

CityMomentum: An Online Approach for Crowd Behavior Prediction at a Citywide Level

Zipei Fan, Xuan Song, Ryosuke Shibasaki
Center for Spatial Information Science,
The University of Tokyo, Japan
fanzipei@iis.u-tokyo.ac.jp,
{songxuan, shiba}@csis.u-tokyo.ac.jp

Ryutaro Adachi
Zenrin DataCom Co'Ltd
Tokyo, Japan
r_adachi@zenrin-datacom.net

ABSTRACT

Human movements are difficult to predict, especially, when we consider rare behaviors that deviate from normal daily routines. By tracing the behavior of a person over a long period, we can model their daily routines and predict periodical behaviors, whereas rare behaviors, such as participating in the New Year's Eve countdown, can hardly be predicted readily and thus they have usually been treated as outliers of the daily routines in most existing studies. However, for scenarios such as emergency management or intelligent traffic regulation, we are more interested in rare behaviors than daily routines. Using human mobility Big Data, the rare behavior of each individual in a social crowd is no longer rare and thus it may be predicted when we analyze the crowd behavior at a citywide level. Therefore in this study, instead of predicting movement based on daily routines, we make short-term predictions based on the recent movement observations. We propose a novel model called CityMomentum as a predicting-by-clustering framework for sampling future movement using a mixture of multiple random Markov chains, each of which is a Naive Movement Predictive model trained with the movements of the subjects that belong to each cluster. We apply our approach to a big mobile phone GPS log dataset and predict the short-term future movements, especially during the Comiket 80 and New Year's Eve celebration. We evaluate our prediction by a Earth Mover Distance (EMD) based metric, and show our approach accurately predicts the crowd behavior during the rare crowd events, which makes an early crowd event warning and regulation possible in the emergent situations.

Author Keywords

human mobility; urban computing; emergency management; big data

ACM Classification Keywords

H.4 Information Systems Applications: Miscellaneous; H.2.8 Database Management: Database Applications - data mining, Spatial databases and GIS

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
UbiComp '15, September 7-11, 2015, Osaka, Japan.

Copyright is held by the owner/author(s). Publication rights licensed to ACM.
ACM 978-1-4503-3574-4/15/09...\$15.00.
<http://dx.doi.org/10.1145/2750858.2804277>

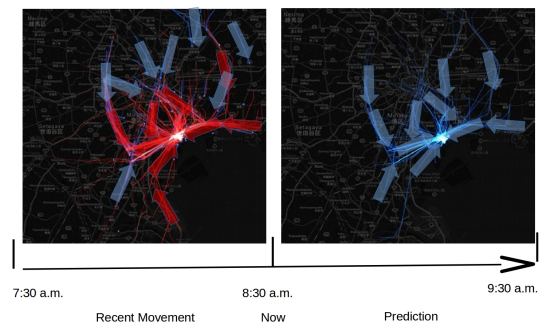


Figure 1. *During rare crowd event such as Comiket or New Year's Eve celebration, can we predict the human movement in next one hour? Our prediction model is based on recent and contemporary crowd movement. The gathering movements that trains the predictive model are shown in red and the potential observed movements of arrivals are shown in blue in the left panel. Our objective is to identify the potential arrivals and the routes they will take (right panel).

INTRODUCTION

For many years, building a smart city has long been a difficult but attractive objective of governments, companies, and researchers in a various areas. This problem is difficult because collecting, integrating and responding to human mobility data at a city-wide level in real-time was almost impossible before the popularization of cell phones and devices with a localization function (e.g. GPS-equipped taxis), which can be regarded as distributed localization sensors of for tracking human mobility.

Although it may be difficult, building a smart city is an attractive prospect because it could provide a digitalized framework that solves traditional urban engineering problems in a novel and more effective manner. Recently, an explosion of studies utilizing the data collected from localization devices has led to the development of a various novel and socially beneficial applications such as disaster emergency management [18], gas station recommendation [14], air condition prediction [22], and noise level prediction [23]. In addition, recent social tragedies have also led the researchers to develop new solutions to prevent their repetition. In particular, on New Year's Eve in Shanghai, a dazzling light show attracted a huge crowd to celebrate the arrival of 2015. However, the crowd

*We used Congestion Statistics (R) by ZDC co'Ltd. Congestion statistics is generated based on user agreement of GPS mobile phones. The statistics data is anonymized and aggregated so that individual information can never be retrieved. The anonymization and aggregation are made by NTT Docomo based on the request of Zenrin DataCom co'Ltd.

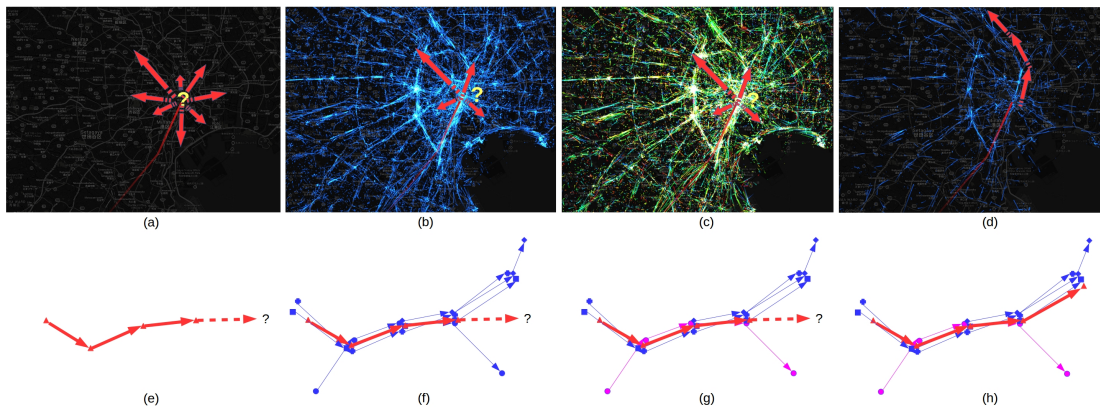


Figure 2. *Overview of the key processes in our CityMomentum model. We predict the future movement of each subject based on the recent movements of other subjects (b,f). To further improve the accuracy of the prediction, we partitioned the subjects movements into clusters (c,g) and predict future movements based on the best-matching cluster (d,h).

density was uncontrolled and a tragedy occurred, where 36 people died and 47 were injured in a stampede and the police admitted that they had underestimated the crowd's density [3]. Reconsidering the causes the tragedy led us to ask whether we can predict a crowd's density before it becomes too dangerous. In addition, can we take measures to prevent such tragedies from happening?

Human movements are extremely complex in the real world. Therefore, to compute these movements, we can simplify them by making the following assumptions:

- Social crowds gather gradually so we can predict the late arrivals based on the information provided by those who arrived early to attend gathering.
- Subjects who sharing similar recent trajectories will have a similar bifurcating pattern in the short-term future (the bifurcating behavior can be drawn from the same distribution).
- The bifurcating behavior can be assumed to be invariant during a short period.

Given these assumptions, we propose a novel online people movement prediction model, which integrates the GPS logs of mobile phone users based on the current time to predict their future movements at a citywide level. We refer to our predictive model as “CityMomentum”, where the physical term “momentum” describes the momentary motion of an object, because our predictive model captures the overall momentary people movements of people in a city. Figure 1 shows the recent movements of the attendees at a large crowd event, i.e., the Comiket 80 in red in the left panel. The model was trained based on recent attendees to distinguish potential new arrivals (blue in the left panel) and their predicted routes (blue in the right panel). We consider that obtaining correct predictions of movements for 1h would facilitate an early intervention, such as traffic regulation, to prevent a crowd event from becoming a tragedy.

CityMomentum is a novel human flow prediction model, which may facilitate smarter government by exploiting the

cell phone data to obtain better solutions than existing emergency management systems. The main contributions of this study are as follows:

- We propose a novel online subject clustering algorithm and a human movement prediction system based on a random walk sampler.
- The CityMomentum model is proposed to facilitate an early detection of the escalation of the potential crowd events as well as providing essential information to effectively regulate traffic growth in a emerging situation.
- We apply our method to a big GPS log dataset and validate our algorithm in real-world scenarios.

OVERVIEW

In this section, we explain the design of the probabilistic formulation of online clustering and predictive model construction. First, we introduce the big human mobility dataset used in this study. Next we give a mathematical description of the input data, the expected output, and some important intermediate concepts, which are described in a formal manner. We also provide an overview of the method employed in this study and we illustrate our algorithm by using a graphical model to explain each module in the algorithm and their relationships.

Methodology Overview

An overview of our proposed CityMomentum model is shown in Figure 2, which depicts an example based on our real-world dataset (Figure 2a-d) and an abstract illustration (the bottom row) to illustrate the key points in the design of human movement prediction. Based on observations of the recent trajectory of a subject, it is difficult to predict where they will move next (Figure 2a). However, after considering the movements of other subjects, we can make a prediction based on the last location of the subject and the destinations of others from the same location (Figure 2b, f). If we only make our prediction based on the last location of a subject, we would not fully exploit the trajectory of the subject. For example,

a subject arriving from the south-west is more likely to continue traveling to the north-east, but very unlikely to make a sudden turn and return to the south-west. Therefore, to encode the information related to the past trajectory, we cluster the subjects and refer to the cluster assignment as the latent variable so subjects within each cluster share similar trajectories (Figure 2c). By identifying the subject's cluster, we can reconsider the potential movement from the last location by considering the movements within the cluster into account, and thus the predicted movement based on a random walk procedure (Figure 2d).

The prediction of future movements are based on a bifurcating pattern derived from the early arrivals. As shown in the bottom row of Figure 2(e-h), the subjects' momentary movements in the cluster can be trained to obtain a current prediction model that predicts the branch a late arrival (blue) will select with a probability that is proportional to the number of choices made by early arrivals. Imagine that we are taking a bus and we want to predict how many people will get off at each station, then if we have little additional information, the best assumption would be proportional to the bus ahead (Figure 2f). Although the distribution of people who get off the bus may differ significantly between the morning and evening, the behavior are most probably similar between two consecutive buses. Moreover, this prediction can be improved further based on our assumption 2, by making the prediction conditional on where people board the bus (regarded as the latent cluster).

Data

To address the problem of real-world human mobility and to develop innovative applications that may allow governments to accelerate the promotion of smart cities, we collected an anonymous GPS log dataset from about 1.6 million real mobile-phone users in Japan over a three-year period (August 1, 2010 to July 31, 2013)*. The dataset stored about 30 billion GPS records with a total file size exceeding 1.5 terabytes using a Hadoop cluster. The data were collected by a mobile operator and a private company under an agreement with the mobile phone users.

By default, the positioning function in mobile phones is activated and uploaded about every 5 min. However, data acquisition is not stable and robust due to several factors, such as the loss of signal or battery power. In addition, the positioning function is also turned off automatically when no motion is detected from the mobile phone (e.g. the phone is placed evenly on the table). The characteristics of the dataset also cause handling problems, so we designed the CityMomentum model to address these challenges.

Preliminaries

Definition 1 (GPS log data): The original dataset can be described by a set of 4-tuple:

$$X = \{(u, \tau, lat, lon)\}$$

where u , τ , lat , and lon are the unique mobile phone user ID, time stamp, latitude and longitude of the GPS record respectively.

Definition 2 (Momentary movement): We denote X_t as the subset of the original GPS log dataset that contains all the records with time stamps ranging from $t - \Delta t$ to t :

$$X_t = \{\mathbf{x} \in X | t - \Delta t < \mathbf{x}.\tau \leq t\} \subset X$$

Definition 3 (Predicted movement): Similar to the definition of momentary movement, the predicted movement at time t is the collection of predicted records from t to $t + \Delta t$, which share the same data structure of GPS log data.

$$\hat{X}_t = \left\{ \left(\hat{u}, \hat{\tau}, \hat{lat}, \hat{lon} \right) \mid t < \hat{\tau} \leq t + \Delta t \right\}$$

Definition 4 (Displacement): We are concerned with the dynamic patterns of human movements, we define an operator T to extract the displacement segments of each subject in terms of momentary movement X_t .

$$T_u(X_t) = \{(\mathbf{x}_o, \mathbf{x}_d) \mid \mathbf{x}_o, \mathbf{x}_d \in X_t \wedge \mathbf{x}_o.u = \mathbf{x}_d.u = u \wedge C\}$$

where x_o and x_d denote the records of the origin and destination, respectively, and C is the consequential condition that guarantees each transition is defined based on two temporally consecutive records from the same subject.

$$C = (\mathbf{x}_o.\tau < \mathbf{x}_d.\tau) \wedge (\neg \exists \mathbf{x} \in X_t \rightarrow \mathbf{x}_o.\tau < \mathbf{x}.\tau < \mathbf{x}_d.\tau)$$

Definition 5 (Subject cluster): To maximize the predictability of human flow, we determine the clusters of mobile phone users with similar momentary movements at time t , which we denote as Z_t :

$$Z_t = \{z_u^t = k \mid u \in U_t, k \in \{1, \dots, K\}\}$$

where the allocation of the subject u to cluster k ($k = 1, \dots, K$) at time t is denoted as z_u^t . K is the predefined number of clusters.

Definition 6 (CityMomentum model): The CityMomentum model is the online updating predictive model of human movement, which gives the momentary movement that we observed and the latent cluster assignment of each subject. The future human movement is predicted by a sampling process that considers the conditional distribution, i.e. $p(\hat{X}_t | Z_t, X_t)$

Graphical Model Description

Our proposed model can be formulated as the a generative model described by a graph structure, as shown in Figure 3. The generative process comprises two main phases: online clustering and prediction phases. The Online clustering phase considers $\{X_t\}$, $\{Z_t\}$, and the links between them, where its structure is a first-order Markov model with latent variables, as used widely for modeling sequential data. The CityMomentum prediction model samples human movements in the short-term to construct each time interval t , which is conditioned on both the cluster assignments Z_t and momentary movement data X_t at time t only. Next, the joint probability

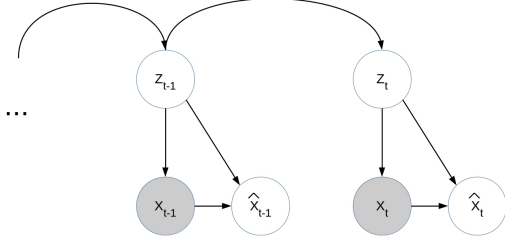


Figure 3. Graphical representation of the CityMomentum human movement prediction model. Shaded nodes denote the observed data and unshaded nodes denote latent variables.

can be decomposed into the product of conditional distributions as follows:

$$p(\hat{X}_{1:t}, X_{1:t}, Z_{1:t}) = \left(\prod_{\tau=1}^t p(X_\tau | Z_\tau) \right) \cdot \left(p(Z_1) \prod_{\tau=2}^t p(Z_\tau | Z_{\tau-1}) \right) \cdot \left(\prod_{\tau=1}^t p(\hat{X}_\tau | Z_\tau, X_\tau) \right) \quad (1)$$

Intuitively, we can separate the product into three modules by the brackets, the first and second of which characterize the movement similarity and online smoothing terms of the clustering phase respectively, while the last is the CityMomentum model that predicts human movements in the short-term future. In the following sections, we explain the estimation of the conditional distribution and inference of the latent variables.

NAIVE MOVEMENT PREDICTIVE (NMP) MODEL

In this section, we introduce a preliminary CityMomentum model. To address the limitations of the first-order Markovian model, CityMomentum can be regarded as an extension of the NMP model that includes the concept of online clustering and prediction using a mixing model of each cluster rather than a single naive predictive model.

Each subject is conditionally independent by giving the momentary movement. To determine the predicted movement for each subject, we utilize a random walk on a mobility graph constructed by X_t . The transition distribution is calculated as follows:

$$p(\hat{l}_{i+1}, \Delta\tau | \hat{l}_i, X_t) = \frac{n_{\hat{l}_{i+1}, \hat{l}_i}}{n_{:, \hat{l}_i}} \cdot \frac{n_{\Delta\tau, \hat{l}_i}}{n_{:, \hat{l}_i}} \quad (2)$$

where $n_{\hat{l}_{i+1}, \hat{l}_i}$ is the number of displacement from \hat{l}_i to \hat{l}_{i+1} , and $n_{\Delta\tau, \hat{l}_i}$ is the number of displacements starting from \hat{l}_i and with a time interval of $\Delta\tau$. The symbol “:” denotes the sum over, e.g. $n_{:, \hat{l}_i} = \sum_l n_{l, \hat{l}_i}$. Intuitively, the predicted location \hat{l}_{i+1} is drawn from the multinomial distribution which counts the occurrence of the end points of every displacement with the same start point. As shown in Figure 2f, the next displacement of the blue subject we sample should be with the probability of $\frac{3}{4} = 0.75$ upwards and $\frac{1}{4} = 0.25$ downwards if we do not take subject cluster into consideration. Similarly, we can draw the time intervals of each displacement from the

distribution based on the time interval statistics of all the displacements starting from \hat{l}_i .

However, the NMP model relies on the first-order Markovian assumption, which has the natural limitation of multi-step prediction. For example, people driving a car on the same road are likely to travel in the same direction, rather than making a sudden U-turn. However, a random walk under a first-order Markovian assumption would give an equal probability of moving forward and backward if the traffic flows are equivalent on both sides of the road.

Therefore, to maximize the predictability of human movements, we propose the CityMomentum model based on a mixing Markov chains instead of a single Markovian assumption. Each of the basic Markov chains is trained and it makes a prediction in the same manner as the NMP model, but X_t is partitioned into clusters $\{X_t^k\}$ to train the basic predictors separately. In the next section, we will describe the clustering algorithm that partitions X_t to maximize the predictability, before we revisit the prediction phase to extend the NMP model and propose our predicting-by-clustering CityMomentum model.

CITYMOMENTUM MODEL

To overcome the predictability limitations of the NMP model and to encode the momentary movement of each subject, we devise the CityMomentum model to make prediction of future movement conditioned of the latent cluster assignment of the subjects. The details of our model-based subject clustering algorithm is given in the first subsection and the predictive model using the intermediate results of clustering is introduced in the latter subsection.

Model-based Clustering

Model-based clustering [7] involves a widely used family of clustering algorithms, which assume that the data are drawn from a distribution that comprises a mixture of finite components. Each component includes a cluster of data, a model trained by the cluster of data, and the weight of each component. However, determining the optimal cluster assignment for each subject is a very complex problem. When using a Bayesian formulation to infer the latent variables, Monte Carlo Markov Chain (MCMC) has been proved to be a powerful technique for effectively estimating the posterior distribution of latent variables. In this study, we employ the Gibbs sampling method [8], which is recognized as the best variation of MCMC, to infer the cluster assignments of subjects.

In general, a Gibbs sampler iteratively takes out one variable before re-sampling from the posterior distribution by conditioning on all remaining variables. In this study, we define the posterior distribution of each component k after removing one subject u from U_t as:

$$p(z_u^t = k | Z_{t,-u}, Z_{t-1}, X_t) \propto p(X_{t,u} | z_u^t = k, X_{t,-u}^k) \cdot p(z_u^t = k | Z_{t,-u}, Z_{t-1}) \quad (3)$$

where $X_{t,u}$ is the momentary movement of subject u and $X_{t,-u}^k$ is the momentary movement of all the subjects that belong to cluster k without subject u . The first term in Eq. 3

indicates the likelihood function used to estimate the likelihood of subject u belonging to cluster k whereas the latter is the weight of component k . The estimates of these two terms are explained in detail in the following subsections.

Likelihood

If we recall our definition of the NMP model, it might appear that it can be adapted to an estimator of the likelihood function. However, given that the sample rate of our original dataset is not stable and the presence of fast-moving subjects in cars or trains, even small differences in velocity or events such as stopping at traffic lights could cause a significant geographical disparities with each other. To address this problem, in the design of our likelihood function, we consider the direction and speed of displacement to cluster the subjects with similar momentary movement.

$$p(\mathbf{x}_d | \mathbf{x}_o, Z_{t,-u}^k, X_{t,-u}^k) = \frac{n_{I(\mathbf{x}_o), D(\mathbf{x}_d, \mathbf{x}_o)}^k + \gamma}{n_{I(\mathbf{x}_o), :}^k + L_{I(\mathbf{x}_o)} \gamma} \quad (4)$$

where $I(x)$ is the index of the location of GPS record \mathbf{x} and $D(\mathbf{x}_d, \mathbf{x}_o)$ indicates the type of displacement. A hyper-parameter γ governs the sparsity of the model and L_i denotes the number of end point locations for all the displacements starting from location indexed by i . In our experiment, we categorize the displacements into three speed levels (low, medium, high) and four directions (north-east, north-west, south-west and south-east). Thus, by considering the combination of speed and direction, we can distinguish nine types of displacement (we do not consider the direction when the speed is low)

Thus, we construct a random Markov chain with the transition distribution defined in Eq. 4. Using a random Markov chain, the likelihood of each subject's trajectory given the momentary movement of cluster k is formulated as a random walk on the chain.

$$p(X_{t,u}^t | z_u^t = k, X_{t,-u}^k) \propto \prod_{(\mathbf{x}_o, \mathbf{x}_d) \in T_u(X_t)} p(\mathbf{x}_d | \mathbf{x}_o, Z_{t,-u}^k, X_{t,-u}^k) \quad (5)$$

Thus, we define the prior likelihood function for our clustering algorithm. In the next subsection, we describe how to estimate the weight of each component, or the prior distribution of each component in Bayesian view.

Prior Cluster Distribution

We define the prior cluster distribution as the prior distribution used to draw a cluster label k for every subject. A widely used prior distribution is the discrete Dirichlet distribution, which includes a hyper-parameter that describes prior knowledge of the prior distribution. To consider the temporal continuity, we utilize the cluster distribution at the previous time with a damping factor of α by adding up β as the hyper-parameter of the discrete Dirichlet distribution:

$$p(z_u^t = k | Z_{t,-u}, Z_{t-1}) = \frac{n_{t,-u}^k + \alpha n_{t-1}^k + \beta}{n_{t,-u} + \alpha n_{t-1} + \beta K} \quad (6)$$

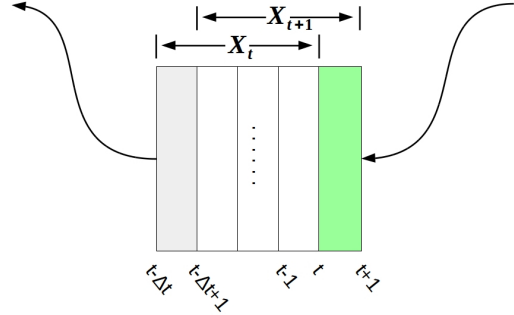


Figure 4. Online updating of the momentary movement model. From time t to $t+1$, the records in the leftmost slot (grey) are polled from the queue and those in the rightmost slot (green) are offered to the queue.

where $n_{t,-u}^k, n_{t,-u}$ is the number of subjects belong to cluster k and the total number of subjects at time t except subject u respectively.

Until now, we have characterized the Gibbs sampling based subject clustering process. A pseudo-code of the entire inference process is given in Algorithm 1. “InitializeModel” and “UpdateModel” in the pseudo-code are used to calculate or update (a response to adding or removing a subject from a cluster) the intermediate parameters, i.e. $n_{t,-u}^k, n_{t,-u}, n_{I(\mathbf{x}_o), D(\mathbf{x}_d, \mathbf{x}_o)}^k$ and $n_{I(\mathbf{x}_o), :}^k$.

We conclude this section with an interpretation of the “similarity” of the subjects that belong to the same cluster. We do not consider it is necessary for the movements of two subjects are to be “similar” if they share similar trajectories or they are geographical close to each other, but instead we interpret “similar” as less uncertainty resulting from flow bifurcation or a conflict with the movement of other subjects in the same cluster. We provide an illustrative interpretation using our experimental results in a later section.

Predicting-by-clustering

Next, we extend the NMP model to predict future movements from a mixing model based on the predictability-maximum partition of the X_t , i.e. $\{X_t^k\}$, rather than the entire X_t . Considering the conditional independence, we formulate our predicting-by-clustering distribution as:

$$p(\hat{X}_t | Z_t, X_t) = \prod_{u \in U_t} p(\hat{X}_{u,t} | X_t^k) \quad (7)$$

where $k = z_u^t$ is the cluster at time t to which subject u belongs, and X_t^k is the subset of the momentary movement at time t that contains all the GPS records of the subjects from cluster k . Therefore the CityMomentum model involves the mixing of NMP models, each of which is a random walk on the mobility graph constructed by X_t^k (see Eq. 2). The pseudo-code of the random walk sampling process is given in Algorithm 2.

CityMomentum is an online predictive model, where according to the definition of the momentary movement, X_t is a queue that buffers the GPS records in recent Δt , and which updates the queue as time elapses. As shown in Figure 4, we

Input: Subject clusters Z_{t-1} ; Momentary movement X_t
Output: Subject clusters Z_t
 /* Random initialization */
for u in U_t **do**
 $z_{u,t} = \text{RandInt}(K)$;
end
 InitializeModel (Z_t);
 /* Iterative random inference */
for $i = 1$ to ITERATION **do**
 for u in U_t **do**
 UpdateModel ($-u, z_{u,t}$);
 for k to K **do**
 Calculate *likelihood* from Eq. 5;
 Calculate *weight* from Eq. 6;
 $p(k) = \text{likelihood} \cdot \text{weight}$
 end
 Draw $z_{u,t}$ from the distribution $p(k)$;
 UpdateModel ($+u, z_{u,t}$);
 end
end

Algorithm 1: Gibbs Sampling Inference

Input: Subject Clusters Z_t , Momentary movement X_t
Output: Predicted movement \hat{X}_t
 Calculate $n_{\hat{d},o}^k, n_{\Delta\tau}^k, n_{:,o}^k$ for all $k, \Delta\tau, o, \hat{d}$;
 $\hat{X}_t = \emptyset$;
for u in U_t **do**
 $k = z_{u,t}$;
 $\hat{o} = \text{LastLocation}(X_t^u)$;
 $\hat{\tau} = t$;
 while $\hat{\tau} < t + \Delta t$ **do**
 Draw $\Delta\hat{\tau}, \hat{d}$ from Eq. 2 (conditioned on X_t^k);
 $\hat{\tau} = \hat{\tau} + \Delta\hat{\tau}$;
 $\hat{o} = \hat{d}$;
 $\hat{X}_t = \hat{X}_t \cup (u, \hat{\tau}, \hat{d}.lat, \hat{d}.lon)$;
 end
end

Algorithm 2: Random walk sampling predicted people movement

update the momentary movement queue from time t to $t + 1$ by offering the records during the time period from t to $t + 1$ and poll the out-of-date records during the time period from $t - \Delta t$ to $t - \Delta t + 1$. Accordingly, X_t is updated to X_{t+1} , such that Z_{t+1} can be inferred and the predicted movement \hat{X}_{t+1} follows naturally.

EXPERIMENTAL RESULTS

In this section, we present the results of extensive experiments performed using Congestion Statistics (R) by ZDC co'ld*. We evaluate our algorithm by both a qualitatively based on selected big crowd events in Tokyo and quantitatively over a few days before and after the event to verify its robustness. We select the New Year countdown and the Comiket (a major comics market festival in Tokyo) as big crowd events because the people movements are difficult to predict at these events

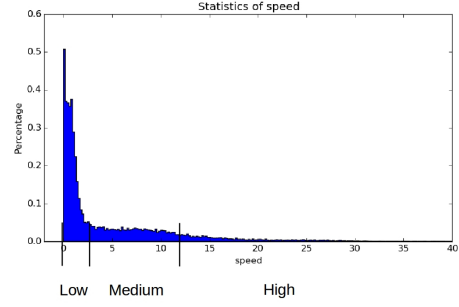


Figure 5. Determination of the thresholds to categorize the speed using the histogram for speed of all displacements.

using existing methods.

Experimental Setup

For simplicity, the time unit use in this study was 10 min, i.e. the momentary movements are updated every 10 min. Accordingly, Δt is set to 6 and thus we cache the momentary movements each hour to predict the movements in the next hour.

We crop a subset of our data from the Tokyo area in the range of $35.5 \leq latitude \leq 35.8$ and $139.4 \leq longitude \leq 139.9$ and discretize the coordinates by $(\text{floor}(latitude/0.004), \text{floor}(longitude/0.005))$, with the grid size of about $450m \times 450m$.

The Speed is categorized as “Low” if the speed is below $3m/s$, “Medium” if the speed is $3m/s$ to $12m/s$, and “High” if the speed exceeds $12m/s$. As shown in Figure 5, these thresholds are determined based on a statistical analysis of the speed of each movement.

As constants, we set the number of clusters $K = 10$, the damping factor $\alpha = 0.01$, $\beta = 1000$ and $\gamma = 1$.

Clustering Results

In this subsection, we present the intermediate results of online subject clustering, which are helpful for understanding how movements are predicted using our CityMomentum model. In Figure 6, the first row shows the movements for two representative clusters. The “shooting-star” trail of each target indicates the directions and speed of movement. To provide a more intuitive visualization, we also emphasize the mainstreams of human flow with arrows. Using our clustering algorithm, we group the subjects by minimizing the flow bifurcation and conflict uncertainty. It is assume that if we only know the initial location and the cluster to which a subject belongs, then we can determine their possible trajectory with high confidence based on the trajectories of other subjects in the cluster. Therefore, each bifurcation and conflict in the human flow would confound the possible next step. As shown in the top row in Figure 6, for each cluster, the occurrence of flow bifurcation and conflict is reduced greatly after clustering such that whenever a location is designated as the start point of a new target, we can predict the future movement by following the human flow without much confusion.

In the second row, we provide another view of our clustering



Figure 6. *The top row shows the visualization of two examples of human movements clusters. For a more intuitive view, we emphasize the dominating mainstreams of human movement by arrows. The bottom row tracks the subjects within a rectangular region (around Tokyo station) at 8 a.m. as well as showing their locations 20 min before and 20 min later. Different clusters are labeled with different colors, ranging from blue to red.

results. We select a rectangular region at 8:00 a.m. around Tokyo station and track the subjects in this region to show where they were located 20 min before and where they will go 20 min later, where we distinguish each cluster by different colors. As shown in Figure 6, the cluster of subjects labeled with dark blue came mainly from the north before passing through Tokyo station and heading south-west, whereas subjects labeled in pale blue come from similar sources but they are distributed around Tokyo station after 20 min. The subjects who come from the south-west, and head to northern areas are colored red, yellow and orange.

Note that the geographical differences among clusters guarantee the predictability of our CityMomentum model. If we apply the NMP model to predict future movements, which is a degenerate version of the CityMomentum model with a cluster number of 1, and we assume that a subject is at Tokyo station, then their future movement distribution is highly scattered area as shown in the bottom-right of Figure 6 (regardless of the color). However, after considering the clustering results, the possible destinations of the subject are narrowed down and more concentrated. For example, for a subject at Tokyo station in the dark blue cluster, their most likely destination is only in the south-west as shown in the figure. This is very important because actual applications require accurate predictions of the future density of crowds, but they also require early profiling by an intelligent surveillance system that to provide warnings about crowd event as well as the profiles

of the participants in a crowd and their most likely route in order to push congestion warnings or recommend other route to potential arrivals and regulate traffic.

Prediction Results

In this subsection, we present the results of experiments where we apply our CityMomentum model to the real-world GPS log dataset in an online updating process using 1 h observations of movements to predict the movements in the next 1 h. Apart from some rare events such as the Comiket (Comic market), which is held twice each year and that attracts a tremendous number of visitors, there are very few people around Tokyo Big Sight otherwise (e.g. the day before Comiket). Figure 8 shows the obvious difference in the population density at 10 a.m. around Tokyo Big Sight on the day before the Comiket 80 (C80) and the first day of C80 (August 12, 2011, i.e. Friday, a workday). For such crowd event, Human movements are very difficult to predict because little prior knowledge would work in this situation. Our model is designed to address this challenge so we provide qualitative and quantitative analysis of two examples, i.e. C80 and the New Year's Eve celebrations as follows.

Comiket 80

Comiket is the biggest comic fair in the world and it attracts about 590,000 visitors, which is more than the population of Macau and 2.5 times the attendance at Summer Sonic [10], the most famous Rock festival in Japan. The fair opened at

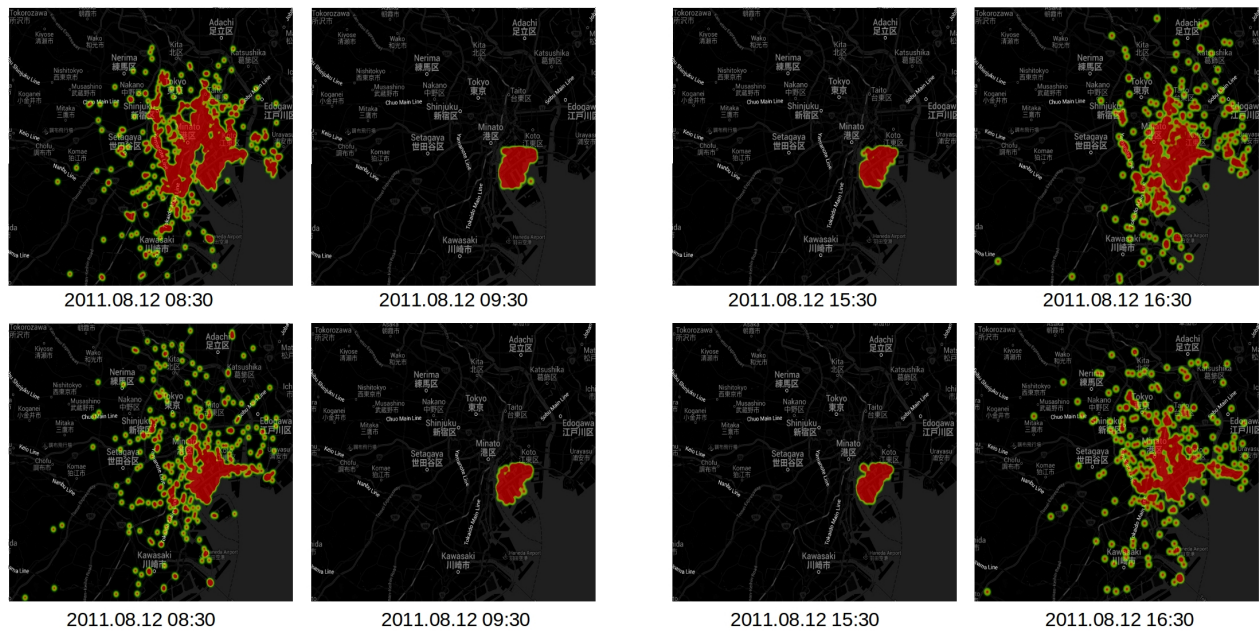


Figure 7. *Predicted sources of visitors to C80 at 1 h later and predicted 1-h-later locations of the visitors at C80 (in the top row) and the actual sources of and 1-h-later locations of visitors in reality (in the bottom row)

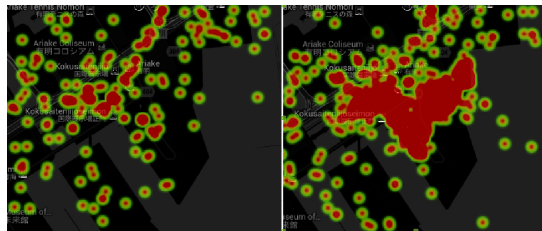


Figure 8. *Population density around Tokyo Big Sight at 10 a.m. on the day before C80 (left) and on the first day of C80 (right).

10 a.m and closed at 4 p.m. which was a big challenge for the local transportation system. The left panel of Figure 7 shows the predicted human movements between 8:30 and 9:30 (top row) based on the momentary movements at 8:30 and a comparison with the ground truth (bottom row). To visualize our prediction of the potential visitors to C80, we select the region around Big Sight at 9:30 and trace back the subjects in this region to find out their locations at 8:30. Our algorithm successfully predict the geographical distribution of the visitors to C80. The bottom row shows that the main sources of the visitors are the two railways to the north-west of Big Sight. These two railways are the most convenient way for reaching Big Sight and they pass through the most important transport hubs in Tokyo (Shinjuku, Tokyo, and Shibuya stations). The results shown in the top row illustrate that most of our predictions agreed with ground truth movements. However, our prediction overestimate the arrivals from the two railways to the north-west. This is because we make the prediction based on the observation from 7:30 to 8:30, so the potential for movement is exaggerated because the rush hour has already passed, and thus the stations (especially around Tokyo station) have already been transformed from a trans-

port hub into an office area. This limitation will be further discussed in the Conclusion section.

At 4 p.m., C80 closed and a large crowd of people were leaving from the Big Sight. In fact, most of the visitors tended to leave earlier because the majority of the best-sellers sold out within the first few hours. As a result, we can obtain some insights into the dispersion behavior of the visitors based on those early leavers. The right panel of Figure 7 provides an intuitive verification of the predicted of dispersion behavior based on a comparison between the predicted destinations of C80 visitors and their actual destinations.

New Year Eve Celebration

There are a number of celebrations of New Year Countdown celebrations in Tokyo, such as Disney Land New Year party, Shibuya square countdown and “Hatsumode” (the first visit in Buddhist temple or shrine) in Harajuku. Besides, most of the railway lines operate overnight on the New Year’s Eve, thus makes it more difficult to predict the human movements. To quantitatively evaluate the prediction results, as shown in Table 1, we select a few most famous places for New Year’s Eve countdown, and compare the predicted number of subjects in our dataset compared with ground truth.

Quantitative Evaluation

This subsection offers a detailed quantitative evaluation of our algorithm. In particular, we discuss about the metrics we use and then quantitatively evaluate the performance of CityMomentum human mobility predictor, compared with the NMP model.

Metric

A good human mobility predictor should be intended to simulate the *movement patterns* as similar as in the real data, rather

Location	Number of subjects(predicted)	Number of subjects(ground truth)
Disneyland Resort	1540	1558
Shibuya	256	366
Harajuku	185	168
Zojoji	301	321
Asakusa	256	263
Odaiba	151	143

Table 1. *Predicted number of subjects (as of in our dataset) compared with ground truth at 0:00 on New Year’s Eve.

than simply the population density. For example, considering two places A and B, if 100 subjects move from A to B while 100 subjects move from B to A at the same time, although the population density does not change, the movement patterns is significantly different from no movement switching between A and B. Therefore, to assess the performance of our approach, inspired by [11], we adapt Earth Mover Distance (EMD) [13] and define an more sophisticated metric to characterize the distance between movement patterns we predicted and the ground truth.

By tracking the subjects U_t^o at time t within region o , we can obtain the spatial distribution $p(\mathbf{d}|t + \Delta t, U_t^o)$ and $\hat{p}(\hat{\mathbf{d}}|t + \Delta t, U_t^o)$ of these tracked subjects at time $t + \Delta t$ from real data (used as the ground truth) and our prediction results respectively, where \mathbf{d} is the discrete random variable for the potential locations (grid cells) for the subject at time $t + \Delta t$. We denote the distance of the movement patterns between our prediction and the ground truth all over the city at time t as $CityEMD(t)$, and define it as the sum of the distance between $p(\mathbf{d}|t + \Delta t, U_t^o)$ and $\hat{p}(\hat{\mathbf{d}}|t + \Delta t, U_t^o)$ with respect to all the regions o at time t :

$$CityEMD(t) = \sum_o dist\left(p(\mathbf{d}|t + \Delta t, U_t^o), \hat{p}(\hat{\mathbf{d}}|t + \Delta t, U_t^o)\right) \quad (8)$$

where $dist(*, *)$ is the distance between two discrete distributions. To take the spatial distance among the grid cells into account, EMD turns out to be the most widely used and state-of-the-art histogram comparison measure which allows the probability to flow among the bins rather than a simply bin-by-bin comparison. Therefore, EMD is a powerful metric which well reflects the movement pattern dissimilarity caused by geographical shift.

Experimentally, we use the source code of fastEMD from [13], and defines the distance matrix as the geographical distance (km) between grid cells and the threshold distance as 10 km.

Overall Evaluation

Fig. 9 shows the logarithmic of the $CityEMD$ in Tokyo during Comiket 80 (one day before Comiket 80 and the three days fair) and New Year’s Eve (Dec. 31th and Jan. 1th). We run a dummy predictor(D) which assumes no subjects move after time t to indicate the intensity of subjects movement

(from the definition of EMD, we can easily conclude that the dummy predictor would have a higher $CityEMD$ when people movement becomes intenser and vice versa). In addition, the $CityEMD$ from dummy predictor gives an intuition of to what degree our prediction methods can cover the subjects spatial variance between time t and $t + \Delta t$ induced by the subjects movement (The area under dummy predictor indicates the variance in total while the area between dummy predictor and our method indicates the variance covered).

In the legend of the Fig. 9, the curves are denoted by “ $\Delta t +$ method”, where “C” stands for CityMomentum model, “N” for NMP and “D” for dummy predictor. For example, the prediction of 40 min later time t using CityMomentum model is named as “40C”. Note that the $CityEMD$ from dummy predictor is about 10^2 larger than our predictors, we use a logarithmic axis for $CityEMD$ to have a more clear performance comparison. As the results show, our CityMomentum model outperforms the NMP model and significantly cover the overall population variance at each time t .

Regional Evaluation for C80

During a big crowd event, the regional prediction accuracy at where the event takes place would be more crucial to the real world applications. A simple extension to Eq. 8 would help us gauge the the regional prediction accuracy:

$$RegionEMD(t) = \sum_{o \in R} dist\left(p(\mathbf{d}|t + \Delta t, U_t^o), \hat{p}(\hat{\mathbf{d}}|t + \Delta t, U_t^o)\right) \quad (9)$$

where R is the subset of all locations that within the region of interest. Fig. 10 shows the regional evaluation from one day before C80 to the end of this fair. We can see that our CityMomentum model still outperforms the NMP model at the same time successfully predict the overall population variance at each time t .

RELATED WORKS

Urban Computing: GPS log data from mobile phones [17, 5, 18], location-based online social networking data [1, 9], and Bluetooth sensed data [20] have emerged and are increasing explosively. The emergence of a wide-range of such big data has shed light on a wide-range of social problems that could hardly be solved in a traditional approach, e.g. disaster evacuation [17], taxi sharing [12] and gas consumption [14]. A comprehensive review written by Y. Zheng [21] provides an excellent overview of the recent researches, methodologies and challenges of urban computing using human mobility data.

Human Mobility Prediction: Human mobility is an attractive research topic in the fields of traffic engineering [2], behavioral sciences [4] and computer vision field [24]. Particularly, at a citywide level, an accurate prediction of human mobility is very important for the government for an early preparation of any emergent situation. J. Zheng [19] proposes an unsupervised learning algorithm that discovers the latent patterns of daily routines and predicts the people’s location based on the model. C. Song [15] explores the upper bound of the predictability of the human mobility using call detail

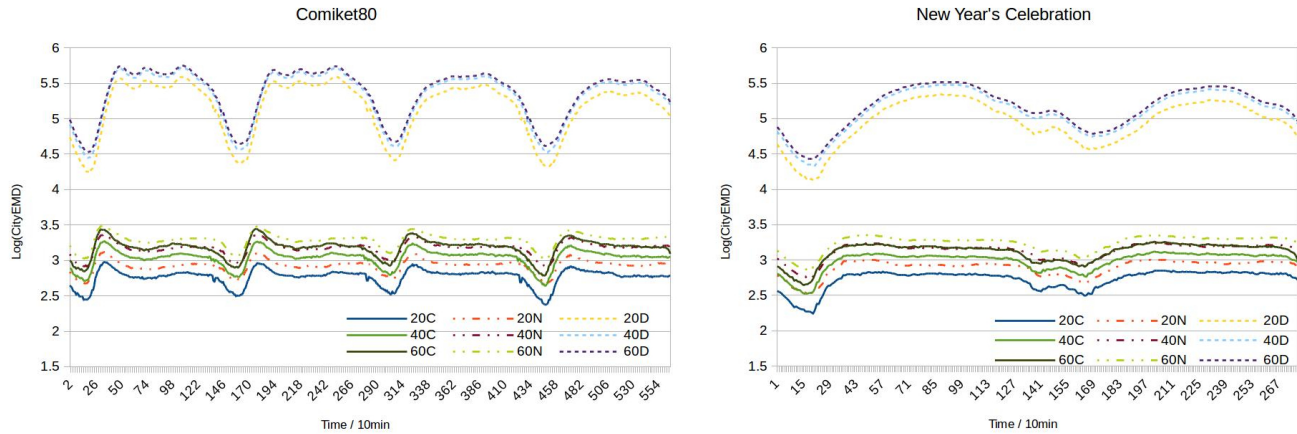


Figure 9. Evaluation on our prediction for human movements during Comiket 80 (left) and during New Year's Eve celebration (right).

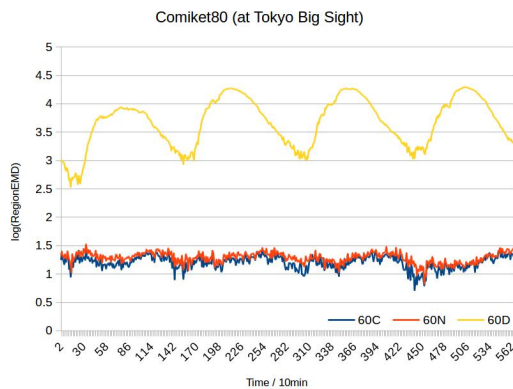


Figure 10. Regional evaluation on our prediction for human movements during C80.

records based on a long period of observations. However, although most of the behaviors are predictable by referring their daily routines, human movements during rare crowd events can hardly be predicted. X. Song [17] predicts the evacuation behavior after earthquake. In this paper, to model the rare crowd behaviors such as C80 gathering and dispersion, our CityMomentum model captures the momentary patterns of movement that based on the recent observations rather than a long period.

Trajectory Clustering: Trajectory cluster is an effective tool to analyze the semantic scene through moving targets. [24, 16] construct online-learned dynamic model to predict short-term prediction of future movement. However, the movements they study are sensed by video camera or laser scanner, where the scope is much smaller ($10^1 \sim 10^2$ m level) and the scene is relatively simpler comparing with our work, which predicts at a citywide level (10^4 m level). A relevant and interesting study is conducted by N. Ferreira [6], who considers the crowd movements as a mixture of vector field, and develops an interesting crowd movement visualization algorithm that discovers and colors each vector field. However, in our study, to maximize the predictability rather than visualize

the movement, we propose our clustering algorithm for the subjects based on the minimizing the movement bifurcation and conflict in each cluster.

CONCLUSION

In this study, to model rare crowd behavior, we propose a predicting-by-clustering framework for simulating a short-term future human movement. We validate our approach on the real-world data, especially on the New Year Countdown celebration and C80.

We note several limitations of our work. First, a more sophisticated preprocessing will be included in our further research. Particularly, we will apply map matching algorithm to interpolate the trajectories and improve the clustering results. Moreover, we predict future movements based on the current transition pattern (Assumption 3), but it may deviate from the movements in reality if the transition pattern changes suddenly. Some sudden changes, such as the beginning and ending of rush hours in Tokyo, we could improve our prediction by a heterogeneous prediction model considering both the momentary movements and habitual movements. However, when a baseball match is over, the nearest station or bus stop will suffer a sudden increase pressure, which can hardly be predicted by our model, nor by any model that only considers the movements. Thus to overcome this limitation, we could extend our model by fusing an additional information source such as twitter data or the match schedule data.

ACKNOWLEDGEMENT

This work was partially supported by, JST, Strategic International Collaborative Research Program (SICORP); Grant-in-Aid for Young Scientists (26730113) of Japan's Ministry of Education, Culture, Sports, Science, and Technology (MEXT); Microsoft Research collaborative research (CORE) program; and DIAS/GRENE project of MEXT. We specially thank ZENRIN DataCom CO., LTD for their supporting.

REFERENCES

1. Allamanis, M., Scellato, S., and Mascolo, C. Evolution of a location-based online social network: analysis and

- models. In *Internet Measurement Conference* (2012), 145–158.
2. Ashok, K., and Ben-Akiva, M. E. Estimation and prediction of time-dependent origin-destination flows with a stochastic mapping to path flows and link flows. *Transportation Science* 36, 2 (May 2002), 184–198.
 3. CNN. 'I failed to protect you' – Details emerge of victims in deadly Shanghai stampede.
<http://edition.cnn.com/2015/01/02/world/asia/china-shanghai-new-years-stampede/>.
 4. Eagle, N., and Pentland, A. Eigenbehaviors: Identifying structure in routine. Tech. rep., IN PROC. OF UBICOMP06, 2006.
 5. Fan, Z., Song, X., and Shibasaki, R. Cityspectrum: A non-negative tensor factorization approach. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing, UbiComp '14*, ACM (New York, NY, USA, 2014), 213–223.
 6. Ferreira, N., Klosowski, J. T., Scheidegger, C. E., and Silva, C. T. Vector field k-means: Clustering trajectories by fitting multiple vector fields. *CoRR abs/1208.5801* (2012).
 7. Fraley, C., and Raftery, A. E. Model-based clustering, discriminant analysis, and density estimation. *JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION* 97 (2000), 611–631.
 8. Gelfand, A., and Smith, A. Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association* 85, 410 (1990), 398–409.
 9. Han, J., Xie, X., and Woo, W. Context-based microblog browsing for mobile users. *JAISE* 5, 1 (2013), 89–104.
 10. Infogr. How big is Comiket.
<https://infogr.am/how-big-is-comiket>.
 11. Isaacman, S., Becker, R., Cáceres, R., Martonosi, M., Rowland, J., Varshavsky, A., and Willinger, W. Human mobility modeling at metropolitan scales. In *Proceedings of the 10th International Conference on Mobile Systems, Applications, and Services, MobiSys '12*, ACM (New York, NY, USA, 2012), 239–252.
 12. Ma, S., Zheng, Y., and Wolfson, O. Real-time city-scale taxi ridesharing. *IEEE Transaction on Knowledge Discovery and Data Engineering* (June 2015).
 13. Pele, O., and Werman, M. Fast and robust earth mover's distances. In *ICCV* (2009).
 14. Shang, J., Zheng, Y., Tong, W., Chang, E., and Yu, Y. Inferring gas consumption and pollution emission of vehicles throughout a city. In *KDD 2014*, ACM (August 2014).
 15. Song, C., Qu, Z., Blumm, N., and Barabsi, A.-L. Limits of predictability in human mobility. *Science* 327, 5968 (2010), 1018–1021.
 16. Song, X., Shao, X., Zhao, H., Cui, J., Shibasaki, R., and Zha, H. An online approach: Learning-semantic-scene-by-tracking and tracking-by-learning-semantic-scene. In *The Twenty-Third IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2010, San Francisco, CA, USA, 13-18 June 2010* (2010), 739–746.
 17. Song, X., Zhang, Q., Sekimoto, Y., Horanont, T., Ueyama, S., and Shibasaki, R. Modeling and probabilistic reasoning of population evacuation during large-scale disaster. In *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining*, ACM (2013), 1231–1239.
 18. Song, X., Zhang, Q., Sekimoto, Y., and Shibasaki, R. Prediction of human emergency behavior and their mobility following large-scale disaster. In *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '14*, ACM (New York, NY, USA, 2014), 5–14.
 19. Zheng, J., and Ni, L. M. An unsupervised framework for sensing individual and cluster behavior patterns from human mobile data. In *The 2012 ACM Conference on Ubiquitous Computing, UbiComp '12, Pittsburgh, PA, USA, September 5-8, 2012* (2012), 153–162.
 20. Zheng, J., and Ni, L. M. An unsupervised learning approach to social circles detection in ego bluetooth proximity network. In *Proceedings of the 2013 ACM International Joint Conference on Pervasive and Ubiquitous Computing, UbiComp '13*, ACM (New York, NY, USA, 2013), 721–724.
 21. Zheng, Y., Capra, L., Wolfson, O., and Yang, H. Urban computing: Concepts, methodologies, and applications. *ACM Transaction on Intelligent Systems and Technology* (October 2014).
 22. Zheng, Y., Liu, F., and Hsieh, H.-P. U-air: When urban air quality inference meets big data. In *KDD 2013*, ACM (August 2013).
 23. Zheng, Y., Liu, T., Wang, Y., Liu, Y., and Zhu, Y. Diagnosing new york city's noises with ubiquitous data. In *UbiComp 2014*, ACM (September 2014).
 24. Zhou, B., Tang, X., and Wang, X. Learning collective crowd behaviors with dynamic pedestrian-agents. *International Journal of Computer Vision* 111, 1 (2015), 50–68.