Human Mobility Synchronization and Trip Purpose Detection with Mixture of Hawkes Processes

Pengfei Wang*
CNIC, Chinese Academy of Sciences
Beijing, China
wpf@cnic.cn

Yanjie Fu*†
Missouri U. of Sci. and Tech.
MO, USA
fuyan@mst.edu

Guannan Liu
Beihang University
Beijing, China
liugn@buaa.edu.cn

Wenqing Hu
Missouri U. of Sci. and Tech.
MO, USA
huwenqing.pku@gmail.com

Charu Aggarwal
IBM T. J. Watson Research Center
NY, USA
charu@us.ibm.com

ABSTRACT

While exploring human mobility can benefit many applications such as smart transportation, city planning, and urban economics, there are two key questions that need to be answered: (i) What is the nature of the spatial diffusion of human mobility across regions with different urban functions? (ii) How to spot and trace the trip purposes of human mobility trajectories? To answer these questions, we study large-scale and city-wide taxi trajectories; and furtherly organize them as arrival sequences according to the chronological arrival time. We figure out an important property across different regions from the arrival sequences, namely *human mobility synchronization effect*, which can be exploited to explain the phenomenon that two regions have similar arrival patterns in particular time periods if they share similar urban functions. In addition, the arrival sequences are mixed by arrival events with distinct trip purposes, which can be revealed by the regional environment of both the origins and destinations. To that end, in this paper, we develop a joint model that integrates Mixture of Hawkes Process (MHP) with a hierarchical topic model to capture the arrival sequences with mixed trip purposes. Essentially, the human mobility synchronization effect is encoded as a synchronization rate in the MHP; while the regional environment is modeled by introducing latent *Trip Purpose* and *POI Topic* to generate the Point of Interests (POIs) in the regions. Moreover, we provide an effective inference algorithm for parameter learning. Finally, we conduct intensive experiments on synthetic data and real-world data, and the experimental results have demonstrated the effectiveness of the proposed model.

KEYWORDS

Human mobility, synchronization, trip purpose, Hawkes process, variational inference

*These two authors contributed equally to this work.
†Corresponding author.

KDD'17, August 13–17, 2017, Halifax, NS, Canada.
 ISBN 978-1-4503-4887-4/17/08…$15.00
DOI: http://dx.doi.org/10.1145/3097983.3098067

1 INTRODUCTION

The pervasiveness of GPS-equipped devices have accumulated large-scale human mobility data, such as taxicab GPS trajectories, cellphone traces, geo-tagged posts and photos. Such fine-grained trajectory data, including where and when people leave, where and when they arrive, represent an invaluable source of information for understanding the nature of and exploring the semantics of human mobility, which can be benefit many applications such as smart transportation, city planning and governance, and urban economics [7, 21, 26, 27].

In most cases, people's mobility behaviors after they leave the taxi remains un-tracked, and we have no clues about their specific destinations and afterward activities. Therefore, in order to extend the analysis of human mobility, the focus of this paper is to develop a general probabilistic model to describe the arrival events at a region and spot trip purposes simultaneously. Then, the following two key questions need to be answered: (i) What is the spatial diffusion of human mobility across functional regions? (ii) How to spot and trace the latent trip purposes of taxi trajectories?

To answer these questions, we study the large-scale and city-wide taxi trajectories. We observe and identify a very important property of human mobility, which is, *human mobility synchronization*. In other words, if two regions share similar spatial configurations and urban functions in a particular time period, the two regions are likely to have similar patterns of arrival events. For instance, in the Greater New York area, there are two famous international airports (Newark and JFK) providing domestic and international flight services. It is likely that trips to the two places usually share similar purposes: to fly out or to pick up others; as Figure 1 shows, the distributions of arrival events around the two places in a common workday shares similar temporal patterns, with peaks in the early morning and afternoon.

In particular, a trajectory can be regarded as a mobility arrival event consisting of three elements: the destination region, the origin region, and the time period and each arrival event is associated with a trip purpose. However, it is difficult to identify the trip purpose of a single arrival event with high confidence, if each trip is analyzed individually and independently. Alternatively, we organize arrival events as a sequence in terms of chronological arrival times, so as to obtain higher confidence for identifying trip purposes at an aggregate level. For instance, if we observe

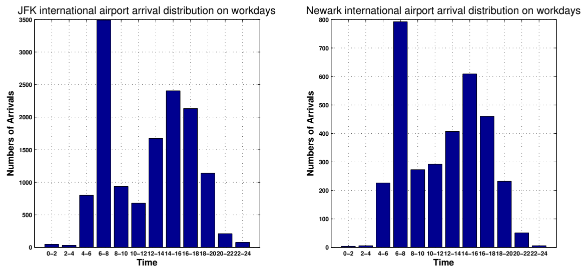


Figure 1: Taxi arrival distribution of JFK Airport and Newark Airport

massive arrival events from 7AM to 9AM in a specific region, it is *likely* that the trip purposes are to work. Moreover, suppose this region has many companies, banks, office buildings, then the function of the region is likely to be office function, and we have *higher confidence* to identify the trip purposes of arrival events as to work. Furthermore, assuming there are two regions of office function, we observe massive arrival events in both regions from 7AM to 9AM. Based on the synchronization property, we have *much higher confidence* to identify the trip purposes of the arrival events in both regions are to work. While the above intuition demonstrates the correlations among mobility arrivals, functional regions, and trip purposes, how to jointly model the correlations for describing mobility arrivals and tracing trip purposes is still a technical question.

In addition, trip purposes are not only relevant to the urban function of the destinations but also related to the origins. For example, a trajectory from a residential region to an office region is likely to be associated with a working purpose. However, a region in urban areas contains different types of POIs and thus is usually mixed with multiple functionalities [33], making it difficult to exploit functional regions for spotting trip purposes. For example, the Times Square is one famous scenic spot mixed with numerous office buildings. The arrival events in Times Square then can be attributed to various purposes. To overcome the challenge, we exploit a topic modeling method, where origin and destination regions are regarded as documents and POIs are regarded as words, for profiling the portfolios of POI topics in each region.

Along this line, we provide a unique perspective of modeling human mobility data as a stochastic process for simultaneously capturing mobility arrival events and trip purposes. Specifically, we develop a unified probabilistic model by combining a multivariate Hawkes mixture model with a hierarchical POI topic model. First, the Hawkes mixture model exploits the synchronization property to describe mobility arrivals and trace trip purposes. In particular, we first organize trajectories as a sequence of chronological arrival events, with each having a specific trip purpose. We then introduce a mixture of Hawkes processes to track the arrival sequences with distinct trip purposes, by assuming each Hawkes process corresponds to a kind of trip purposes. Essentially, in modeling the arrival sequences, the emergence of arrival events with a specific purpose at a region can possibly synchronize to other similar regions, such effect is modeled with a synchronization rate. Second, the trip purpose of an arrival can be inferred by its POI topics of origin and the POI topics of destination. Then, we introduce a latent variable *POI Topic* to model the mixture of regional functionalities with regards to POI distributions. Therefore, a hierarchical topic model can be extended by establishing connections between POI Topic and the latent trip purpose. Also, we provide an effective learning algorithm for estimating the parameters. Finally, we conduct intensive experiments on both synthetic data and real-world data, and the experimental results have demonstrated the effectiveness of our proposed model.

2 PRELIMINARY

2.1 Arrival Sequences

In order to distinguish the traffic flow into different purposes and identify the corresponding arrival patterns in aggregate, we firstly transform the independent arrival events into arrival sequences. To be more specific, we firstly divide the map according to the latitude and longitude, resulting in multiple divided small regions in the city. Then, each taxi trajectory from one region to another can be regarded as an event, with each consisting of the origin and destination, as well as the time. Considering the arrival time chronologically, we can organize the events as an arrival sequence.

Usually the events in time sequences are homogeneous and only the underlying temporal patterns can be extracted from the sequences. Nevertheless, the arrival events are assumed to have distinct types or goals, and in order to reveal the hidden purposes of the taxi trips, we resort to the regional environment, which can be further depicted by the POIs located in the regions, since the POIs in a region can be viewed as a set of independent functionalities, and together reflecting the functional mixtures of the regions. Therefore, the arrival events are heterogeneous with the information about the POIs around the origins and destinations, as well as the arrival time, which can be defined as a three-element tuple: (POI^{origin}, $POI^{destination}$, arrival time).

Furthermore, the events are assumed to be inter-correlated with each other because different regions with similar functionality may synchronize with each other on the arrival events. For example, in holiday seasons, when we see people flock in one particular business area of a city, the other region with shopping malls assembled has higher probability to have a great number of arrival events at the same time. In this paper, we define such effect across different regions as *human mobility synchronization effect*.

Definition 1. ***(Human mobility synchronization)*** *Human mobility synchronization refers to the phenomenon that the arrival events with a particular trip purpose at two or more regions has similar arrival rate at the same time.*

By discovering the synchronization effect across different regions, the arrival patterns with a specific trip purpose can be mutually learned from the synchronized regions, and to better reveal the distribution of urban functions in the city. Therefore, the arrival sequence with several different trip purposes contains heterogeneous events which are also self-correlated.

2.2 Problem Definition

Given a city map and taxi trajectories, along with the accurate positions of POIs on the map, we can organize the taxi trajectories as arrival sequences accordingly. The arrival sequence consists of a set of arrival events $E = \{E_1, E_2, \cdots, E_N\}$, in which each event corresponds to a three-element tuple (W_n^o, W_n^d, t_n), representing the POIs around the origin and the destination, and the arrival time respectively. Here $W_n^o = \{w_{n1}^o, w_{n2}^o, \cdots, w_{ni}^o, \cdots\}$ indicates the set of POIs in the regions of the origin point O_n, with each element in the set w_{ni}^o representing one particular POI in the region. Assume that there are M trip purposes underlying all the taxi trajectories, and then the problem can then be defined as follows, i) modeling the heterogeneous and self-correlated arrival sequences; ii) identifying the M types of trip purposes from the arrival sequences; iii) revealing the traffic flow with different trip purposes at different time periods for each region.

3 MODEL

3.1 Hawkes Process

Before we introduce the model for the arrival sequences, we briefly introduce the Hawkes Process [12]. Point process models the events in a time sequence with N events $\{E_1, \cdots, E_N\}$. It generally assumes that historical events before time t, which is $\mathcal{H}(t)$, can influence the future events with a conditional intensity function as follows,

$$\lambda(t|\mathcal{H}(t)) = \lim_{\Delta t \to 0} \frac{\mathbb{E}[N(t+\Delta t)|\mathcal{H}_t]}{\Delta t}. \tag{1}$$

The above intensity function simply depicts the expected rates conditioned on the past events. Obviously, Hawkes Process does not follow Markovian property since the events at t are not only influenced by the events at time $t-1$. More formally, the univariate Hawkes Process $N(t)$ is defined by,

$$\lambda(t) = \mu(t) + \int_{-\infty}^{t} g(t-s)\mathrm{d}N(s), \tag{2}$$

where μ is the base intensity, $g(\cdot)$ is a kernal function describing the time decay effect of past events on current events. With the Hawkes Process, the *self-exciting* property can be captured, and such property is indeed a common phenomenon in different types of sequences. In our previously defined arrival sequences, such property may not hold if the events are treated as a univariate Hawkes Process with same intensity function, because the influence across different regions cannot be encoded through the *self-exciting* process.

Obviously, the univariate Hawkes Process can be extended to handle multi-dimensional random processes, which is know for multivariate Hawkes Process (MHP). In the multivariate case, different dimensions are correlated with each other, and a kernal function $g_{d'd}$ is introduced to model the triggering effect between a pair of dimensions. To be more specific, suppose the sequence can be described by a D-dimensional variable, then for each dimension d, the intensity function can be defined separately as follows,

$$\lambda_d(t) = \mu_d(t) + \sum_{d'=1}^{D} \int_{\infty}^{t} g_{d'd}(t-s)\mathrm{d}N_{d'}(s). \tag{3}$$

Considering the arrival sequences, each region i can be treated as a separate dimension; however, the arrival events at the same region may show distinct arrival patterns if trip purposes are considered separately, thus we have to exploit mixture of Hawkes Process (MHP) to incorporate the latent trip purposes in modeling the heterogeneous arrival sequences.

3.2 Mixture of Hawkes Processes in Modeling Arrival Sequences

In this paper, we aim to model the previously defined arrival sequences in terms of the arrival time at different regions of the city, and meanwhile with distinct trip purposes. Moreover, the intercorrelations between the events, or the *synchronization effect* across different regions are considered to uncover the arrival patterns of the trip. Since we assume that there are M types of trip purposes underlying all the arrival sequence, we then exploit MHP to model the heterogeneous arrival sequences.

We partition the city map into different regions, and let the "regions" be labeled as $i = 1, 2, ..., I$. Then, the arrival sequence can be regarded as a series of events at these I regions. We assume that there are M trip purposes underlying the taxi trajectories, then for each region i, the arrival events with purpose m can be modeled as one particular Hawkes Process. Therefore, we introduce a set of latent variables $\{G_n\}$ to represent the purpose of each event E_n respectively. Specifically, $G_n = \{G_{n1}, ..., G_{nM}\}$ is a binary vector with $G_{nm} = 1$ if and only if the arrival event E_n belongs to purpose m.

Meanwhile, according to the definition of *human mobility synchronization*, the arrival patterns of trips with similar purposes can synchronize across different regions. Therefore, suppose there is a *synchronization structure* for trip purpose m between the regions, which can be represented as a set of non-negative weights $\{\alpha_{ij}^m\}$, $i, j = 1, 2, ..., I$, $m = 1, 2, ..., M$ over all the regions distributed in the map.

Therefore, we model the arrival sequence with *synchronization structure* in the framework of mixture of Hawkes Process [32]. Specifically, we have a multivariate Hawkes Process $\{N_{m,i}(t)\}$, representing the number of arrival events at region i, with trip purpose m at time t, and the intensity can be defined as,

$$\lambda_{i,m}(t) = \mu_{i,m} + \sum_{j=1}^{I} \alpha_{ji}^m \int_{-\infty}^{t} g(t-s)dN_{m,j}(s) . \tag{4}$$

Here $\mu_{i,m}$ specifies the rate that region i got visited spontaneously with trip purpose m. The base rate of visiting with purpose m is given by γ_m, and the specific rate related to the region with node i is μ_i, thus $\mu_{i,m} = \mu_i \gamma_m$. The function $g(t) : \mathbb{R}_+ \to \mathbb{R}_+$ is a kernel function expressing the time-decay effect. We can pick $g(t)$ to be a function that is equal to zero when $t \geq 2$ (hours), such that we capture the temporal regularity of human mobility.

We can further extend Eq. (4) in a more practical form as follows,

$$\lambda_{i,m}(t) = \mu_i \gamma_m + \sum_{t_l \leq t} G_{lm} \alpha_{i_l i}^m g(t-t_l) . \tag{5}$$

3.3 Hierarchical Topic Modeling for Trip Purposes

As discussed previously, the trip purpose can be reflected by the neighborhood environment of both the origin place and the destination place, because the POIs in the neighborhood corresponds to travelers' afterward activities when they leave the taxi in an aggregate level. However, the regional environment is usually mixed with multiple categories of POIs, in order to better uncover the urban functionality of a region, we introduce a latent variable *POI*

Topic$\in \{1, \cdots, r, \cdots, R\}$ to represent the mixture of POIs, which is indeed in spirit of traditional topic model such as LDA [1]. Then, the POIs W^o located in the origin neighborhood and the POIs W^d in the destination neighborhood can be regarded as words in documents, which can be generated from *POI Topic* specific probability distribution.

Then, the purpose of a trip G_n can be recovered from the POI Topics of both ends of the trajectories. For example, if the origin place is identified to be a residential area with houses, gardens, and supermarkets, while the destination region consists of many office buildings, the trip purpose can be inferred as pairwise POI Topics ("*residential*"→ "*office*"), which can be semantically explained as a working purpose trip. Since the trip purpose is a latent variable itself, the POI Topics of both ends are generated from a purpose-specific probability distribution respectively. Furthermore, we assume that each G_n is sampled i.i.d from a multinomial distribution parameterized by a multinomial distribution $\pi = [\pi_1, ..., \pi_M]$.

In a nutshell, we can construct a hierarchical topic model with the two latent variables, *i.e.*, trip purpose and POI Topic. Then, for each trip from origin point o to destination point d, the POIs locate in the origin and destination regions are generated as follows,

- Generate a purpose $m \sim \texttt{Multi}(\pi)$
- Generate the POI Topic for the origin $z_o \sim Multi(\Phi_{mz})$
  - For each POI w^o in the origin neighborhood
  - Generate the POI $w^o \sim Multi(\beta_{zw})$
- Generate the POI Topic for the destination $z_d \sim Multi(\Phi_{mz})$
  - For each POI w^d in the origin neighborhood
  - Generate the POI $w^d \sim Multi(\beta_{zw})$

Therefore, the POIs around the origin point o and destination point d generated by the trip purpose G can be written as,

$$p(\mathbf{W}^o, \mathbf{W}^d|G) = \sum_{z^o} \sum_{z^d} p(\mathbf{W}^o|z^o)p(\mathbf{W}^d|z^d)p(z^o, z^d|G). \quad (6)$$

3.4 Joint MHP for Arrival Sequences

Considering both the mixture of Hawkes Processes and the hierarchical topic modeling for trip purposes, we can jointly model the arrival sequences by incorporating the temporal arrival events and the regional environment. Here, we provide a toy example to show how the arrival sequences with multiple trip purposes are modeled. As shown in Figure 2, in modeling the arrival sequences at *Time Square* and *Wall Street*, both the regional environment of the origins and destinations are taken into account, which are generated from the latent trip purposes. Because both *Time Square* and *Wall Street* are mixed with scenic spots and office buildings, the arrival sequences show different patterns with regards to specific latent trip purposes. For example, in the early morning, the arrival events are mostly for working purpose coming from residential areas such as *Newport*, *Brooklyn*, etc. While in later time of the day, travelers may also arrive at the spots for sightseeing purpose, and their origin places are usually mixed with hotel, and other scenic spots. Furthermore, since the arrival events are similar at the two places, a *synchronization effect* between the two areas will collaboratively influence the arrival events of each other.

Specifically, given the sequence E, each event n is represented by $\{t_n, i_n, <\mathbf{W}_n^o, \mathbf{W}_n^d>\}$, and suppose the latent trip purpose G_n is given, then the likelihood of the arrival sequences can be written

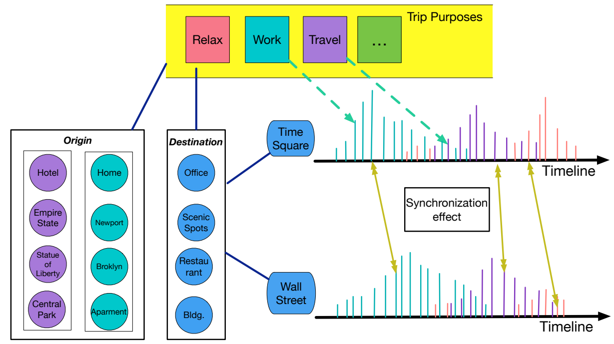


Figure 2: Toy example of Joint MHP for modeling arrival sequences

as,

$$L(G, t, \mathbf{W}) = \prod_{n=1}^{N} p(G_n)p(\mathbf{W}_n^o, \mathbf{W}_n^d|G_n)p(t_n|G_n). \quad (7)$$

3.5 Model Inference

Assuming that the latent variable G and then z are observed, then the likelihood function for the complete data $(G, z, t, W) = \{(G_n, z_n, t_n, W_n)\}$, with $t_0 = 0$ and $t_N = T$, is given by

$$\begin{aligned} L(G, z, t, W) &= \textstyle\prod_{n=1}^{N} \prod_{m=1}^{M} p(\mathbf{W}_n^o, \mathbf{W}_n^d, z_o, z_d|G_{nm}) \\ &\times [\pi_m \lambda_{i_n,m}(t_n)]^{G_{nm}} \times \exp\left(-\sum_{i=1}^{I} \textstyle\sum_{m=1}^{M} \displaystyle\int_0^T \lambda_{i,m}(s)ds\right). \end{aligned} \quad (8)$$

We introduce a mean-field fully factorized variational distribution parametrized by multinomial distribution ϕ's and ζ's as follows:

$$q(G_{1:N}, Z|\phi, \zeta) = \prod_{n=1}^{N} q(G_n|\phi_n) \times \prod_{m=1}^{M} q(z_n^o|\zeta_m^o) \times q(z_n^d|\zeta_m^d). \quad (9)$$

Thus, by using the concavity of log-function, a tight lower bound of the log likelihood can be introduced

$$\begin{aligned} \mathcal{L}(t, W) &\equiv \log\left(\int_{\{(G,z)\}} L(G, z, t, W)d\{(G, z)\}\right) \\ &\geq \int_{\{(G,z)\}} \log\left(L(G, z, t, W)\frac{d\{(G, z)\}}{dq(\{(G, z)\})}\right) dq(\{(G, z)\}) \\ &= \mathbf{E}_q[\mathcal{L}(G, z, t, W)] + \mathcal{E}[q] \equiv \mathfrak{L}, \end{aligned} \quad (10)$$

where $\mathcal{L}(G, z, t, W) = \log L(G, z, t, W)$ is the log-likelihood and

$$\begin{aligned} \mathcal{E}[q] &= -\mathbf{E}[\log q(\phi)] - \mathbf{E}[\log q(\zeta^o)] - \mathbf{E}[\log q(\zeta^d)] \\ &= -\sum_{n=1}^{N} \sum_{m=1}^{M} \phi_{nm} \log \phi_{nm} \\ &- \sum_{m=1}^{M} \sum_{r^o=1}^{R} \zeta_{mr^o}^o \log \zeta_{mr^o}^o - \sum_{m=1}^{M} \sum_{r^d=1}^{R} \zeta_{mr^d}^d \log \zeta_{mr^d}^d \end{aligned} \quad (11)$$

The right-hand side $\mathfrak{L}$ in (10) is the so-called evidence lower bound (ELBO), which will be used as the surrogate to the true log-likelihood in inference and in learning. We then have,

$$
\begin{aligned}
\mathfrak{L} &= \sum_{n=1}^{N}\sum_{m=1}^{M}\phi_{nm}(\log\pi_m + \mathbf{E}_q[\log\lambda_{i_n,m}(t_n)] + \mathbf{E}_q\log p_m(W_n)) \\
&- \sum_{i=1}^{I}\sum_{m=1}^{M}\int_0^T \mathbf{E}_q[\lambda_{i,m}(s)]ds + \mathcal{E}[q] \ .
\end{aligned} \tag{12}
$$

To calculate (12) more explicitly, we make use of the formula (5) and we further have

$$
\begin{aligned}
& \sum_{i=1}^{I}\sum_{m=1}^{M}\int_0^T \mathbf{E}_q[\lambda_{i,m}(s)]ds \\
= & \sum_{i=1}^{I}\sum_{m=1}^{M}\left[\gamma_m\mu_i T + \int_0^T \sum_{t_l\le s}\phi_{lm}\alpha^m_{i_l i}g(s-t_l)ds\right] \\
= & \ T\sum_{i=1}^{I}\sum_{m=1}^{M}\gamma_m\mu_i + \sum_{n=1}^{N}\mathcal{G}(T-t_n)\sum_{m=1}^{M}\phi_{nm}\sum_{i=1}^{I}\alpha^m_{i_n i} \ .
\end{aligned} \tag{13}
$$

To deal with the sum involving $\mathbf{E}_q[\log\lambda_{i_n,m}(t_n)]$, we make use of the formula (5) and we write $\mathbf{E}_q[\log\lambda_{i_n,m}(t_n)] = \mathbf{E}_q[\log(\gamma_m\mu_i + \sum_{t_l<t_n} G_{lm}\alpha^m_{i_l i_n}g(t_n-t_l))]$, after that we apply Jensen's inequality:

$$
\begin{aligned}
& \mathbf{E}_q[\log\lambda_{i_n,m}(t_n)] \\
= & \ \mathbf{E}_q[\log(\gamma_m\mu_i + \sum_{t_l<t_n} G_{lm}\alpha^m_{i_l i_n}g(t_n-t_l))] \\
\ge & \ \eta^m_{nn}\log\left(\frac{\gamma_m\mu_i}{\eta^m_{nn}}\right) + \sum_{l=1}^{n-1}\phi_{lm}\eta^m_{ln}\log\left(\frac{\alpha^m_{i_l i_n}g(t_n-t_l)}{\eta^m_{ln}}\right) \\
= & \ \eta^m_{nn}\log(\gamma_m\mu_{i_n}) + \sum_{l=1}^{n-1}\phi_{lm}\eta^m_{ln}\log(\alpha^m_{i_l i_n}g(t_n-t_l)) \\
& -\eta^m_{nn}\log(\eta^m_{nn}) - \sum_{l=1}^{n-1}\phi_{lm}\eta^m_{ln}\log(\eta^m_{ln}) \ .
\end{aligned} \tag{14}
$$

Here the set of branching variables $\{\eta^m, m = 1,2,...,M\}$ are chosen such that each η^m is a $N\times N$ lower–triangular matrix with n–th row $\eta^m_n = [\eta^m_{1n},...,\eta^m_{nn}]^T$. The branchings η also satisfies the following conditions

$$
\eta^m_{ln} \ge 0 \ , \ l = 1,2,...,n \ , \tag{15}
$$

$$
\eta^m_{nn} + \sum_{l=1}^{n-1}\phi_{lm}\eta^m_{ln} = 1 \ . \tag{16}
$$

While for $\mathbf{E}_q[\log p(\mathbf{W}^o_n, \mathbf{W}^d_n|G_n)]$, since we introduce the fully factorized variational distribution (9), we now have

$$
\begin{aligned}
& \mathbf{E}_q[\log p(\mathbf{W}^o_n, \mathbf{W}^d_n, z_o, z_d|G_n)] \\
= & \ \mathbf{E}_q[\log p(z_o|G_n)] + \mathbf{E}_q[\log p(\mathbf{W}^o_n|z_o)] \\
+ & \ \mathbf{E}_q[\log p(\mathbf{z}_d|G_n)] + \mathbf{E}_q\log p(\mathbf{W}^d_n|z_d) \\
= & \sum_{z_o=1}^{R}\zeta^o_{mz_o}\log\zeta^o_{mz_o} + \sum_{z_o=1}^{R}\sum_{w=1}^{C}\zeta^o_{m,z^o}\log\beta_{z^o,c^o_w} \\
+ & \sum_{z_d=1}^{R}\zeta^d_{mz^d}\log\zeta^d_{mz^d} + \sum_{z_d=1}^{R}\sum_{w=1}^{C}\zeta^d_{m,z^d}\log\beta_{z^d,w} \ .
\end{aligned} \tag{17}
$$

Summarizing (12), (13), (14), and (17), we can maximize the ELBO by introducing the constraints $\sum_{m=1}^{M}\phi_{nm} = 1$ as Lagranging multiplier. Then, the variational distribution ϕ can be inferred as follows,

$$
\begin{aligned}
& \phi_{nm} \propto \pi_m : \text{prior} \\
& \times\prod_{r_o=1}^{R}(\zeta^o_{mr^o})^{\zeta^o_{mr^o}}\prod_{r^d=1}^{R}(\zeta^d_{mr^d})^{\zeta^d_{mr^d}} : \text{POI topics} \\
& \times\prod_{r^o=1}^{R}\prod_{c^o=1}^{C}(\beta^o_{r^o c^o})^{\zeta^o_{mr^o}W^o_{nc^o}}\prod_{r^d=1}^{R}\prod_{c^d=1}^{C}(\beta^d_{r^d c^d})^{\zeta^d_{mr^d}W^d_{nc^d}} : \text{POIs} \\
& \times(\gamma_m\mu_{i_n})^{\eta^m_{nn}} : \text{ self triggering} \\
& \times\prod_{l=1}^{n-1}(\alpha^m_{i_l i_n}g(t_n-t_l))^{\phi_{lm}\eta^m_{ln}} : \text{ influences from the past} \\
& \times\prod_{l=n+1}^{N}(\alpha^m_{i_n i_l}g(t_l-t_n))^{\phi_{lm}\eta^m_{nl}} : \text{ influences to the future} \\
& \times\exp\left(-\mathcal{G}(T-t_n,t_n)\sum_{i=1}^{I}\alpha^m_{i_n i}\right) : \text{ influences by trip purpose.}
\end{aligned} \tag{18}
$$

We can see from (18) that the inference for ϕ can be decomposed to different building blocks of the joint model including the influence in MHP and the purposes inferred from the regional environment.

Furthermore, we can infer the variational multinomial ζ^o and ζ^d by taking the derivative with respect to ELBO with the normalization condition as Lagrangian, and let $\epsilon = \frac{\sum_{n=1}^{N}\phi_{nm}}{1-\sum_{n=1}^{N}\phi_{nm}}$, we can get

$$
\zeta^o_{m,r} \propto \prod_{c=1}^{C}(\beta_{rc})^{\epsilon W^o_{nc}} \ , \tag{19}
$$

$$
\zeta^o_{m,r} \propto \prod_{c=1}^{C}(\beta_{rc})^{\epsilon W^d_{nc}} \ . \tag{20}
$$

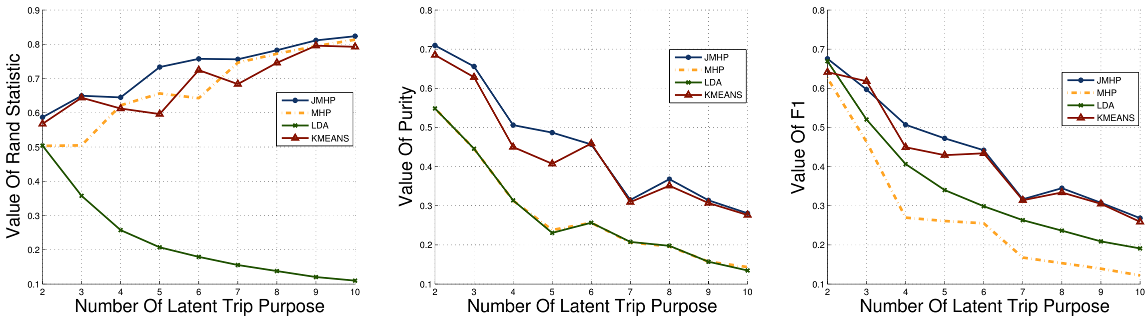
After updating the variational parameters ϕ and ζ, we can proceed to estimate and update the parameters involved in the joint model including the self-instantaneous visit rates μ_i, the per-region visit rates γ_i, the synchronization rates $\alpha^m_{ij} \in \mathbb{R}^{12}_+$, the regularization prior for sampling the purposes π_m, and the multinomial distribution of POIs over topics β_{rc},

$$
\pi_m \propto \sum_{n=1}^{N}\phi_{nm} \ , \tag{21}
$$

$$
\mu_i \propto \frac{\sum_{m=1}^{M}\phi_{nm}\sum_{n=1}^{N}\delta_{i_n,i}\eta^m_{nn}}{\sum_{m=1}^{M}\gamma_m T} \ , \tag{22}
$$

$$
\gamma_m \propto \frac{\sum_{n=1}^{N}\phi_{nm}\eta^m_{nn}}{\sum_{i=1}^{I}\mu_i T} \ , \tag{23}
$$

$$
\alpha^m_{ij} = \frac{\sum_{n=1}^{N}\sum_{l=1}^{n-1}\phi_{nm}\phi_{lm}\eta^m_{ln}\delta_{i_l i}\delta_{i_n j}}{\sum_{n=1}^{N}\mathcal{G}(T-t_n)\phi_{nm}\delta_{i_n i}} \ , \tag{24}
$$



(a) (b) (c)

Figure 3: Experiments on synthetic data. (a): Comparison of Rand Statistic. (b): Comparison of purity. (c): Comparison of F1-Value

$$\beta_{rc} \propto \sum_n \sum_m \phi_{nm}(\zeta^o_{m,c} W^o_{m,c} + \zeta^d_{m,c} W^d_{m,c}). \tag{25}$$

4 EXPERIMENT

4.1 Synthetic Data

We simulate data with pre-defined M trip purposes by running Ogata’s modified thinning algorithm [23]. Specifically, given model dimensions (M, I, T), the parameters $(\mu, \gamma, \alpha, \beta)$ can be drawn. Then, we can randomly generate the distribution of the arrival events, including the arrival time, and the corresponding POI distribution of the origin and destination points. We mix the trips with different purposes into one arrival sequence by ordering the events chronologically. The POIs located in the regions were sampled from the hierarchical topic models. In this experiment, 2,000-50,000 events involving $M = 10$ purposes, $Z = 10$ latent topics, $C = 50$ words and $I = 50$ nodes are generated.

Since the simulated data has the “label” of trip purposes, we can then testify whether our model can accurately distinguish traffic flow with particular purposes from all the trajectories. More specifically, the parameter ϕ inferred from the joint model reveal the trip purpose distribution of each arrival event, and we can label an event by the purpose with the largest value of ϕ, *i.e.*, $m^*_n = \text{argmax}_m\, \phi_{n,m}$. Thus, the task can be regarded as clustering the trajectories into different types.

Baseline Methods. We compare our proposed approach with the following baseline methods:

- MHP: In MHP, we only model the time sequence without considering the particular POIs of the regions. Similar to the joint MHP model, we can also estimate the parameters concerning the purpose distribution of each trip.
- LDA: Given the POIs in the origin and destination region, we can regard each trip as a document with each POI as a word. Then, we can use topic model LDA to learn the topic distribution of each trip and categorize the trip into one particular cluster.
- K-means: K-means is exploited to cluster the related POI vectors of each trip.

Evaluation metrics. Since the task of identifying trip purposes can be regarded as clustering, and we also simulate the labels, we can use several external measures to evaluate the performance.

Purity. Purity represents the extent to which a cluster can represent a particular labeled class, which can be computed by

$$\text{purity} = \sum_j^K \frac{N_j}{N} max(\frac{N_{ij}}{N_j}), \tag{26}$$

where $N_{ij} = |P_i \bigcap C_j|$ and $N_j = |C_j|$, P_i is a set of data which are labeled with class i, and C_j is the jth cluster in the clustering results.

F1-Measure. We firstly calculate the precision and recall with respect to the clustering results, and

$$F_1 = 2 * precision * recall/(precision + recall) \tag{27}$$

Rand Statistics. Rand statistics can be computed by

$$\text{RS} = \frac{2 * (N_{11} + N_{00})}{N(N-1)}, \tag{28}$$

where N_{11} is the number of data pairs that are in the same class both in the origin data set and the clustering outcome, N_{00} is the number that events pair are in the different clusters both in the origin data set and the clustering results, and N is the number of data points.

In Figure 3, we show the performances of different methods in identifying distinct trip purposes with regards to the above evaluation metrics. It is shown that the joint MHP model achieves the best performance among all the methods. Meanwhile, we also vary the number of latent trip purposes to show the performance. Obviously, the performance of the proposed model still performs best.

4.2 NYC Data

We also apply the model to a real world dataset, which consists of taxi trajectories and the POI information collected from 01/2015 to 06/2015 in New York City. In the raw data set, there are more than 70,000,000 trajectories from about 700,000 pickup points to more than 1,000,000 dropoff points. First, we divide the NYC map into regions which consists of a square with 0.008 longitude as width and 0.008 latitude as height, and we also filter out the obviously incorrect trajectories and points. Then, we retain the regions which

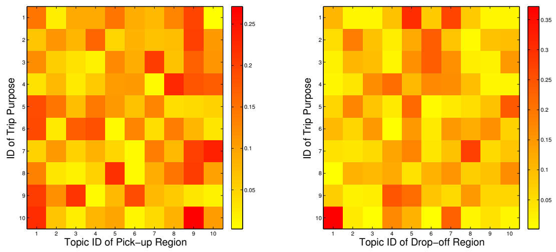


Figure 4: POI Topic distribution over latent trip purposes for origin and destination

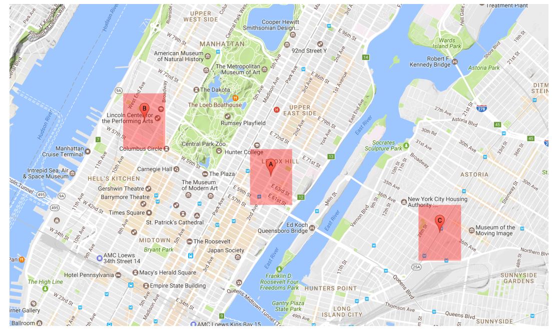


Figure 5: The locations of regions A, B, and C on the city map of NYC.

have at least 10 POIs and 36000 arrival events during the time period. Thus, we obtain about 55,000,000 taxi trajectories and 152 regions. In addition, the arrival events are sampled from the original dataset randomly, resulting in 31630 arrival events between the 152 regions in total.

Identifying POI Topics. In order to discover the underlying semantic meanings of POI Topics, we analyze the parameters β_r^c, and list the most representative POIs for each POI topic, which are shown in Table 1. We can see from Table 1 that each POI Topic is mixed with several related POIs but show particular urban functionality in overall. Specifically, we see representative POIs such as bar, nightclub and restaurant in Topic 1, which refers to a nightlife related POI Topic. Both Topic 2 and Topic 3 can be inferred as dining related topics, but there still exist some differences regarding the types of restaurants. Topic 4 is mixed with office, hotel and entertainment places, which corresponds to the fact that some regions in the city play the role of both working areas and scenic spots, *e.g.*, Time Square. Topic 5 indicates shopping related urban functions and Topic 6 is regarded as school oriented because it mostly consists of POIs concerning schools. Obviously, Topic 7 should be identified as a mixture of travelers' visited places including scenic spots, traffic centers, etc. Topic 8 can be viewed as a general entertainment function while Topic 9 consists of fast food related POIs, as well as offices; since a working place is usually blended with many fast food restaurants. Topic 10 refers to the residential areas with apartment buildings and dining restaurants.

Spotting Trip Purposes. As discussed previously, the trip purpose can be detected by integrating the regional environment of both the origins and destinations, which can be estimated from the POI Topic distribution over trip purposes. As shown in Figure 4, we can see that each trip purpose focuses on particular POI topics, and the purpose can be inferred in combination of the topics at both ends of the trips. Then, several interesting trip purposes are identified as follows:

Shopping purpose. In purpose 1, we see topics of the origins concentrate on Topic 1 and 9, while the destinations are mainly composed of Topic 5 and 7. Considering the urban functionality of both ends, such trips can be inferred as shopping and outdoors activities.

Dining purpose. Purpose 2 consists of office related topics (Topic 4 and 9) as the origins and the restaurant related topics (Topic 6) as destinations, thus such trips can be inferred as for dining purpose.

Homing purpose. In purpose 5, the topics of the destinations are composed of topic 10, which corresponds to residential areas, while the topics of the origins are much more diverse. Then, the trips annotated with purpose 5 can be regarded as homing from various regions in the city.

Synchronization Effect. We further use a case to illustrate the *synchronization effect* discovered from the model. In the joint MHP model, we introduce a synchronization effect to when the correlation index is bigger, the relationship between the pair of regions is bigger. We choose three regions A, B, and C and plot them on the map 5. Region A and B are located at the two sides of the Central Park, while region C locates in Brooklyn. By estimating the synchronization rate between these regions, we can obtain $\alpha_{AB} = 8.27066832, \alpha_{AC} = 0.00711464$, indicating that region A and B has a strong synchronization effect while A and C has weak connections in terms of arrival events. Then, we further analyze the POI distribution and the arrival sequences at these regions. As shown in Figure 6, the POI distribution of region A and B are very similar, which are mainly mixed with educational buildings, residential apartments, restaurants, and shops. However, the POIs in region C show different distribution with few shopping related POIs. In view of the arrival sequences, region A and B show extremely similar patterns, while region C has a different arrival patterns. This case illustrates that the synchronization effect is strong among the regions with similar urban functions.

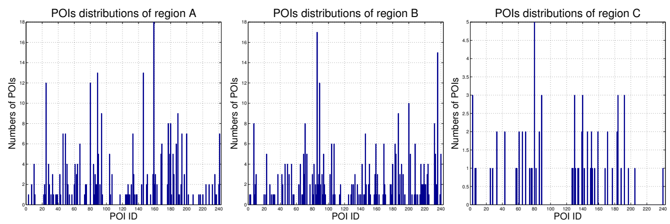
5 RELATED WORK

This paper is related to the following categories of prior work including discovering the urban functionality and trip purposes from human mobility data, traffic flow prediction, as well as the application of Hawkes Process in various fields.

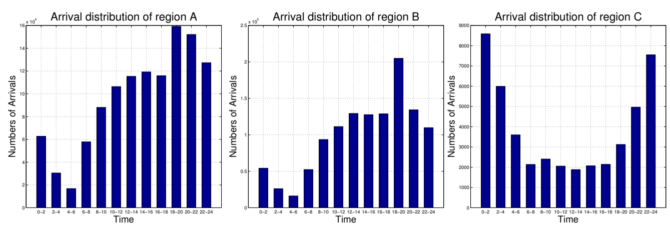
Urban functionality and trip purpose discovery. With the rapid development of mobile and GPS technologies, large-scale footstep data has been accumulated and attracted lots of studies in discovering functional blocks and trip purposes [36]. The work in [30] proposed a semantic model, along with a computation and annotation platform for developing a semantic approach to understand the trip purpose better. Rinku *et al.* [3] proposed a method to determine the destinations of trips using only speed and time data from real world driving trips without accessing GPS traces. The work in [8] proposed a method to automatically annotate raw trajectories with the activities of users during their movement when tracked by a GPS device. The work in [37] proposed an approach

Table 1: Representative POIs for POI Topics

TOPIC 1	prob.	TOPIC 2	prob.	TOPIC 3	prob.	TOPIC 4	prob.	TOPIC 5	prob.
Bar	0.1884	Chinese Rest.	0.1286	Bar	0.0933	Office	0.3331	Clothing Store	0.0995
Home	0.0953	Italian Rest.	0.0913	Italian Rest.	0.0565	General Entertain	0.1035	Cafe	0.0693
Nightclub	0.0571	Asian Rest.	0.0541	American Rest.	0.0442	Hotel	0.1023	Office	0.0574
Event Space	0.0495	Tea Room	0.0481	Wine Bar	0.0373	Building	0.0869	Coffee Shop	0.0535
Cocktail Bar	0.0495	Bar	0.0472	Sushi Rest.	0.0319	Event Space	0.0593	Cosmetics Shop	0.0419
Lounge	0.0495	Spa or Massage Parlor	0.0416	Mexican Rest.	0.0306	Sandwich Place	0.0376	General Entertain	0.0408
Speakeasy	0.0471	Salon or Barbershop	0.0403	Lounge	0.0297	Hotel Bar	0.0342	French Rest.	0.0406
Breakfast Spot	0.0382	Vietnamese Rest.	0.039	Pizza Place	0.0278	Lounge	0.0342	High Tech Outlet	0.0388
French Rest.	0.0334	Art Gallery	0.0342	Coffee Shop	0.0256	Other Outdoors	0.0298	Salon or Barbershop	0.0368
Boat or Ferry	0.0316	Cocktail Bar	0.0316	Salon or Barbershop	0.0256	Performing Arts Venue	0.0289	Miscellaneous Shop	0.0331
TOPIC 6	**prob.**	**TOPIC 7**	**prob.**	**TOPIC 8**	**prob.**	**TOPIC 9**	**prob.**	**TOPIC 10**	**prob.**
College Acad.	0.0808	Park	0.1343	Art Gallery	0.2773	American Rest.	0.1023	Home	0.2005
Food Truck	0.0756	Other Outdoors	0.1	Park	0.1021	Deli or Bodega	0.0619	Building	0.0591
University	0.0653	Scenic Lookout	0.0767	Other Outdoors	0.0892	Office	0.0569	Deli or Bodega	0.0471
College Library	0.0639	General Travel	0.0753	Cafe	0.0555	Pizza Place	0.0464	Pizza Place	0.0442
General College/University	0.0573	Building	0.074	Playground	0.049	Bar	0.0448	Laundromat or Dry Cleaner	0.0342
College Dorm	0.0565	Airport	0.074	Automotive Shop	0.0386	Food Truck	0.0434	Coffee Shop	0.0317
Cafe	0.0499	Harbor or Marina	0.0616	Event Space	0.033	Sandwich Place	0.0392	Drugstore or Pharmacy	0.0291
Plaza	0.0485	Taxi	0.0534	Strip Club	0.0265	Coffee Shop	0.0346	Chinese Rest.	0.0256
Park	0.0382	Government Building	0.048	Sculpture Garden	0.0241	Burger Joint	0.0326	Mexican Rest.	0.0236
College Classroom	0.0374	Seafood Rest.	0.0343	Plaza	0.0233	Cafe	0.0307	Apartment Building	0.0206



(a) POIs distribution in region A, B and C



(b) Arrival sequences in region A, B and C

Figure 6: Synchronization effect with different synchronization rates (α_{AB} = 8.27066832, α_{AC} = 0.00711464).

which focused on model and infer the purpose of travel, or the activity at the destination of a trip during daily life scenarios. The work in [18] constructed a effective recommendation system by the analysis of the historical check-in data. Specifically, they studied the exploration prediction problem and converted it into a classification problem and proposed a novel future location prediction model called CEPR to make a location prediction more efficiently. The work in [19] proposed a collective iterative classification algorithm to classify the purpose of passengers. The general idea of this work is to first form groups based on the extracted mobility features and collectively classify the purposes of group passengers sharing same destination and same purposes. The work in [28] used the historical mobility data records to spot user preferences to annotate venue to mobility records of mobile users.

Traffic flow prediction. Traffic flow prediction has been one of major challenges for developing intelligent transportation system [35]. [15] uses regime-switching space-time models and selection operator (LASSO) to predict the real-time road traffic and verifies the method on a real data set. [20] compares a few typical methods on forecasting traffic flow and then proposes a seasonal support vector machine function to improve the accuracy and computing efficiency. [11] uses a method, adaptive Kalman filter, to improve the performance of short term traffic flow forecasting. [13] constructs framework with a deep belief network at the bottom to achieve the effective features and a multi-task regression layer at the top. The work in [25] proposed four methods by using Kalman filtering theory to solve the multi-step ahead predicting of short-term traffic flow and evaluated the approaches with other prediction methods.

Applications of Hawkes Processes. Recently Hawkes process [12] has been widely used in many fields, including earthquake [22, 24], finance [5], crime [2], streaming texts [4, 31, 32, 34], social network [6, 9, 10, 14, 17], and marketing [29], etc. For example, [22] describes a generalized model which focuses on space-time clustering for historical earthquakes by Hawkes process. Also, [24] applies the Hawkes process for earthquake occurrences with clustering and interpretation. Yang et al [32] propose a generalized model to track the meme and the corresponding diffusion network by combining mixture of Hawkes processes and a language model. Li et al. [16] combine topic model with Hawkes process to simultaneously identify and label the searching tasks. [34] constructs a self-exciting point process model with the re-sharing information in the social network to predict the popularity of Tweets.

6 CONCLUSIONS

Summary. Modeling mobility arrivals and spotting trip purposes have been two key challenges in the semantic exploration of human mobility data. By studying large-scale human mobility data, we observe and identify the synchronization property of human mobility. Human mobility synchronization indeed jointly captures the correlations among functional regions, mobility arrivals, and trip purposes; that is, if two regions share similar spatial configurations and urban functions, given a particular time period, the two regions show arrival events with similar arrival rates with respect to specific trip purposes. In addition, we construct the latent factor POI topics of both the origin and destination to reflect the latent trip purpose. Along this line, we develop a probabilistic method via modeling human mobility data as stochastic point processes, by exploiting both human mobility synchronization and POI topics of origin and destination regions. This probabilistic model combines both a mixture of Hawkes point processes and a topic modeling where POIs are regarded as words and regions are regarded as documents. The model can effectively model mobility arrivals, spot trip purposes, and learn the spatiotemporal synchronization rates among region pairs through the variational inference. Finally, we present extensive experiments on synthetic data and real world human mobility data of NYC to demonstrate the effectiveness of the proposed model.

Acknowledgment. This research was partially supported by University of Missouri Research Board (proposal number: 4991). This work was also partially supported by the National Key Research Program of China under Grant No. 2016YFB1000600, No. 2016YFB0501900 and Natural Science Foundation of China under Grant No. 61402435.

REFERENCES

[1] David M Blei, Andrew Y Ng, and Michael I Jordan. 2003. Latent dirichlet allocation. *Journal of machine Learning research* 3, Jan (2003), 993–1022.

[2] Aaron Clauset, Ryan Woodard, and others. 2013. Estimating the historical and future probabilities of large terrorist events. *The Annals of Applied Statistics* 7, 4 (2013), 1838–1865.

[3] Rinku Dewri, Prasad Annadata, Wisam Eltarjaman, and Ramakrishna Thurimella. 2013. Inferring trip destinations from driving habits data. In *Proceedings of the 12th ACM workshop on Workshop on privacy in the electronic society*. ACM, 267–272.

[4] Nan Du, Mehrdad Farajtabar, Amr Ahmed, Alexander J Smola, and Le Song. 2015. Dirichlet-hawkes processes with applications to clustering continuous-time document streams. In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, 219–228.

[5] Paul Embrechts, Thomas Liniger, Lu Lin, and others. 2011. Multivariate Hawkes processes: an application to financial data. *Journal of Applied Probability* 48 (2011), 367–378.

[6] Eric W Fox, Martin B Short, Frederic P Schoenberg, Kathryn D Coronges, and Andrea L Bertozzi. 2016. Modeling e-mail networks and inferring leadership using self-exciting point processes. *J. Amer. Statist. Assoc.* just-accepted (2016).

[7] Yanjie Fu, Guannan Liu, Spiros Papadimitriou, Hui Xiong, Yong Ge, Hengshu Zhu, and Chen Zhu. 2015. Real estate ranking via mixed land-use latent models. In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, 299–308.

[8] Barbara Furletti, Paolo Cintia, Chiara Renso, and Laura Spinsanti. 2013. Inferring human activities from GPS tracks. (2013).

[9] Manuel Gomez-Rodriguez, Jure Leskovec, and Bernhard Schölkopf. 2013. Modeling Information Propagation with Survival Theory.. In *ICML (3)*. 666–674.

[10] Manuel Gomez Rodriguez, Jure Leskovec, and Bernhard Schölkopf. 2013. Structure and dynamics of information pathways in online media. In *Proceedings of the sixth ACM international conference on Web search and data mining*. ACM, 23–32.

[11] Jianhua Guo, Wei Huang, and Billy M. Williams. 2014. Adaptive Kalman filter approach for stochastic short-term traffic flow rate prediction and uncertainty quantification. *Transportation Research Part C Emerging Technologies* 43 (2014), 50–64.

[12] Alan G Hawkes. 1971. Spectra of some self-exciting and mutually exciting point processes. *Biometrika* 58, 1 (1971), 83–90.

[13] W. Huang, G. Song, H. Hong, and K. Xie. 2014. Deep Architecture for Traffic Flow Prediction: Deep Belief Networks With Multitask Learning. *IEEE Transactions on Intelligent Transportation Systems* 15, 5 (2014), 2191–2201.

[14] David Hunter, Padhraic Smyth, Duy Q Vu, and Arthur U Asuncion. 2011. Dynamic egocentric models for citation networks. In *Proceedings of the 28th International Conference on Machine Learning (ICML-11)*. 857–864.

[15] Yiannis Kamarianakis, Wei Shen, and Laura Wynter. 2012. Real-time road traffic forecasting using regime-switching space-time models and adaptive LASSO. *Applied Stochastic Models in Business and Industry* 28, 4 (2012), 297–315.

[16] Liangda Li, Hongbo Deng, Anlei Dong, Yi Chang, and Hongyuan Zha. 2014. Identifying and labeling search tasks via query-based hawkes processes. In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 731–740.

[17] Liangda Li and Hongyuan Zha. 2013. Dyadic event attribution in social networks with mixtures of hawkes processes. In *Proceedings of the 22nd ACM international conference on Conference on information & knowledge management*. ACM, 1667–1672.

[18] Defu Lian, Xing Xie, Vincent W Zheng, Nicholas Jing Yuan, Fuzheng Zhang, and Enhong Chen. 2015. CEPR: A collaborative exploration and periodically returning model for location prediction. *ACM Transactions on Intelligent Systems and Technology (TIST)* 6, 1 (2015), 8.

[19] Youfang Lin, Huaiyu Wan, Rui Jiang, Zhihao Wu, and Xuguang Jia. 2015. Inferring the travel purposes of passenger groups for better understanding of passengers. *Intelligent Transportation Systems, IEEE Transactions on* 16, 1 (2015), 235–243.

[20] Marco Lippi, Matteo Bertini, and Paolo Frasconi. 2013. Short-Term Traffic Flow Forecasting: An Experimental Comparison of Time-Series Analysis and Supervised Learning. *IEEE Transactions on Intelligent Transportation Systems* 14, 2 (2013), 871–882.

[21] Yu Liu, Zhengwei Sui, Chaogui Kang, and Yong Gao. 2014. Uncovering patterns of inter-urban trip and spatial interaction from social media check-in data. *PloS one* 9, 1 (2014), e86026.

[22] Fabio Musmeci and David Vere-Jones. 1992. A space-time clustering model for historical earthquakes. *Annals of the Institute of Statistical Mathematics* 44, 1 (1992), 1–11.

[23] Yosihiko Ogata. 1981. On Lewis' simulation method for point processes. *IEEE Transactions on Information Theory* 27, 1 (1981), 23–31.

[24] Yosihiko Ogata. 1998. Space-time point-process models for earthquake occurrences. *Annals of the Institute of Statistical Mathematics* 50, 2 (1998), 379–402.

[25] Luis Leon Ojeda, Alain Y Kibangou, and Carlos Canudas De Wit. 2013. Adaptive Kalman filtering for multi-step ahead traffic flow prediction. In *American Control Conference (ACC), 2013*. IEEE, 4724–4729.

[26] Meng Qu, Hengshu Zhu, Junming Liu, Guannan Liu, and Hui Xiong. 2014. A cost-effective recommender system for taxi drivers. In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 45–54.

[27] Zuchao Wang, Min Lu, Xiaoru Yuan, Junping Zhang, and Huub Van De Wetering. 2013. Visual traffic jam analysis based on trajectory data. *IEEE Transactions on Visualization and Computer Graphics* 19, 12 (2013), 2159–2168.

[28] Fei Wu and Zhenhui Li. 2016. Where did you go: Personalized annotation of mobility records. In *Proceedings of the 25th ACM International on Conference on Information and Knowledge Management*. ACM, 589–598.

[29] Lizhen Xu, Jason A Duan, and Andrew Whinston. 2014. Path to purchase: A mutually exciting point process model for online advertising and conversion. *Management Science* 60, 6 (2014), 1392–1412.

[30] Zhixian Yan, Dipanjan Chakraborty, Christine Parent, Stefano Spaccapietra, and Karl Aberer. 2013. Semantic trajectories: Mobility data computation and annotation. *ACM Transactions on Intelligent Systems and Technology (TIST)* 4, 3 (2013), 49.

[31] Jaewon Yang and Jure Leskovec. 2011. Patterns of temporal variation in online media. In *Proceedings of the fourth ACM international conference on Web search and data mining*. ACM, 177–186.

[32] Shuang-Hong Yang and Hongyuan Zha. 2013. Mixture of Mutually Exciting Processes for Viral Diffusion. *ICML (2)* 28 (2013), 1–9.

[33] Jing Yuan, Yu Zheng, and Xing Xie. 2012. Discovering regions of different functions in a city using human mobility and POIs. In *ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 186–194.

[34] Qingyuan Zhao, Murat A Erdogdu, Hera Y He, Anand Rajaraman, and Jure Leskovec. 2015. Seismic: A self-exciting point process model for predicting tweet popularity. In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, 1513–1522.

[35] Yu Zheng. 2015. Trajectory data mining: an overview. *ACM Transactions on Intelligent Systems and Technology (TIST)* 6, 3 (2015), 29.

[36] Yu Zheng, Licia Capra, Ouri Wolfson, and Hai Yang. 2014. Urban computing: concepts, methodologies, and applications. *ACM Transactions on Intelligent Systems and Technology (TIST)* 5, 3 (2014), 38.

[37] Zack Zhu, Ulf Blanke, and Gerhard Tröster. 2014. Inferring travel purpose from crowd-augmented human mobility data. In *Proceedings of the First International Conference on IoT in Urban Space*. ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering), 44–49.