main spatio-temporal influence relations among latent components.

(b) Visualization of $\{\phi^*_{m,k}\}$.

Fig. 5: Analysis results for the spring data.

component dependent temporal influence decay and time-slot varying influence degree, and demonstrate the effectiveness of the proposed model.

5.3 Analysis of Geographical Attention Dynamics

By applying the proposed method, we examine the properties of the geographical attention dynamics in Kyoto. Here, we only report the analysis results for the spring data (see Fig. 5).

Figure 5a displays the geographical locations of the latent Gaussian components $C_1, \ldots, C_8$ estimated, which represent the major sightseeing areas of Kyoto in the spring. Figure 5b gives a visualization result of the estimated parameters $\{\phi^*_{m,k}\}$, where each $\phi^*_m = (\phi^*_{m,1}, \ldots, \phi^*_{m,8})$ indicates the popularity distribution among components within time slot TS_m. We observe that C_1, C_2 and C_3 are always popular, and C_5 and C_6 are also popular in some time slots to a certain degree. Here, C_1, C_2 and C_3 correspond to neighborhoods of Kyoto Imperial Palace (Kyoto Gosho)[5]. Heian-jingu Shrine[6] and Kiyomizu-dera Temple[7], respectively, and they are located near Kyoto's downtown.

---

[5] https://www.japan.travel/en/spot/1168/
[6] https://www.japan.travel/en/spot/1195/
[7] https://www.japan.travel/en/spot/2199/

Also, C_5 corresponds to a neighborhood of Kinkaku-ji Temple[8] featuring a shining golden pavilion, which is located in a suburban area of Kyoto city. C_6 corresponds to Arashiyma area[9], which is a touristy district on the western outskirts of Kyoto, and famous as a place of scenic beauty.

Figure 5c shows the estimated parameters $\boldsymbol{\beta}^* = (\beta_1^*, \dots, \beta_4^*)$, which represent the influence structure depending on time slots. We can see that events occurred during morning TS_1 were the most influential, while events occurred during night-time and early-morning TS_4 were the least influential. For the estimated parameters $\{\gamma_k^*\}$, there was little difference among C_1, C_2, C_3 and C_5. Figure 5d displays the temporal influence decay functions estimated for components C_1 and C_6. This implies that the influence of events occurred in C_6 more rapidly decayed than that of events occurred in C_1.

Figure 5e shows the estimated influence relation $R_{k,\ell}$ from C_ℓ to C_k for $k, \ell = 1, \dots, 8$, where the color of the entry in the kth row and the ℓth column indicates the value of $R_{k,\ell}$. We see that the influence relations among C_1, C_2 and C_3 were substantially strong compared to the others. Figure 5f displays the main influence relations among components, where for $k, \ell = 1, \dots, 8$, an arrow from C_ℓ to C_k is drawn if $R_{k,\ell}$ is greater than the average value of $\{R_{k',\ell'}\}$. This reveals people's primary movement patterns for Kyoto's sightseeing in the spring. Here, note that the spatio-temporal influence relations $\{R_{k,\ell}\}$ significantly changed for the autumn data. Like these, the proposed method can provide interesting analysis results for Kyoto's sightseeing during a specified season. These analysis results are expected to contribute a foundation for tourism marketing.

6 Conclusion

We dealt with modeling of *geographical attention dynamics*, that is, the problem of modeling the occurrence process of POI visit events for a sightseeing city in the setting of a continuous space-time. We have proposed a novel probabilistic model for discovering the spatio-temporal influence structure among major sightseeing areas, and attempted to accurately predict POI visit events in the near future. The proposed model is constructed by combining a Hawkes process with a time-varying Gaussian mixture model in a novel way and incorporating the influence structure depending on time slots as well. We developed an efficient method of inferring the parameters in the proposed model from the observed sequence of POI visit events, and provided an analysis method for the geographical attention dynamics. Using real data of Kyoto, a Japanese sightseeing city, we demonstrated that the proposed model significantly outperforms the conventional HG model and the baseline model in terms of predictive accuracy, and revealed the spatio-temporal influence relation structure among major sightseeing areas in Kyoto from the viewpoint of geographical attention dynamics.

In this paper, we focused on Kyoto's data obtained from Flickr, a photo-sharing site. Clearly, it is possible to apply the proposed method to other sightseeing cities and geographical regions including several sightseeing cities. Our immediate future work is to evaluate the proposed method for various sightseeing cities around the world and to explore POIs of variable geographical scales. We also supposed that the latent Gaussian

---

[8] `https://kyoto.travel/en/shrine_temple/132`
[9] `https://www.japan.travel/en/spot/1142/`

components extracted by the proposed method represent major tourism topics and can be identified with major sightseeing areas. Our future work includes exploring spatial distributions other than Gaussian mixture. In several photo-sharing services, there are many photos that are annotated not only with GPS locations and time-stamps but also with text documents, and a method of detecting spatio-temporally exclusive topics from those data is investigated (see [17]). By applying such a method, we also plan to develop a framework of easily interpreting those latent components in terms of tourism topics.

References

1. Aalen, O., Borgan, O., Gjessing, H.: Survival and event history analysis: A process point of view. Springer (2008)
2. Chen, C., Yang, H., Lyu, M., King, I.: Where you like to go next: successive point-of-interest recommendation. In: Proceedings of IJCAI'13. pp. 2605–2611 (2013)
3. Chen, S., Yuan, X., Wang, Z., Guo, C., Liang, J., Wang, Z., Zhang, X., Zhang, J.: Interactive visual discovering of movement patterns from sparsely sampled geo-tagged social media data. IEEE Transactions on Visualization and Computer Graphics **22**(1), 270–279 (2016)
4. Daneshmand, H., Gomez-Rodriguez, M., Song, L., Schölkopf., B.: Estimating diffusion network structures: Recovery conditions, sample complexity & soft-thresholding algorithm. In: Proceedings of ICML'14. pp. 793–801 (2014)
5. Du, N., Dai, H., Upadhyay, U., Gomez-Rodriguez, M., Song, L.: Recurrent marked temporal point processes: Embedding event history to vector. In: Proceedings of KDD'16. pp. 1555–1564 (2016)
6. Farajtabar, M., Du, N., Gomez-Rodriguez, M., Valera, I., Zha, H., Song, L.: Shaping social activity by incentivizing users. In: Proceedings of NIPS'14. pp. 2474–2482 (2014)
7. Farajtabar, M., Wang, Y., Gomez-Rodriguez, M., Li, S., Zha, H., Song, L.: Coevolve: A joint point process model for information diffusion and network evolution. Journal of Machine Learning Research **18**(41), 1–49 (2017)
8. Feng, S., Li, X., Zeng, Y., Cong, G., Chee, Y., Yuan, Q.: Personalized ranking metric embedding for next new poi recommendation. In: Proceedings of IJCAI'15. pp. 2069–2075 (2015)
9. Frey, B., Dueck, D.: Clustering by passing messages between data points. Science **315**(5814), 972–976 (2007)
10. Gao, H., Tang, J., Hu, X., Liu, H.: Exploring temporal effects for location recommendation on location-based social networks. In: Proceedings of RecSys'13. pp. 93–100 (2013)
11. Gao, S., Ma, J., Chen, Z.: Modeling and predicting retweeting dynamics on microblogging platforms. In: Proceedings of WSDM'15. pp. 107–116 (2015)
12. Gomez-Rodriguez, M., Leskovec, J., Krause, A.: Inferring networks of diffusion and influence. In: Proceedings of KDD'10. pp. 1019–1028 (2010)
13. Hawkes, A.: Spectra of some self-exciting and mutually exiting point process. Biometrika **58**(1), 83–90 (1971)
14. Liu, Y., Liu, C., Liu, B., Qu, M., Xiong, H.: Unified point-of-interest recommendation with temporal interval assessment. In: Proceedings of KDD'16. pp. 1015–1024 (2016)
15. Ogata, Y.: On lewis' simulation method for point processes. IEEE Transactions on Information Theory **27**(1), 23–31 (1981)

16. Shen, H., Wang, D., Song, C., Barabási, A.L.: Modeling and predicting popularity dynamics via reinforced poisson processes. In: Proceedings of AAAI'14. pp. 291–297 (2014)
17. Shin, S., Choi, M., Choi, J., Langevin, S., Bethune, C., Horne, P., Krnonenfeld, N., Kannan, R., Drake, B., Park, H., Choo, J.: Stexnmf: Spatio-temporally exclusive topic discovery for anomalous event detection. In: Proceedings of ICDM'17. pp. 435–444 (2017)
18. Stephens, M.: Bayesian analysis of mixture models with an unknown number of components: An alternative to reversible jump methods. The Annals of Statistics **28**(1), 40–74 (2000)
19. Wang, D., Song, C., Barabási, A.L.: Quantifying long-term scientific impact. Science **342**(6154), 127–132 (2013)
20. Yuan, Q., Cong, G., Ma, Z., Sun, A., Magnenat-Thalmann, N.: Time-aware point-of-interest recommendation. In: Proceedings of SIGIR'13. pp. 363–372 (2013)
21. Zhang, J., Chow, C.: Spatiotemporal sequential influence modeling for location recommendations: a gravity-based approach. ACM Transactions on Intelligent Systems and Technology **7**(1), 11:1–11:25 (2015)
22. Zhao, Q., Erdogdu, M., He, H., Rajaraman, A., Leskovec, J.: Seismic: A self-exciting point process model for predicting tweet popularity. In: Proceedings of KDD'15. pp. 1513–1522 (2015)
23. Zhou, K., Zha, H., Song, L.: Learning social infectivity in sparse low-rank networks using multi-dimensional hawkes processes. In: Proceedings of AISTATS'13. pp. 641–649 (2013)
24. Zhou, Z., Matteson, D., Woodard, D., Henderson, S., Micheas, A.: A spatio-temporal point process model for ambulance demand. Journal of the American Statistical Association **110**(509), 6–15 (2015)