AM-PM: Travel Demand Nowcasting

Anonymous
Anonymous

***Abstract*—Predictive models of urban mobility can help alleviate traffic congestion problems in future cities. State-of-the-art in travel demand forecasting is mainly concerned with long and very short term models. Long term forecasts aim at urban infrastructure planning, while short term predictions typically use high-resolution freeway detector/camera data to project traffic conditions in the near future. In this paper, we present a medium term travel demand forecast framework (several hours to days ahead). Our approach is designed to use cellular data that are collected passively, continuously and in real time to predict the intended travel plans of individual travelers.**

To do so we developed a variety of generative sequence learning methods to infer activity models from cellular data. Experimental results show that input-output hidden Markov models (IO-HMMs) used in a semi-supervised manner perform well for location prediction while long short term memory models (LSTMs) are better at predicting temporal day structure patterns thanks to their continuous hidden state space and ability to learn long term dependencies. We validated our predictions by comparing predicted versus observed (1) individual activity sequences; (2) aggregated activity and travel demand; and (3) resulting traffic flows on road networks via a hyper-realistic microsimulation of the predicted travel itineraries.

I. Introduction

Travel demand forecast has been an integral part of most Intelligent Transportation Systems research and applications [38]. Long term forecast provides the basis for transportation planning and scenario evaluation. For example, transportation planers may need to answer the question of: how many people will be affected if a new subway line is introduced? How will travel patterns be changed if a major bridge is upgraded? These studies typically use data collected from travel surveys that are infrequent, expensive, and reflect changes in transportation only after significant delays.

On the other hand, short term prediction studies traffic conditions in a transportation network based on its past behavior, which is critical for many applications such as travel time estimation, real time routing, etc. These studies use high-resolution data, usually collected from sensors and detectors on freeways. However, one main concern is that these studies are limited to regions where high-resolution data is available. Moreover, such forecasts can only inform local operations such as adapting traffic light timing in response to growing queues.

One missing element of comprehensive transportation systems optimization frameworks is medium term forecasting, which, for example, could answer the question: based on observations of early morning or noon traffic, what will traffic be like during the evening commute? This could be a critical piece of knowledge used in the design of demand-responsive congestion mitigation interventions. In this paper, we propose a medium term travel demand forecasting framework to fill this gap. The idea is that given a large volume of partially observed user traces derived from cellular data available at different times of day (e.g., 3:00 am, 9:00 am, 3:00 pm, etc.), we complete the individual daily activity sequences for the remaining period with pre-trained generative mobility models. Cellular data are collected non-invasively, for a large proportion of the population and continuously in time. It is also rich spatially compared to detector/camera data, the main data source for short term prediction, as cellular data is ubiquitous and not limited to freeways or arterials.

To validate the predictions, we compare (1) at individual level: the discrepancies (e.g. differences in number of activities, travel distance, Hamming distance, etc.) between predicted sequences and ground truth sequences (observed by the end of a day) per individual; (2) at aggregated level: the hourly travel demand - number of activities, travel distances from all users; and (3) the resulting traffic volumes on all the major freeways within the region of study from predicted sequences and ground truth sequences. Results proves that we can improve the medium term travel demand forecast by incorporating observed information by the time of prediction. The mean absolute percentage error can be less than 5% one hour ahead and around 10% three hours ahead for the regional road network.

The main contributions of this paper lie in three aspects:

- We proposed and solved a medium term travel demand forecast framework which fills the gap between mainstreams of long term travel demand forecast and short term traffic state prediction.
- We improved and compared the state-of-the-art deep generative urban mobility models. Lessons learned from training different types of urban mobility models are summarized for future researchers.
- We explored the predictability of human mobility with parametric sequence learning models as related to using individualized non-parametric "nearest neighbor" approach.

The remainder of this paper is organized as follows. Urban mobility models are reviewed in Section II-A. Section III depicts the framework of medium term travel demand forecast. Section IV improves the state-of-the-art deep generative urban mobility models using co-training input-output hidden Markov models (IO-HMM) (Section IV-A) and long short term memory (LSTM) (Section IV-B). Technical details on sequence completion from partially observed sequence are presented in Section V. In Section VI, we report on experiments, model selection, and validation results. We conclude the present work

and offers discussions in Section VII. We emphasize that no personally identifiable information (PII) was gathered or used in conducting this study. The mobility data that was analyzed was anonymous and aggregated in strict compliance with the carrier's privacy policy.

II. Related Work

A. Human Mobility Modeling and Prediction

Urban mobility models characterize multiple aspects of individuals' travel patterns. Large amount of work focuses on the activities (trip purpose), such as the spatial (location, [1, 11, 17, 27, 34]) - temporal (start time and duration, [31]) choices of a single activity, or activity patterns (daily/weekly activity scheduling, [2, 4, 7, 13, 14, 25, 32, 40, 42, 44]). Another branch of research considers trips linking these activities, studying trajectories [22, 26, 35], travel mode [3, 6, 18, 29, 36, 39, 45, 46], by applying map matching and route choice [9, 37].

State-of-the-art can also be classified by the data sources used to model individual urban mobility. Early studies mainly used travel surveys [3, 4, 7]. In the recent decade, with the mobile phone data more available, passively collected data such as GPS [1, 2, 6, 9, 17, 18, 21–25, 27, 29, 31, 35–37, 39, 43–46], CDR [8, 11–14, 19, 26, 28, 32, 33, 40, 42] and location-based social networks (LBSN) data [10, 41] has provided grounds for new approaches in urban mobility studies. GPS data is granular in both spatial and temporal resolution. However, the availability of such granular data is usually limited to hundreds of travelers. LBSN data is exact in locations, and may provide additional social relation, comments and reviews on the venues for larger samples of travelers. However, LBSN data is limited by its discontinuity and sparsity in time. CDR data provides a trade-off between spatial-temporal resolution and ubiquity, while covering millions of travelers.

Studies that are not concerned with predictive or generative methods fall into two categories: first category tends to purely understand generic human mobility laws using descriptive statistics [8, 19, 23, 33], the other category focus on the problem of recognition (activity, travel mode, [6, 9, 21, 24, 28, 32, 37]) rather than prediction. The second category focus on mobile phone data since activity type and travel mode are not explicitly observed from the data itself. One illustration of the distinction between recognition and prediction is the modeling of travel mode. Recognition of travel mode from mobile phone data takes a segment of single mode trajectory and classify its travel mode based on the extracted features such as travel speed, acceleration, distances, heading change rate, etc. [18, 29, 36, 39, 46]. One cannot use this model to determine the travel mode until observing the trajectory. On the other hand, discrete choice models from travel surveys can be used to predict the travel mode without having the trips happened using just features of alternative modes and characteristics of individuals [3]. For studies that do focus on predictive (generative) power, most work was focused on predicting only next location (or duration) since it is a well formulated task that is also easier to validate. Some researchers make prediction by assuming Markov properties [1, 12, 34, 41]; other researchers treat prediction of next location as a classification (regression) problem using supervised learning [11, 17]; and some researchers used trajectory matching techniques to make the prediction [27, 43]. However, not much research has been done on models able to predict a sequence of locations (durations) for the full day or longer.

Another observation is that most of the previous studies focus on only one aspect of urban mobility (such as location, duration, travel mode), or model these several aspects separately. Not many studies have focused on modeling daily activity patterns and scheduling that fuse activity type, location and duration together, which enables the model to generate a sequence of samples. Eagle and Pendland [13], Farrahi and Gatica-Perez [14], and Zheng et al. [44] used unsupervised techniques such as PCA and topic models to clustering daily activity patterns. However, they only included primary activity types such as "home" and "work", all other activities are categorized as "other". Liao et al. unified the process of map matching, place detection, and significant activity inference through a hierarchical conditional random field (CRF) using GPS data [25]. However, their model is discriminative in nature and is most suitable for recognition, rather than generating new sequences. Widhalm et al. [40] used an undirected relational Markov network to infer urban activities with CDR data. However, they did not model activity transitions due to the lack of cliques for consecutive activities. A similar work by Yin et al. [42] improved the modeling of activity patterns (spatial-temporal profiles of primary and secondary activity) with explicit modeling of contextual dependent activity transition probabilities. They validated the activity inference and the reliability of new activity chain generation with three independent data sources.

To summarize, existing literature has focused on long term travel demand and short term traffic state forecasts, while current methods of individual human mobility modeling have got limitations that make them only partly useful for medium-term forecasting. In this paper, we fill this gap with sequence learning methods applied to build generative urban mobility models from cellular data.

III. Modeling Framework

The developed data processing and modeling pipeline is presented in Fig. 1. Historical CDR data are processed to unlabeled historical activity sequences [42]. Urban mobility models are built upon these historical activity sequences. In this paper, we improved the state-of-the-art urban mobility models including interpretable IO-HMM models, as detailed in Section IV-A, and deep LSTM models, as detailed in Section IV-B.

On a target day, we receive streaming CDR data at different time of day (e.g. 3:00 am, 9:00 am, 3:00 pm, etc.), which are then processed to partially observed activity sequences. These partially observed sequences, along with the pre-trained parametric urban mobility models, are sent to the sequence predictor. The sequence predictor predicts and completes the

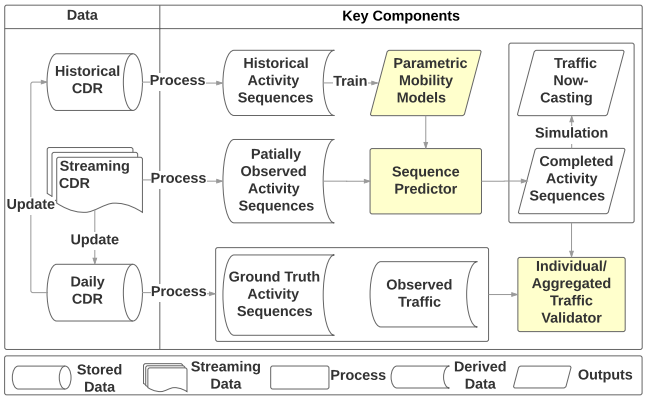


Fig. 1: Modeling framework diagram. The left column represents the input to the algorithms and the right column represents the model components. Our key contribution of improved deep urban mobility models, sequence predictor, and validation are shown in shaded yellow.

activity sequences for the rest of the day based on the observed information, as detailed in Section V. The completed activity sequences are sent to MATSim, a state-of-the-art agent-based traffic micro-simulation tool that performs traffic assignment. MATSim generates the predicted traffic conditions for the day.

By the end of the day, full day CDR are observed and processed to ground truth activity sequences. These ground truth activity sequences are validated against the predicted activity sequences at both individual level and aggregated level at different times of day. We also validate the resulting traffic from predicted activity sequences versus ground truth sequences, as detailed in Section VI. Finally, historical CDR database is updated with the new day's CDR, and urban mobility models can be updated and re-trained overnight.

IV. Urban Mobility Models from Cellular Data

Yin et al. [42] developed an end-to-end pipeline from raw CDR data to urban mobility models, which characterizes heterogeneous context-dependent activity transition probabilities and spatial-temporal profiles of primary and secondary activities ("shopping", "leisure", etc.). The pipeline can be summarized in the following steps: (1) processing raw CDR trajectories to stay locations; (2) identifying primary activity (home and work) locations; (3) recognizing activity patterns using input-output hidden Markov model (IO-HMM); (4) validating activity recognition with small set of ground truth activities based on short range distributed antenna systems (DASs). In this section, we improve the urban mobility models in two aspects: (1) use ground truth activities for co-training the IO-HMM models; and (2) apply deeper LSTM models for activity sequences. We then compare how these models perform in the prediction tasks.

A. Co-training IO-HMM models

Supervised learning of activity types requires data with labeled ground truth. In urban mobility, the ground truth activities are derived by either manually labeled [17, 25], or collected for a small group of participants from a survey accompanying GPS data [21]. Privacy concerns and spatial resolution of CDR data precludes us from obtaining extensive ground truth labels. While fully unsupervised models can be used to cluster activities with similar temporal and spatial profiles, the recognized activities may not correspond to conventional activity types. In this subsection, we propose to use semi-supervised learning to reach a compromise – we use a small set of ground truth activities based on short short range distributed antenna systems (DASs) to direct the learning process. Short range antennas usually serve only a small range of area, which have relatively high spatial resolution. These short range antennas provide us the opportunity to label "ground truth" activities with Point of Interest (POI) information and domain knowledge. The collection process of ground truth activities is documented in [42].

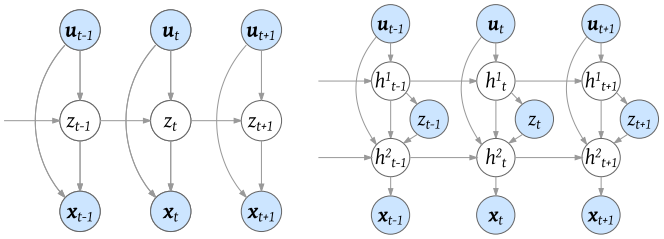
Traditionally, semi-supervised learning is used to improve classifier performance, that is, to use "cheap" unlabeled data to assist training of labeled data. In our work, we adopt another view of semi-supervised approach, that is, we use labeled data to help direct the pattern recognition from unlabeled data. Zhu [47] did a thorough literature review on semi-supervised learning methods, including self-training, co-training, graph-based methods and Expectation-Maximization (EM) in generative models. In our work, we took the advantage of EM in generative models and co-training to improve the activity pattern recognition performance.

The idea behind co-training is that one uses two views of a sample that inform the learning algorithms by teaching one another. Ideally each sample is represented by two independent sets of features, which is however unlikely to exist [15]. Co-training can also be applied by using the same set of features but two different classifiers, which has been proven to perform well by [16]. It is expected to be less sensitive to mistakes than self-training.

In this work, we choose to use a semi-supervised IO-HMM with EM algorithm as the generative classifier, and a decision tree (DT) classifier as its discriminative counterpart. With this combination, we have both the classification power of discriminative model and the generative power of IO-HMM models.

Algorithm 1 Co-training of urban activities

Input: Labeled data L, unlabeled sequences S, confidence thresholds θ_1 and θ_2.
Output: IO-HMM model m_1 and DT model m_2.
Initialisation: $L_1 = L_2 = L$
while L_1, L_2 changes **do**
  Train semi-supervised IO-HMM m_1 from S and L_1.
  Train DT model m_2 from L_2.
  Classify the unlabeled data with m_1 and m_2 separately.
  Add data labeled by m_1 with confidence $\geq \theta_1$ to L_2.
  Add data labeled by m_2 with confidence $\geq \theta_2$ to L_1.
end while
return m_1, m_2.



(a) IO-HMM (b) LSTM

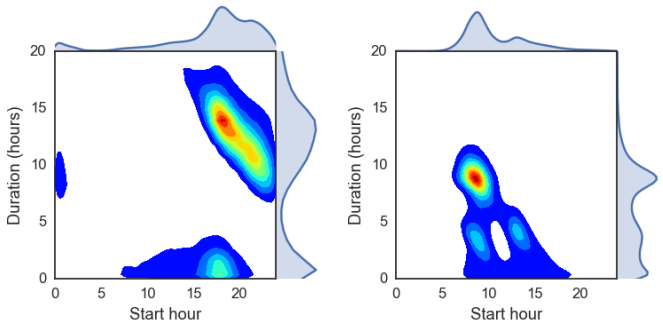
Fig. 2: Deep Urban Mobility Architectures, IO-HMM (left) and LSTM (right). The solid nodes represent observed information, while the transparent (white) nodes represent latent random variables. The top layer contains the *observed* input variables $\boldsymbol{u_t}$; the middle layer contains categorical variables z_t (latent in IO-HMM since we include secondary activities while observed in LSTM since we only include "home", "work", and "other"); and the bottom layer contains observed output variables $\boldsymbol{x_t}$. h_t are LSTM cells in the LSTM architecture.

1) Semi-Supervised IO-HMM Model with EM: Hidden Markov Models (HMMs) have been extensively used in the context of action recognition and signal processing. IO-HMM extends standard HMM by relaxing the assumption of homogeneous transition and emission probabilities. For instance, if a user engages in a home activity on a weekday, and departs for the next activity in the morning, she is likely going to work. If she departs in the evening, the trip purpose is likely to be recreation or shopping. In Fig. 2a, the solid (blue) nodes represent observed information, while the transparent (white) nodes represent latent random variables z_t corresponding to unobserved activity types.

Input features: The top layer $\boldsymbol{u_t}$ contains the *observed* contextual variables that have to be known prior to a transition: (1) a binary variable indicating whether the day is a weekend; (2) five binary variables indicating the time of day that the activity starts, morning (5 to 10am), lunch (10am to 2pm), afternoon (12 to 2pm), dinner (4 to 8pm) or night (5pm to midnight); and (3) for the users with identified work location, the number of hours the user has spent at work this day. This variable contains accumulated knowledge on the past activities.

Output features: The bottom layer $\boldsymbol{x_t}$ contains *observed* output variables that are available during training of the models but not when generating activity sequences: (1) $x^{(1)}$, the distance between the current stay location and the user's home; (2) $x^{(2)}$, the distance between the current stay location and the user's work place; (3) $x^{(3)}$, the duration of the activity; and (4) $x^{(4)}$, whether the user has visited this stay location cluster previously.

The difference between IO-HMM and semi-supervised IO-HMM lies in the forward-backward algorithms. If we have ground truth activity (hidden states z) for timestamp t, then we will use $I_{j,t}$ to replace $\varphi_{ij,t}$ where $I_{j,t}$ is 1 if the hidden state $z_t = j$ at timestamp t in the labeled data, 0 otherwise, since $\Pr(z_t = j \mid z_{t-1} = i)$ reduces to $\Pr(z_t = j)$ with observed information. A summary of the differences between HMM, IO-HMM and semi-supervised IO-HMM is presented in TABLE I.



(a) Home (b) Work

Fig. 3: Joint distribution plot of duration and start hour for home (left) and work (right).

2) Decision Tree Counterpart: Decision trees are interpretable classifiers that are capable of generating arbitrarily complex decision boundaries. They have been used successfully in many diverse areas [30]. In this work, we use CART (Classification and Regression Trees) classifier. The features we include are the combination of input and output features in IO-HMM.

3) Model Selection: Model selection for co-training includes the choice of hidden states. The choice should come directly from the collection of ground truth activities. Yin et al. [42] collected ground truth activities for "Food/Shop", "Stop in Transit", "Recreation", "Personal Business", and "Travel". In this paper, we follow their convention and include these five secondary activities in the hidden states.

We further noticed a significant heterogeneity within home and work activities. Temporal profile of home activities in Fig. 3a has two major clusters. The upper cluster indicates regular overnight home activities (H_1) and the lower cluster indicates short stay at home before going to some other activities (H_2). The temporal profile of work activities in Fig. 3b has three clusters. The upper cluster indicates regular "9 to 5" work activities without a break (W_1). The lower left cluster represents the morning work activities (W_2) and the lower right cluster represents the afternoon work activities (W_3). Considerably, the transition probability from H_2 to work is lower, and the transition probability from W_2 to "Food/Shop" should be higher but to "Recreation" should be lower than the transition probability from W_1 or W_3. By separating home and work activities into sub-activities, we expect to get better contextual-dependent transition probabilities. A more rigorous definition of sub-activities is:

1) H_1: cross day home activity that starts before 3:00 am and end after 3:00 am.
2) H_2: other home activities.
3) W_1: work activity if it is the only work activity in a day.
4) W_2: first work activity if there are more than one.
5) W_3: second work activity if there are more than one.
6) W_4: other work activities.

TABLE I: Highlights of comparison between an HMM vs. IO-HMM vs.semi-supervised IO-HMM ($\boldsymbol{u_t}$, z_t, $\boldsymbol{x_t}$ denote input, hidden and output variables respectively, i is an index of a hidden state, t is a sequence timestamp index, $I_{j,t}$ is 1 if the hidden state $z_t = j$ at timestamp t in the labeled data, 0 otherwise).

	HMM	IO-HMM	semi-supervised IO-HMM
initial state probability π_i	$\Pr(z_1 = i)$	$\Pr(z_1 = i \mid \boldsymbol{u_1})$	
transition probability $\varphi_{ij,t}$	$\Pr(z_t = j \mid z_{t-1} = i)$	$\Pr(z_t = j \mid z_{t-1} = i, \boldsymbol{u_t})$	
emission probability $\delta_{i,t}$	$\Pr(\boldsymbol{x_t} \mid z_t = i)$	$\Pr(\boldsymbol{x_t} \mid z_t = i, \boldsymbol{u_t})$	
forward variable $\alpha_{i,t}$	$\delta_{i,t} \sum_l \varphi_{li,t} \alpha_{l,t-1}$, with $\alpha_{i,1} = \pi_i \delta_{i,1}$		$\delta_{i,t} I_{i,t} \sum_l \alpha_{l,t-1}$, with $\alpha_{i,1} = I_{i,t} \delta_{i,1}$, if t is observed
backward variable $\beta_{i,t}$	$\sum_l \varphi_{il,t} \beta_{l,t+1} \delta_{l,t+1}$, with $\beta_{i,T} = 1$		$\sum_l I_{l,t+1} \beta_{l,t+1} \delta_{l,t+1}$, with $\beta_{i,T} = 1$, if $t+1$ is observed
complete data likelihood L_c	$\sum_i \alpha_{i,T}$		
posterior transition probability $\xi_{ij,t}$	$\varphi_{ij,t} \alpha_{i,t_1} \beta_{j,t} \delta_{j,t} \,/\, L_c$		
posterior state probability $\gamma_{i,t}$	$\alpha_{i,t} \beta_{i,t} \,/\, L_c$		

We compare experimentally the basic and extended specifications (one with 7 activities and the other with 11 activities) in Section VI-A.

B. LSTM models

LSTM models have been extensively used for modeling complex sequences, including natural language, videos and handwriting trajectories. We design a 2-layer LSTM model structure for modeling activity sequences as shown in Fig. 2b.

The top layer models activity transitions between "home", "work", and "other" (we treat all secondary activities as "other" since we do not have full ground truth labels for all secondary activities). $\boldsymbol{u_t}$ represents the input contextual features similar to the ones specified in IO-HMM models. The only difference is that we include the observed previous activity (one of "home", "work", and "other") in this feature vector. The reasons are (1) in LSTM models, the previous activity type is observed prior to transition to a new activity, and (2) for generating new activity based on the previous activity, we need to include this previous activity in the training phase. Note that in IO-HMM models, we use dynamic programming to get the probabilities of previous activity, as detailed in Section V-A. h_t^1 represent the first layer of LSTM cells and z_t represents the observed current activity type. The loss function for this top layer is:

$$L_1(\boldsymbol{\theta_1}) \quad = \quad -\sum_{t=1}^{T} \sum_j (z_t = j) \cdot \log \phi\left(h_t^1; \boldsymbol{\theta_1}\right)_j$$

where ϕ is the softmax function, $\boldsymbol{\theta_1}$ is the collection of parameters for this LSTM neural network, and j belongs to one of the activity types "home", "work" and "other".

The bottom layer is a mixture density network (MDN) which models the **distributions** of spatial (location) and temporal (duration) variables $\boldsymbol{x_t}$ associated with each activity type z_t. MDN was first described in [5] and was further developed for handwriting synthesis tasks [20]. The contextual vector $\boldsymbol{u_t}$, first layer LSTM cells h_t^1, second layer LSTM cells from previous timestamp h_{t-1}^2, and the current activity type z_t are the inputs to the second layer LSTM cells h_t^2, which generates the coefficients of the mixture distributions (in our task we assume Gaussian distribution for each output feature) $\{\hat{\boldsymbol{\pi}}, \hat{\boldsymbol{\mu}}_{d_h}, \hat{\boldsymbol{\mu}}_{d_w}, \hat{\boldsymbol{\mu}}_{\text{st}}, \hat{\boldsymbol{\mu}}_{\text{dur}}, \hat{\boldsymbol{\sigma}}_{d_h}, \hat{\boldsymbol{\sigma}}_{d_w}, \hat{\boldsymbol{\sigma}}_{\text{st}}, \hat{\boldsymbol{\sigma}}_{\text{dur}}, \hat{\boldsymbol{\rho}}_{\text{st, dur}}\}$. At each timestamp t, $\hat{\boldsymbol{\pi}}_t$ is an M by 1 array representing the mixture component weights, M is the number of mixture components. $\hat{\boldsymbol{\mu}}_{d_h,t}$, $\hat{\boldsymbol{\mu}}_{d_w,t}$, $\hat{\boldsymbol{\mu}}_{\text{st},t}$, and $\hat{\boldsymbol{\mu}}_{\text{dur},t}$ are M by 1 arrays representing the component means of the distance to home, distance to work, start time, and duration. $\hat{\boldsymbol{\sigma}}_{d_h,t}$, $\hat{\boldsymbol{\sigma}}_{d_w,t}$, $\hat{\boldsymbol{\sigma}}_{\text{st},t}$, and $\hat{\boldsymbol{\sigma}}_{\text{dur},t}$ are M by 1 arrays representing the component standard deviations of the distance to home, distance to work, start time, and duration. $\hat{\boldsymbol{\rho}}_{\text{st, dur},t}$ represents the correlation between start time and duration. This second layer mixture networks is meant to divide "home", "work", and "other" activities into smaller and finer components, each has its local spatial-temporal distributions. The loss function for this bottom layer is:

$$L_2(\boldsymbol{\theta_2}) \quad = \quad \sum_{t=1}^{T} -\log \sum_{i}^{M} \pi_t^i \mathcal{N}(\boldsymbol{x_t} | \hat{\boldsymbol{\mu}}_t^i, \hat{\boldsymbol{\sigma}}_t^i, \hat{\boldsymbol{\rho}}_t^i)$$

where $\boldsymbol{\theta_2}$ is the collection of parameters of the neural network used to generate the mixture density distribution coefficients $\{\hat{\boldsymbol{\pi}}, \hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\sigma}}, \hat{\boldsymbol{\rho}}\}$, i is the index of the mixture component. $\mathcal{N}$ is the Gaussian probability density function.

This two-layer structure extends Lin et al. [26] as we moved the modeling of activity types into the first layer. Otherwise we keep the same model specifications and loss functions as in that paper.

V. Urban Mobility Prediction

The problem we are solving in this section is to predict the activity sequence of the rest of day, given partially observed sequences at a cut time (e.g. 9:00 am). This problem can be tackled by breaking it into two inferential sub-problems: (1) what an individual has done; and (2) what he/she is likely to do. We will show how these two sub-problems are tackled using IO-HMM model and LSTM model, respectively.

A. Prediction using IO-HMM models

1) Filtering: The first step is calculating $\Pr(z_{t-1} = i \mid \boldsymbol{u_{1,\ldots,t-1}}, \boldsymbol{x_{1,\ldots,t-1}})$. Since the next activity to be generated depend on the contextual variables such as time of day and day of week information, as well as the previous hidden activity, we need to understand what is the last observed activity. There are two cases:

1) By the cut time, the last observed activity is completed. That is, the person is traveling to the next activity location. This case is simple since we can use standard forward algorithm to estimate the posterior probability

$\Pr(z_{t-1} = i \mid \boldsymbol{u_{1,\ldots,t-1}}, \boldsymbol{x_{1,\ldots,t-1}})$ of the last observed activity. One thing to note is that we need to sample a travel time that are longer than the observed travel time from the complete of the last activity to respect the fact that no new activities happened before the cut time.

2) By the cut time, the last observed activity is not completed. In this case, we apply a modification to the forward algorithm: the emission probability of duration of last activity is a survival function: $\Pr(x_t > d_t^o \mid z_t = i, \boldsymbol{u_t})$, where d_t^o is the observed duration of the last activity until the cut time. After the filtering, we sample a new duration with the truncated distribution whose lower bound is d_t^o to respect the fact that the activity ends after the cut time.

2) Activity generation: With the last activity inferred, the activity generation algorithm is same as Yin et al. [42]: at the end of this activity the relevant context information $\boldsymbol{u_t}$ is updated and the next activity is selected given the newly obtained transition probabilities. Next, the activity duration is sampled from the conditional distribution given the activity type and the start time. Next, the activity location is selected - if the activity is a home or work activity, the exercise is trivial. If not, we calculate the probability of choosing each cluster in the user's historical location clusters based on the conditional distribution of $x^{(1)}$ distance to home and $x^{(2)}$ distance to work given the activity type. This is different from Yin et al. [42] that since their population is synthetic, they do not have location history of the user thus can only generate a new destination from the choice set of TAZs and a random point within the TAZ. By adopting the historical location clusters of the user, we reduce the variance of the location choice. The process continues until the full daily sequence of activities is generated.

B. Prediction using LSTM models

The procedure is straightforward based on Fig. 2b. The LSTM model first calculates $h^1_{1,\ldots,t-1}$, $h^2_{1,\ldots,t-1}$ based on observed $\boldsymbol{u_{1,\ldots,t-1}}$ and $z_{1,\ldots,t-1}$. To generate the next activity at timestamp t, we first update the contextual vector $\boldsymbol{u_t}$ and top LSTM layer h^1_t. The softmax outputs of the top layer is used for sampling the new activity type z_t. z_t, along with $\boldsymbol{u_t}$, h^1_t, h^2_{t-1} are used in the bottom layer of the model. The sampling of the output variables distance to home, distance to work, and duration from the distributions of mixture density network (MDN) is similar to the ones described in [20, 26]. The rest of the generation process is similar to the generation process of IO-HMM model.

VI. Experimental Results

In this section, we describe two regional experiments of medium term travel demand forecast at different times of day. The master data used in these studies comprise a month of anonymized and aggregated CDR logs collected in Summer 2015 by a major mobile carrier in the US, serving millions of customers in the San Francisco Bay Area. No personally identifiable information (PII) was gathered or used for this study. As described previously, CDR raw locations are converted into highly aggregated location features before any actual modeling takes places.

The first experiment use the City of San Francisco for model selection. We evaluate the prediction performance of different models and validate the predictions at individual and aggregated level. The second experiment scales to whole San Francisco Bay Area where we predict the traffic conditions based on trained models for commuters from each of the 34 super-districts. We evaluate the resulting traffic from micro-simulation and validate it against the resulting traffic of observed ground truth data.

We choose a typical weekday June 10, 2015 as the target day. For each regular commuter with available data on that day, we slice the data by different cut time (e.g. 3:00 am, 4:00 am, ..., 11:00 pm) and predict the activities for the rest of the day based on the observed information by the cut time.

A. Model Comparison

In this subsection, we evaluate the performance of different models and methods.

1) **NN**: Nearest Neighbor model, the benchmark model and the expected upper bound of the performance. NN is a fully personalized model that match the observed trajectory with the trajectory history of the user, and use the matched trajectory as prediction for the rest of day. The distance features we used are (1) difference in day type (weekday or weekend, 0 if equal and 1 if not), and (2) the Hamming distance between observed partial sequence and each historical sequence by cut time. We calculate the Hamming distance by segmenting each sequence into 15-minutes segments. For each 15-minutes segment, we set the distance as 0 if the location clusters in two sequences are same (in most of the 15 minutes) and 1 if not. The total Hamming distance is the sum of each segment. We give the day type feature a high weight (in this case 100) so that NN will search the matching sequence within the same day type. Note that NN model is only used for trajectory matching and does not provide insights and interpretability as other activity models.
2) **IO-HMM-unsupervised-7**: The IO-HMM model with 7 hidden states, with the input and output features specified in Section IV-A1.
3) **IO-HMM-co-training-7**: The co-training IO-HMM model specified in Section IV-A. In this model we treat home and work as two activities, thus with 5 secondary activities there are 7 states in total. The threshold parameters for both semi-supervised IO-HMM model with EM (θ_1) and Decision Tree (θ_2) are 0.9. This threshold is chosen based on literature and validation accuracies on secondary activity recognition.
4) **IO-HMM-co-training-11**: In this model we separate "home" and "work" to 6 sub-activities defined in Section. IV-A3. Thus there are 11 states in total.
5) **LSTM-3**: The LSTM model specified in Section IV-B. We used 64 hidden unites in each LSTM cell and

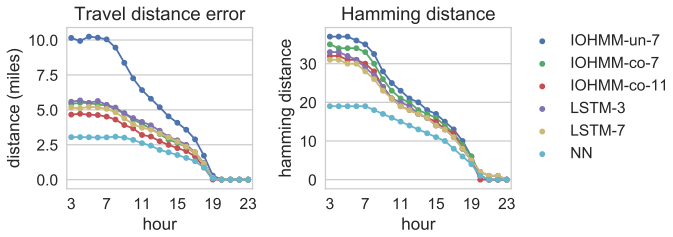
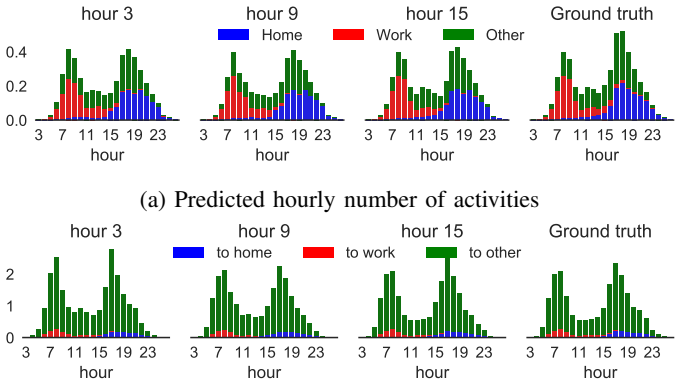


Fig. 4: Models comparison. Two validation metrics are used: median travel distance error (left) and median Hamming distance (right). The x-axis is the prediction hour (cut hour) and the y-axis is the validation error. Each series of points represents the performance of a model.

40 mixture components in the mixture density network (MDN).

6) **LSTM-7**: In this model we separate "home" and "work" to 6 sub-activities thus there are 7 activity types including "other".

In Fig. 4, we plot how the two validation metrics, (1) median travel distance error (left), and, (2) median Hamming distance (right) change for different cut hours using different models. The travel distance error is calculated as the difference between the observed daily travel distance and predicted daily travel distance. The median error of all users are used in the plot. The travel distance error mainly captures the spatial location choice performance of models. The Hamming distance is calculated as in NN models by segmenting the daily sequence into 96 discrete 15-minutes segments. The median error of all users are used in the plot. The Hamming error mainly captures the temporal day structure performance of models. From Fig. 4, we can see that: (1) NN models performs best among all models because it is a fully personalized non-parametric model; (2) IO-HMM models are better at spatial performance than LSTM models since we used co-training to direct the learning of secondary activity profiles. This is also proven by comparing the unsupervised model performance with the co-training results; (3) LSTM models are better at capturing the day structures. Hamming error captures the performance of day structures such as "home", "work", and important secondary activities. LSTM models slightly outperforms IO-HMM models in this metric because it is more flexible and deeper in modeling activity transitions and long term dependencies; (4) By separating "Home" and "Work" into smaller sub-activities, we get better spatial-temporal performance in both IO-HMM models and LSTM-models. This proved our assumption that by separating these primary activities, we can better learn the activity transitions between primary activities and between primary activities and secondary activities; (5) We can explore the limit of the predictability of human mobility. The median travel distance error at the beginning of the day using fully personalized model is about 3 miles, and this number is about 5 miles using non-parametric group models. The median Hamming error is 20 at the beginning of the day using fully personalized model, that is, 5 hours of wrongly predicted activities within a day.



(a) Predicted hourly number of activities

(b) Predicted hourly travel distance in miles

Fig. 5: Predicted aggregated travel demand. The average number of activities (top) and travel distance in miles (bottom) (y-axis) starting in each hour (x-axis). Each of the four subplot represents the prediction at hour 3:00 am, 9:00 am, 3:00 pm, and the observed ground truth.

This error is mainly due to the shift in home and work hours. Since different people has different start hour of work and preferences on the time of going back home, fully personalized model is better at capturing this based on the individual's history.

B. Aggregated Level Evaluation

We validate the predicted versus observed hourly aggregated travel behavior in this subsection. We adopt the IO-HMM-co-training-11 as our urban mobility model. The aggregated pattern is very similar between the best performed IO-HMM and LSTM models.

Fig. 5a shows the average number of activities (y-axis) starting in each hour (x-axis). To make it more informative, we decompose the total number of activities into "home", "work" and "other". We can see that the predicted number of activities of each type is quite comparable to the ground truth observed at the end of the day. The same peak of work activities in the morning and home activities in the evening are observed in all predictions and ground truth. The main difference between our predictions and the ground truth is that we tend to under-predict the number of "other" activities.

Fig. 5b shows the average travel distance in miles (y-axis) in each hour (x-axis). One observation is that the travel distance of "to work" in the morning peak and "to home" in the evening peak are low compared to "to other". This is because a lot of people go for secondary activities before arriving at work and home, as shown in Fig. 5a. The other observation is that though the predicted number of secondary activities is lower, the travel distances to these locations are higher in our predictions. This indicates some inefficiencies in our secondary location choice - people select most convenient locations for secondary activities, and points towards possible improvements in location choice model for secondary activities.

C. Evaluation via Traffic Micro-simulation

In this subsection, we span the scope of the study to the 34 super-districts as defined by the San Francisco Metropolitan Transportation Commission (MTC) to validate the predicted resulting traffic in a region with 7.5M citizens. Since most of the short range DASs are located in urban area such as the City of San Francisco, the ground truth secondary activities are rarely available for other super-districts in Bay Area. Thus we train 34 semi-supervised IO-HMM model with "home" and "work" as ground truth, one for each super-district. For each regular commuter with data available on June 10, 2015, we predict his/her activities for the rest of day based on the activities observed by a cut time. Traffic micro-simulation is a conventional approach in studying performance and evaluating transportation scenarios. MATSim is a state-of-the-art agent based traffic micro-simulation tool that performs traffic assignment for the set of agents with pre-defined activity plans. For each cut time (e.g. 3:00 am, 9:00 am, 3:00 pm, 9:00 pm), we compared the results of the flows produced on the Bay Area network containing all freeways and primary and secondary roads (a total of 24'654 links) from the predicted activity sequences with the ground truth activity sequences. The fit score (1) adjusted R^2; (2) mean absolute percentage error (MAPE, %) are summarized in TABLE II. Fig. 6 plots the volume profiles of two freeway locations, one near the entrance of bay bridge in the eastbound and the other near the crossing of I-880 and US-101. For each location, 4 subplots shows the predictions (in blue) at 3:00 am, 9:00 am, 3:00 pm and 9:00 pm vs the ground truth profiles (in orange). We can see that (1) the predictions get closer to the ground truth volumes with more observed data in the day and (2) our predictions tend to generate slightly higher traffic volumes than ground truth traffic. This is consistent with our previous discussion on the inefficiencies in secondary location choices.

TABLE II: The coefficient of determination (R^2) and mean absolute percentage error (MAPE, %, in the parenthesis) of the predicted versus ground truth resulting traffic counts on 600 locations on the Bay Area road network. The row index is the prediction hour and the column index is the predicted hour. No scores are reported under diagonal because the traffic in the predicted hour is already observed by the prediction hour.

	3	6	9	12	15	18	21	24
3	1 (0)	0.864 (38.1)	0.881 (16.2)	0.876 (18.0)	0.890 (19.1)	0.891 (14.2)	0.924 (14.5)	0.896 (19.7)
9	- -	- -	0.997 (2.9)	0.977 (9.0)	0.947 (14.1)	0.931 (10.8)	0.934 (13.4)	0.937 (15.1)
15	- -	- -	- -	- -	0.995 (4.4)	0.962 (8.8)	0.960 (11.1)	0.955 (13.0)
21	- -	- -	- -	- -	- -	- -	0.999 (2.1)	0.998 (3.8)

TABLE II proves that we can use observed information of the day to improve traffic volume prediction. The coefficient of determination increase and the MAPE decrease with the prediction hour. When we make prediction at the beginning of each hour, we can improve the coefficients of determination

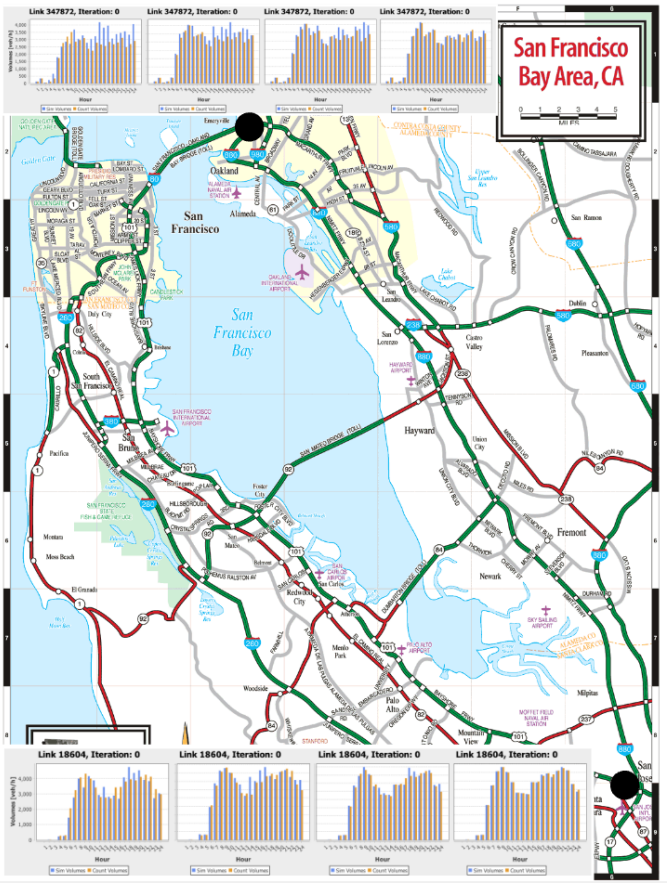


Fig. 6: A fragment of the SF Bay Area road network. Inlet graphs illustrate two sample hourly vehicle volume profiles for observed (orange) and predicted (blue) at 3am, 9am, 3pm, and 9pm.

in that hour to be greater than 0.99 and the MAPE less than 5%. The artifact of perfect prediction of 3:00 am is because we defined the start of the day as 3am, there should be few traffic occurring during that hour. The lower predictability at off-peak hours (e.g. 6:00 am and 12:00 am) is consistent with the observations in [42] of higher variability in travel choices for secondary activities.

VII. Conclusion

In this paper, we proposed a medium term travel demand nowcasting framework. It predicts daily travel demand and traffic conditions at different times of day with partially observed user traces from cellular data and pre-trained urban mobility models. This solution bridges the gap between long term forecast (days, months to years ahead) and short term prediction (seconds to hours ahead), which are the two main-streams of literature in travel demand forecasting.

We improved the state-of-the-art deep generative parametric mobility models using co-training in IO-HMM and LSTMs. We provided partially observed user traces at different times of day to these models and generated the complete daily sequences. We validated the results with the ground truth sequences based on (1) individual level discrepancies; (2) ag-

gregated level hourly travel demand; and (3) the resulting traffic through micro-simulation. A non-parametric individualized nearest neighbor model was explored as the practical limit of predictability of individual's daily travel. We demonstrated that parametric models trained at aggregated group level (due to privacy concern) can approach this limit in terms of prediction accuracy. Among the generative models we compared, IO-HMM models are interpretable and has the power of activity recognition as a range of travel choices might depend on the activity types. Co-training applied to IO-HMM models performs better at secondary activity location choices since we used the ground truth activities to direct the learning process. LSTM models are better at learning day structures since they use continuous hidden state space and are expected to be better at learning long term dependencies. Future research will focus on incorporating activity types in LSTM models and using existing ground truth labels to direct the learning process of LSTM models.

We consider San Francisco residents as a group in the first experiment and each super-district as a group in the second experiment. We trained one urban mobility model for each group. However, certain heterogeneity in activity patterns exists among different sub-groups. Correctly partitioning the population into sub-groups should help us better approach the limit of the predictability in human mobility. We acknowledge it as a current limitation of the paper.

In terms of traffic volumes, our experiments show promising results of medium term forecast. We have reached a MAPE of less than 5% one hour ahead and 10% three hour ahead. Results also show that we can improve the prediction accuracy by incorporating more of the observed data by the time of prediction. Our prediction of traffic conditions is available not only for freeways and arterial where high-resolution detectors data are available from direct observations. Our framework provides accurate prediction for the whole network, detailed in terms of activities and travel itineraries of citizens, providing an actionable model to improve performance of regional transportation systems and inform interventions towards reducing negative impacts of congestion.

References

[1] D. Ashbrook and T. Starner. Using gps to learn significant locations and predict movement across multiple users. *Personal and Ubiquitous computing*, 7(5):275–286, 2003.

[2] M. Baratchi, N. Meratnia, P. J. Havinga, A. K. Skidmore, and B. A. Toxopeus. A hierarchical hidden semi-markov model for modeling mobility data. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pages 401–412. ACM, 2014.

[3] M. E. Ben-Akiva and S. R. Lerman. *Discrete choice analysis: theory and application to travel demand*, volume 9. MIT press, 1985.

[4] C. R. Bhat and S. K. Singh. A comprehensive daily activity-travel generation model system for workers. *Transportation Research Part A: Policy and Practice*, 34(1):1–22, 2000.

[5] C. M. Bishop. Mixture density networks. 1994.

[6] W. Bohte and K. Maat. Deriving and validating trip purposes and travel modes for multi-day gps-based travel surveys: A large-scale application in the netherlands. *Transportation Research Part C: Emerging Technologies*, 17(3):285–297, 2009.

[7] J. L. Bowman and M. E. Ben-Akiva. Activity-based disaggregate travel demand model system with activity schedules. *Transportation Research Part A: Policy and Practice*, 35(1):1–28, 2001.

[8] F. Calabrese, M. Diao, G. Di Lorenzo, J. Ferreira, and C. Ratti. Understanding individual mobility patterns from urban sensing data: A mobile phone trace example. *Transportation research part C: emerging technologies*, 26:301–313, 2013.

[9] J. Chen and M. Bierlaire. Probabilistic multimodal map matching with rich smartphone data. *Journal of Intelligent Transportation Systems*, 19(2):134–148, 2015.

[10] E. Cho, S. A. Myers, and J. Leskovec. Friendship and mobility: user movement in location-based social networks. In *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 1082–1090. ACM, 2011.

[11] T. M. T. Do and D. Gatica-Perez. Where and what: Using smartphones to predict next locations and applications in daily life. *Pervasive and Mobile Computing*, 12:79–91, 2014.

[12] N. Eagle, A. Clauset, and J. A. Quinn. Location segmentation, inference and prediction for anticipatory computing. In *AAAI Spring Symposium: Technosocial Predictive Analytics*, pages 20–25, 2009.

[13] N. Eagle and A. S. Pentland. Eigenbehaviors: Identifying structure in routine. *Behavioral Ecology and Sociobiology*, 63(7):1057–1066, 2009.

[14] K. Farrahi and D. Gatica-Perez. Discovering routines from large-scale human locations using probabilistic topic models. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 2(1):3, 2011.

[15] V. Frinken, A. Fischer, H. Bunke, and A. Foornes. Co-training for handwritten word recognition. In *Document Analysis and Recognition (ICDAR), 2011 International Conference on*, pages 314–318. IEEE, 2011.

[16] S. Goldman and Y. Zhou. Enhancing supervised learning with unlabeled data. In *ICML*, pages 327–334, 2000.

[17] J. B. Gomes, C. Phua, and S. Krishnaswamy. Where will you go? mobile data mining for next place prediction. In *International Conference on Data Warehousing and Knowledge Discovery*, pages 146–158. Springer, 2013.

[18] H. Gong, C. Chen, E. Bialostozky, and C. T. Lawson. A gps/gis method for travel mode detection in new york city. *Computers, Environment and Urban Systems*, 36(2):131–139, 2012.

[19] M. C. Gonzalez, C. A. Hidalgo, and A.-L. Barabasi. Understanding individual human mobility patterns. *Nature*,

453(7196):779–782, 2008.

[20] A. Graves. Generating sequences with recurrent neural networks. *arXiv preprint arXiv:1308.0850*, 2013.

[21] Y. Kim, F. C. Pereira, F. Zhao, A. Ghorpade, P. C. Zegras, and M. Ben-Akiva. Activity recognition for a smartphone based travel survey based on cross-user history data. In *Pattern Recognition (ICPR), 2014 22nd International Conference on*, pages 432–437. IEEE, 2014.

[22] J. Krumm and E. Horvitz. Predestination: Inferring destinations from partial trajectories. In *International Conference on Ubiquitous Computing*, pages 243–260. Springer, 2006.

[23] K. Lee, S. Hong, S. J. Kim, I. Rhee, and S. Chong. Slaw: A new mobility model for human walks. In *INFOCOM 2009, IEEE*, pages 855–863. IEEE, 2009.

[24] L. Liao, D. Fox, and H. Kautz. Location-based activity recognition. *Advances in Neural Information Processing Systems*, 18:787, 2006.

[25] L. Liao, D. Fox, and H. Kautz. Extracting places and activities from gps traces using hierarchical conditional random fields. *The International Journal of Robotics Research*, 26(1):119–134, 2007.

[26] Z. Lin, M. Yin, S. Feygin, M. Sheehan, J.-F. Paiement, and A. Pozdnoukhov. Deep generative models of urban mobility. 2017.

[27] A. Monreale, F. Pinelli, R. Trasarti, and F. Giannotti. Wherenext: a location predictor on trajectory pattern mining. In *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 637–646. ACM, 2009.

[28] S. Phithakkitnukoon, T. Horanont, G. Di Lorenzo, R. Shibasaki, and C. Ratti. Activity-aware map: Identifying human daily activity pattern using mobile phone data. In *International Workshop on Human Behavior Understanding*, pages 14–25. Springer, 2010.

[29] S. Reddy, M. Mun, J. Burke, D. Estrin, M. Hansen, and M. Srivastava. Using mobile phones to determine transportation modes. *ACM Transactions on Sensor Networks (TOSN)*, 6(2):13, 2010.

[30] S. R. Safavian and D. Landgrebe. A survey of decision tree classifier methodology. *IEEE transactions on systems, man, and cybernetics*, 21(3):660–674, 1991.

[31] S. Scellato, M. Musolesi, C. Mascolo, V. Latora, and A. T. Campbell. Nextplace: a spatio-temporal prediction framework for pervasive systems. In *International Conference on Pervasive Computing*, pages 152–169. Springer, 2011.

[32] C. M. Schneider, V. Belik, T. Couronné, Z. Smoreda, and M. C. González. Unravelling daily human mobility motifs. *Journal of The Royal Society Interface*, 10(84):20130246, 2013.

[33] C. Song, Z. Qu, N. Blumm, and A.-L. Barabási. Limits of predictability in human mobility. *Science*, 327(5968):1018–1021, 2010.

[34] L. Song, D. Kotz, R. Jain, and X. He. Evaluating location predictors with extensive wi-fi mobility data. In *INFOCOM 2004. Twenty-third AnnualJoint Conference of the IEEE Computer and Communications Societies*, volume 2, pages 1414–1424. IEEE, 2004.

[35] X. Song, H. Kanasugi, and R. Shibasaki. Deeptransport: Prediction and simulation of human mobility and transportation mode at a citywide level. IJCAI, 2016.

[36] L. Stenneth, O. Wolfson, P. S. Yu, and B. Xu. Transportation mode detection using mobile phones and gis information. In *Proceedings of the 19th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, pages 54–63. ACM, 2011.

[37] A. Thiagarajan, L. Ravindranath, K. LaCurts, S. Madden, H. Balakrishnan, S. Toledo, and J. Eriksson. Vtrack: accurate, energy-aware road traffic delay estimation using mobile phones. In *Proceedings of the 7th ACM Conference on Embedded Networked Sensor Systems*, pages 85–98. ACM, 2009.

[38] E. I. Vlahogianni, M. G. Karlaftis, and J. C. Golias. Short-term traffic forecasting: Where we are and where were going. *Transportation Research Part C: Emerging Technologies*, 43:3–19, 2014.

[39] P. Widhalm, P. Nitsche, and N. Brändie. Transport mode detection with realistic smartphone sensor data. In *Pattern Recognition (ICPR), 2012 21st International Conference on*, pages 573–576. IEEE, 2012.

[40] P. Widhalm, Y. Yang, M. Ulm, S. Athavale, and M. C. González. Discovering urban activity patterns in cell phone data. *Transportation*, 42(4):597–623, 2015.

[41] J. Ye, Z. Zhu, and H. Cheng. What's your next move: User activity prediction in location-based social networks. In *Proceedings of the 2013 SIAM International Conference on Data Mining*, pages 171–179. SIAM, 2013.

[42] M. Yin, M. Sheehan, S. Feygin, J.-F. Paiement, and A. Pozdnoukhov. A generative model of urban activities from cellular data. *IEEE Transactions in ITS*, 2017.

[43] J. J.-C. Ying, W.-C. Lee, T.-C. Weng, and V. S. Tseng. Semantic trajectory mining for location prediction. In *Proceedings of the 19th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, pages 34–43. ACM, 2011.

[44] J. Zheng and L. M. Ni. An unsupervised framework for sensing individual and cluster behavior patterns from human mobile data. In *Proceedings of the 2012 ACM Conference on Ubiquitous Computing*, pages 153–162. ACM, 2012.

[45] Y. Zheng, Q. Li, Y. Chen, X. Xie, and W.-Y. Ma. Understanding mobility based on gps data. In *Proceedings of the 10th international conference on Ubiquitous computing*, pages 312–321. ACM, 2008.

[46] Y. Zheng, L. Liu, L. Wang, and X. Xie. Learning transportation mode from raw gps data for geographic applications on the web. In *Proceedings of the 17th international conference on World Wide Web*, pages 247–256. ACM, 2008.

[47] X. Zhu. Semi-supervised learning literature survey. 2005.