Latent Allocation Spatiotemporal Models For Indoor Human Mobility

Yiming Gu, Hala Mostafa
United Technologies Research Center
gujack,mostafh@utrc.utc.com

B. Aditya Prakash
Department of Computer Science, Virginia Tech
badityap@cs.vt.edu

ABSTRACT

The ability to understand, model and predict human mobility in high-occupancy buildings like corporate offices and campuses can fundamentally change the way buildings are managed. For example, energy efficiency can be improved using more accurate models of the temporal and spatial aspects of building occupancy. Similarly, responding to emergency situations is more effective and less intrusive if the building system has better knowledge not just of where occupants are, but also of their likely next locations and when they will get there. We propose a novel approach to learn a spatiotemporal model of human mobility from observed trajectories. Our approach posits the existence of different *mobility profiles* that reflect the heterogeneity in the way people move between locations. Our proposed latent allocation model describes the probabilistic relationships between the observed trajectory data and the latent (unobserved) mobility profiles and their parameters. To tackle the problem of inferring these parameters efficiently, we frame the model as a neural network. We engineer the layers of the network to enforce appropriate constraints on the learned spatial and temporal parameters of each profile to best explain the data. We demonstrate our model and learning approach on synthetic data and give initial results from a subset of real data collected from our corporate building.

ACM Reference Format:
Yiming Gu, Hala Mostafa and B. Aditya Prakash. 2018. Latent Allocation Spatiotemporal Models For Indoor Human Mobility . In *UrbComp 2018, August 2018, London, United Kingdom.* ACM, New York, NY, USA, Article 4, 9 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

The study of human mobility has found new sources of data with the proliferation of smart phones, Call Detail Records and wireless devices that can be used to track users [2, 12, 14]. Complementing this broader set of data sources is an increasingly broader set of needs in various domains that can benefit from mobility models. For outdoor city- or campus-scale mobility, approaches for smarter transportation and urban planning efforts can use mobility models to design more streamlined cities. Data from from taxi rides and GPS data can also be useful for next location prediction, which can inform resource management policies on the part of taxi companies and municipalities [28]. In commercial spaces, taxi ride data and Point of Interest data can be combined to infer ride purpose, which can help match supply and demand for different types of commercial spaces and services [34].

For indoor settings, better understanding and models of human mobility can enable a range of improvements in the energy efficiency of buildings, their ease of use and their safety in emergency situations. Predictive models can help a building management system plan HVAC settings and perhaps shape the building's energy load to improve the efficiency and cost of the the building operation [16].

Researchers interested in better understanding human mobility include civil engineers and urban planners trying to understand mobility in cities, social scientists trying to understand interactions among people in public places and how it relates to their interactions in virtual environments, wireless network researchers using mobility and wireless usage data to design better wireless communication protocols, and HVAC engineers trying to design systems with higher energy efficiency and user comfort. These disciplines study and model human mobility at different granularities. For example, occupancy models for buildings are usually aggregated at the level of a room or an HVAC zone. Taxi ride and vehicular GPS data capture occupancy at the level of regions in a city.

In this paper, we start exploring the use of data from card readers in an office building to learn spatio-temporal models of human mobility. Our setting exhibits considerable heterogeneity in the way people move in a building. For example, consider the following *mobility profiles*:

(1) An employee who spends most of his time in his office, except for a few hours spent in meeting rooms each week week and daily visits to the cafeteria.
(2) A high-level manager who is almost always attending meetings, which tend to last about an hour (his time is valuable) and be in various parts of the building (his responsibilities are broad and involve multiple departments).
(3) A janitorial staff member who is responsible for a given building and makes the rounds of this building, visiting each room, but spending little time there.

We view human mobility as a flow-conserving diffusion over the graph whose nodes are the zones of the building accessible through card readers. While various approaches have been proposed to model diffusion over graphs in general and human mobility in particular [2, 18, 25, 28, 31, 33], the state of the art has 3 main shortcomings in the context of our setting:

(1) Existing approaches do not specifically cater to the heterogeneity in mobility profiles in our setting. They typically characterize a homogeneous population of moving objects,

UrbComp 2018, August 2018, London, United Kingdom

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM.
https://doi.org/10.1145/nnnnnnn.nnnnnnn

but allow for variations in mobility by time of day and other features.

(2) Much of the existing spatio-temporal models assume the spatial and temporal aspects are two sides of the same coin where the time and location of an entity's next moves are interdependent [31, 33]. This is a direct consequence of using the exponential distribution for the duration of stay, which essentially ties the spatial and the temporal aspects of mobility.
(3) When using features, existing approaches attach features to nodes and assume high mobility between similar nodes [33], which is not necessarily true in our setting. For example, people do not necessarily move to a conference room after being in one.

To address these shortcomings, we propose a spatiotemporal human mobility model that:

(1) Uses latent variables to posit the existence of multiple *mobility profiles* and the parameters of each.
(2) Separately models the spatial and temporal aspects of mobility for each profile.
(3) Attaches features to transitions, rather than individual nodes, for a richer model.
(4) Uses a neural network formulation to learn the latent parameters.

Our work is inspired by machine learning for natural language processing where latent (hidden) variables represent underlying topic models and each topic gives rise to a distribution over the observed data (words in a corpus). Similarly, we posit the existence of mobility profiles, with each profile responsible for spatio-temporal patterns in mobility.

We present an inference approach to learn the parameters of our model from observed timestamped trajectories. We presents preliminary results on synthetic data, as well as data obtained from access control devices (card readers) in our office building.

This paper is organized as follows: Section 2 reviews related work in the areas of human mobility modeling and diffusion modeling. In Section 3, we give background of specific the works that we base this paper on, present our latent allocation model and explain the inference algorithm we propose to learn it. Section 4 shows our initial experimental results. We conclude and outline the remaining work we will do in Section 5.

2 RELATED WORK

2.1 Human mobility

2.1.1 Settings and goals. The body of work devoted to studying human mobility is larger than can be summarized in this section, so we will present a sample thereof and focus on works that are closer to ours in their goals or approaches.

A large portion of studies of human mobility addresses mobility in cities using data from taxi rides and GPS data. Modeling human mobility for next location prediction can inform resource management policies on the part of taxi companies and municipalities [28]. Human mobility data can also be combined with POI data to infer the purpose of a trip (e.g., shopping, dining, work) [34], which can be useful in planning urban development and commercial space purpose and location. Call Detail Records and smart phone data have also been used to construct mobility models [2, 12, 14]. The insight obtained from these studies can be used to construct more effective policies for crowd control in emergency situations, as well as planning the timing and location of road work and detours.

With the advent of mobile devices, wireless communications are providing increasingly large amounts of data that sheds light on human mobility at sub-city levels like university campuses [18, 26]. In these kinds of studies, the improved understanding of human mobility is a means towards an end; the studies are often conducted by researchers from the wireless networks community who seek to design better communication protocols and infrastructure. A smaller number of papers focuses on human mobility indoors (e.g., in office buildings [5, 23]). One of the main motivations behind studying indoor mobility is to improve energy efficiency of office and residential buildings. Numerous occupancy models have been proposed, but most of them operate at the aggregate level, where they model total occupancy at the level of a room or region without regard to modeling individual trajectories [15, 16, 20].

2.1.2 Approaches. **Non-modeling studies:** Many studies of human mobility are descriptive in nature, especially work using wireless network data. A dataset is analyzed and the salient features therein are discussed, but without proposing a model that fits the observed mobility patterns. One example is the work done on understanding the persistence and prevalence of mobility in office buildings [5]. Persistence reflects session durations whereas prevalence reflects the frequency with which users visit various locations. The work finds that the probability distributions of both measures follow power laws. Another example are works that describes different use loads and traffic categories of a campus wireless network over time [18].

Another descriptive work explores the predictability of whether two persons will interact by analyzing tracing data from academic and office environments [26]. They collect ground truth of social structures and use smart phone data to shed light on mobility by consider which wireless access points were visible to a phone at a given time. Using "virtual locations" derived from these access points, they generate a sequence of locations per user. They discover communities that resemble the communities they found in the social structure data.

Modeling studies: For efforts that aim to construct predictive mobility models, some only focus on the spatial aspect, where the goal is to predict the next location of a user, while others focus on temporal models that predict whether a user will remain in her current location for the next t minutes. More specific temporal models attempt to predict the time at which the user will transition to a new location.

One mobility model is the Trajectory Patterns Tree, a type of decision tree that predicts the next location of a trajectory by finding the path in the tree that best matches the locations on the trajectory so far [28]. Learning depends on the movement of all available objects in a certain area instead of on the individual history of the owner of the trajectory only. This way, data is leveraged for learning across moving objects, but without attempting to only leverage data from objects closet in behavior to the current object.

Prediction of a user's next location has also been cast as a classification problem where the covariates include location, time of arrival, previous location and previous duration of stay. Even the temporal aspect of mobility was cast as a classification problem, where a binary classifier learns a mapping from the predictive features to 0 or 1 indicating whether the person will move in next *n* minutes [23].

Another work that tries to build a model of next location uses dynamic Bayesian networks (DBNs) [12]. The random variables in the DBNs include location, day of the week, and time of day. The work uses two mobility data sets; a Call Detail Records data and a Nokia mobile data which is based on GPS. The authors address the shortcomings of having a single mobility model by proposing three models and using different approaches to combine them.

The idea of increasing prediction accuracy by leveraging mobility data of a person's social acquaintances was demonstrated on the Nokia Data Challenge [13]. The assumption is that especially on university campuses, people in the same social circle will tend to visit the same locations at the same times (e.g., having meals together, or studying in the library at the same time).

This idea was further explored in order to leverage data from people who are not necessarily within the social circle of the user we are modeling. The authors use spatial and temporal similarity measures to find those users, which they call *similar strangers* [2]. A Dynamic Bayesian Network is proposed to capture the dependency of the next location on a number of features, which incorporates aspects of the mobility patterns of similar strangers. They demonstrate their approach on CDR data.

In pedestrian movement prediction [4], individuals are clustered into groups based on their mobility traces. The approach learns a Markov model for each group. To predict a person's next location, the approach identifies the group they belong to and infers the next location based on this group's model.

Like our approach, prior work has looked into both the spatial and temporal modeling of mobility, but using smart phone data. The authors propose a model that attempts to find relevant contextual features that help in predicting the next location and duration of stay [14]. They use ensemble methods to combine models that use different contexts, where these models describe different conditional distributions over the location/duration given contextual variables.

In summary, surveying the state of the art in approaches to modeling human mobility, we find the following alternatives:

- Dynamic Bayesian networks based on individual features [12].
- Dynamic Bayesian networks with locations of similar strangers [2].
- Classifiers for predicting whether a person will move in the next *t* time steps, and which if *n* locations they will move to (SVN, KNN, DT) [23]
- Association rules + matching function to find a matching rule [29]. All the matching functions are based on support and confidence and do not consider spatial and temporal distance.
- Clustering trajectories using some notion of trajectory similarity (like a matching function), which is typically distance-based.
- Decision tree for mining trajectory patterns [28].

2.2 Diffusion models

Studying the progress made in learning diffusion models is very relevant to human mobility, since the movement of humans can be seen as diffusion over a graph whose nodes are the different locations. As with human mobility, learning a diffusion model involves learning a model of the spatial and temporal behavior of the spreading phenomenon.

Learning the diffusion process or the structure of the graph on which a phenomenon spreads has many applications in information diffusion [17, 31], epidemiology [11, 24] and social networks analysis [10, 35]. Again as in human mobility, learning is based on (possibly partial) observations of trajectories; sequences of where the phenomenon has been and when.

Approaches for diffusion learning often do not fully exploit the rich features associated with each transition in a trajectory. In human mobility, these features can include the type of the edge (road or corridor), features of the source and destination locations, and features of the person making the transition (e.g., department and title of the employee in an office building). Works on feature-based diffusion do consider features of the source and destination, but only in terms of the difference between their feature vectors, assuming that the smaller the distance, the more likely an entity is to make the transition [33]. While this approach reflects the high probability of memes and retweets spreading among people with similar interests in social networks, this assumption does not make sense for human mobility, since humans move among locations with different functions and features to fulfill their needs. For example, in an office building, people are not necessarily likely to go to another conference room after being in one.

2.3 The spatial and the temporal

Much of the earlier work studying how phenomena spread or physical objects move were restricted to using discrete time, where the object or phenomenon moves from one location to the next in unit time and immediately moves on. As such, these models are oblivious to the notion of *duration of stay* and only model the spatial aspect of the spread. Examples of these models include Markov chains and a line of diffusion modeling work using independent cascade (IC) models [17]. Data used to learn these *spatial models* consists of trajectories or sequences of cascades which list the locations visited by the phenomenon in the order they were visited.

These earlier models were extended to capture and account for time spent at a location and travel time. These *spatiotemporal models* include continuous time Markov chains (CTMC) and spatiotemporal IC models [31]. They learn from data consisting of time-stamped trajectories or sequences of cascades which list the locations visited and the time at which the visits happened.

The above models focus on the propagation of natural phenomena and phenomena taking place in social networks, like the spread of memes or news. They typically uses an exponential distribution to model the duration of stay at a given location [3, 33]. In these domains, there are two main assumptions or characteristics that make the exponential distribution very suitable:

- The remaining amount of time an object spends at a location is independent of the amount of time already spent. In other words, the distribution of the duration stay is *memoryless*.
- The higher the propagation between two locations i and j, the shorter the duration of stay of an object at i before spreading to j.

When studying human mobility, the above assumptions are questionable because unlike the phenomena studied above, the traveling entities here are humans that make *active decisions* where to go next, unlike a contagion of viruses or the spread of a rumor. For example, the amount of time a person spends in a meeting room is typically very dependent on how much she has been there already; people tend to meet for an hour and if 20 minutes have passed, the distribution over the remaining time should peak around 40 minutes, or can even be bimodal with another peak at 100 minutes to reflect 2-hour meetings.

Similarly, the second assumption is more appropriate in social networks where the fact that person i forwards a lot of content to her friend j, then i tends to do this fairly quickly after i herself receives this content. In human mobility however, a person may be in the habit of always heading for a coffee machine located near a certain meeting room after a meeting, but that does not imply that he moves to the coffee machine location soon after he reaches the meeting room. The model of the expiring alarm clocks that is inherent in exponential distributions of stay is thus broken.

3 LATENT ALLOCATION MOBILITY MODEL

3.1 Problem formulation

We propose a spatiotemporal model to learn human mobility in office buildings based on observed timestamped trajectories obtained from access control devices like card readers.

We model the problem as movement on a graph $G = (V, E)$. We model the spatial aspect of mobility using the matrix $\mathbf{A} = \{\alpha_{jk} | j, k \in V, j \neq k\}$, where α_{jk} is the probability of moving to k given that the person is in j. We model the temporal aspect of mobility using the matrix $\Sigma = \{\sigma_{jk} | j, k \in V, j \neq k\}$, where σ_{jk} is the parameter of the Rayleigh distribution which describes the duration of stay of a person in room j before moving on to room k. We give more details on how these 2 parameters interact later in this section, but note the *independence of the spatial and temporal distributions describing mobility between two nodes.*

The data we learn our model from is in the form of timestamped trajectories, or *cascades*:

$$\pi^c = \{(v_0^c, t_0^c, f_0^c), (v_1^c, t_1^c, f_1^c), ..., (v_{N_c}^c, t_{N_c}^c, f_{N_c}^c)\} \quad (1)$$

where π^c is the c-th cascade, t_i^c is the time at which the entity moved to location v_i^c, N_c is the length of cascade c and f_i^c is the feature vector associated the transition from v_i^c at time t_i^c. One option for f_i^c is the feature vector of the node v_i^c, although a richer feature vector can be used.

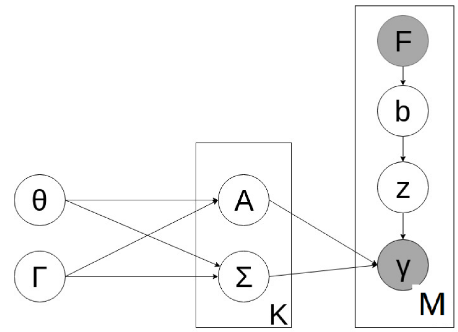

Figure 1: Plate notation for the graphical model of transitions from K mobility profiles. M is the number of transitions per trajectory. Shaded nodes are observed while clear nodes are latent (hidden).

Following the Independent Cascade assumption, the likelihood of the observed data is

$$P(\pi | G; A, \Sigma) = \prod_{c=1}^{C} P(\pi^c | G; A, \Sigma). \quad (2)$$

where C is the number of observed cascades.

To learn a mobility model most consistent with the data, we solve the maximum likelihood estimation problem

$$A^*, \Sigma^* = \arg\max \prod_{c=1}^{C} P(\pi^c | G; A, \Sigma) \quad (3)$$

We now write the specific form of $P(\pi^c | G)$. Assuming each transition is independent of other transitions given its own features and parameters, we have:

$$P(\pi^c | G; A, \Sigma) = \prod_{i=1}^{N_c - 1} P(\gamma_i^c | G; A_{i,i+1}, \Sigma_{i,i+1})$$

where $P(\gamma_i^c | G)$ is the probability of transition γ_i^c which is a tuple $(v_i^c, t_i^c, v_{i+1}^c, t_{i+1}^c, f_i^c)$.

3.2 Latent allocation model for mobility profiles

Our approach posits that the heterogeneity observed in how people move in a building arises from the existence of multiple underlying *mobility profiles*, each giving rise to a distribution over transitions in a trajectory.

We propose a graphical model with latent variables that are similar to the notion of *topics* in the well-known Latent Dirichlet Allocation (LDA) model used extensively in natural language processing [8]. In our setting, we assume that each of our K different "topics" represents a mobility profile that has an associated set of spatial and temporal parameters.

The generative process that generates transitions can be described as a graphical model as shown in Figure 1. The process initially generates distributions $[A, \Sigma] \sim \mathcal{N}(\theta, \Gamma)$ where $\mathcal{N}(\theta, \Gamma)$ is a normal distribution. θ and Γ are hyper-parameters for the spatial transition parameter A and temporal transition parameter Σ.

Table 1: Elements of our latent allocation model. Realizations of F and Y are observed.

θ, Γ	Hyperparameters for the spatial and temporal distributions
$A_k \in R^{N \times N}$	Matrix of transition probabilities of mobility profile k (spatial)
$\Sigma_k \in R^{N \times N}$	Matrix of Rayleigh parameters of mobility profile k (temporal)
$F \in R^d$	Feature space of dimension d
$W \in R^{K \times d}$	Weights per feature per topic
$b \in R^K$	Allocation vector (sums to 1)
$z \in [1..K]$	Sampled topic/mobility profile
$Y \in V^2 \times R^2$	Realizations of transitions of the form (v_i, t_i, v_j, t_j)
K	Number of mobility profiles
N	Number of nodes/locations
M	Number of transitions in a cascade

For each transition from location v_i that was reached at time t_i with feature f_i, the generative process is as follows:

(1) Generate $b \sim S(f_i^T w)$, where $b_j = S_j(f_i) = \frac{e^{f_i^T w_j}}{\sum_k e^{f_i^T w_k}}$, $j \in [1..K]$ and $S_j(f_i)$ is a softmax function. Notice that $S(f_i)$ can be viewed as a neural embedding [21] for feature f_i and could be a deep neural network as long as the final layer is a softmax function. b is thus the allocation vector describing the probability of allocation of the transition to each topic or mobility profile.
(2) Sample the topic/mobility profile $z \sim multi(b)$ where $multi(b)$ is a single draw multinomial distribution parametrized by vector b.
(3) Given topic z, matrices A_z and Σ_z are the generating parameters for the spatial and temporal aspects of the transition, respectively.
(4) Sample the next node to visit $v_j \sim multi(A_{z,v_i})$, where A_{z,v_i} is the v_i-th row of A_z.
(5) Sample $d_{ij} \sim Ray(\Sigma_{z,v_i,v_j})$ where $Ray(\sigma)$ is the Rayleigh distribution

$$Ray(x; \sigma) = \frac{x}{\sigma_2} e^{-\frac{x^2}{2\sigma^2}}$$

The arrival time at location v_j is $t_i + d_{ij}$.
(6) The transition y_i is then $(v_i, t_i, v_j, t_j), f_i)$.

The likelihood of a trajectory π^c according to the above generative process is therefore:

$$P(\pi^c|G) = \prod_{k=1}^{K} N(A_k, \Sigma_k; \theta, \Gamma)$$
$$\prod_{m=1}^{M} S(b_m|f_m) \;\; \text{multi}(z_m = k|b_m) \;\; \text{multi}(v_{m+1}|A_{z_m,v_m})$$
$$Ray(t_{m+1} - t_m|\Sigma_{z_m,v_m,v_{m+1}}) \quad (4)$$

where $S()$ is a softmax function, multi() is a multinomial function, $\Sigma_{k,i,j}$ is the ij entry of matrix Σ_k and $Ray()$ is the Rayleigh distribution.

3.3 The inference

We now outline the inference algorithm for learning the parameters of the proposed model. We use variational inference, which originated in physics [6] and was first introduced by Michael Jordan [22] to the machine learning field. Mean field variational inference is widely applied in latent variable graphical models such as LDA [8].

Denote the unknown latent variables collectively by $L = A, \Sigma, z, W$ and the observation data by π. The Bayesian inference problem for finding the parameters in Equation 4 is:

$$P(L|\pi) = \frac{P(\pi|L)P(L)}{\int_L P(\pi|L)P(L)}$$

It is obvious that the integral in the denominator is intractable for any reasonable sized problem, since the overall size of the latent variable is $2 * N^2 + M$, where N is the number of nodes and M is the number of observed transitions. Instead of computing $P(L|\pi)$ directly, variational inference aims to compute an alternative variational distribution $q(L;\phi)$. The optimal parameter of the variational distribution is the one that minimizes the KL-divergence between the original posterior $P(L|\pi)$ and the variational distribution $q(L;\phi)$:

$$q^*(L;\phi) = \arg\min_{\phi} \text{KL}(q(L;\phi)||P(L|\pi)) \quad (5)$$

Note that in variational inference, instead of minimizing the KL-divergence directly (which is still intractable), we minimize the equivalent term:

$$\mathbb{E}_q(\log(q(L;\phi))) - \mathbb{E}_q(\log(p(L,\pi))) \quad (6)$$

since

$$KL(q(L;\phi)||P(L|\pi)) = \mathbb{E}(\log(q(L;\phi))) - \mathbb{E}(\log(p(L,\pi))) + \log(p(\pi))$$
$$\geq \mathbb{E}(\log(q(L;\phi))) - \mathbb{E}(\log(p(L,\pi))) \quad (7)$$

which is the original KL-divergence minus a constant with respect to q.

A transformation of Equation 6 shows that

$$\mathbb{E}_q(\log(q(L;\phi))) - \mathbb{E}_q(\log(p(L,\pi))) =$$
$$KL(q(L)||p(L)) - \mathbb{E}[log(p(\pi|L))] \quad (8)$$

The first term indicates that the optimized variational distribution should be close to the prior, and the second term indicates that it should put its mass on the configurations of latent variables that explains the data. Thus variational inference encodes the trade-off between likelihood and prior knowledge commonly seen in Bayesian statistics.

Our Bayesian inference problem is therefore transformed to a two step task:

(1) Specify a class of variational distributions $q(L|\phi)$.
(2) Solve the optimization problem in Equation 6.

3.4 Function approximation using neural networks

Neural Networks have been successfully applied as universal function approximators [19] to various functions such as the classification function in object detection [30], the regression function in sequence-to-sequence translation [32], and probability density functions [27] [7].

In this paper, we assume the variational distribution is approximated by a feed-forward neural network:

$$q(A, \Sigma, b, W; \phi) = Q_A(A; \phi_A, I_A) Q_\Sigma(\Sigma; \phi_\Sigma, I_\Sigma) Q_b(b; \phi_b, F, \Sigma_b, W)$$

The details of the neural network are shown in Figure 2, where the layers of the network are described as follows:

(1) Input layer: I_Σ and I_A are identity vectors, and F is a matrix of training features. $\{I_\Sigma, I_A, F\}$ act as input for the feed-forward neural network.
(2) Fully connected layers: The input $\{I_\Sigma, I_A, F\}$ goes through three fully connected layers. Since the spatial transition parameter A and and temporal transition parameter Σ are independent, using three distinct fully connected layers reduces the parameter dimension and increases scalability.
(3) Intermediate output: h_Σ, h_A, h_b are the unconstrained outputs of the fully connected layers.
(4) Variable-specific layers: these layers encode assumptions or constraints on the variables Σ, A, b. For example, for the constraint $\Sigma > 0$, we use an exponential activation function, and for the constraint that rows of A are in the interval $[0, 1]$ and sum to 1, we use a softmax activation function.
(5) Latent variable output: these are the parameters of latent variables Σ, A, b. If Σ, A, b are assumed to be point-wise values with probability 1, these nodes can indicate these value themselves. Or if Σ, A, b are assumed to be distributions, these nodes can indicate the parameters of the distribution.
(6) Loss function: the loss function as defined in Equation 6. In addition, in the implementation, we also use: (1) the L_2 regularization on the weights of the neural network; (2) the cross entropy of the distribution on the parameters of different mobility profiles, based on the assumption that the mobility profiles should be distinctive.

Now the inference is reduced to finding the optimal parameters in the above neural network to minimize Equation 6. This can be done using first-order optimization schemes such as Stochastic Gradient Descent [9] very effectively in Tensorflow at scale [1].

4 EXPERIMENTS

4.1 Simulated dataset

We now present experimental results using synthetic data generated from a hypothetical office building with 5 zones as shown in Figure 3. Node 0 is the building entrance and nodes 3 and 4 are exits. There are two mobility profiles; one profile dominates in bad weather, such as wind and rain, where people tend to heavily use the link $1 \longrightarrow 3$ and $2 \longrightarrow 4$. The other is a fair weather profile where people tend to heavily use the link $1 \longrightarrow 4$ and $2 \longrightarrow 3$ to enjoy the sunshine. Additionally, we assume that in good weather, people tend to stay in the office for shorter periods than in bad weather. For simplicity, we assume that this binary weather is the only feature (i.e., $F \in \{0, 1\}$).

Given a binary feature, the sizes of the latent variable are as follows:

(1) For each of the 2 mobility profiles, the spatial transition matrix A_k is 5×5
(2) For each of the 2 mobility profiles, the temporal transmission matrix Σ_k is 5×5
(3) Number of latent variables z is M, where M is the total number of transitions in the data and each transition is allocated 1 of K mobility profiles.
(4) The softmax parameter W to map from F to b is 2×2
(5) The latent variable vector b is of size 2.

Figures 4 and 5 show the true and inferred A and Σ for each mobility profile. As the figures show, our inference algorithm is able to recover the spatial transition parameters fairly accurately, inferring the higher tendency to use edges $1 \longrightarrow 3$ and $2 \longrightarrow 4$ in one mobility profile and edges $1 \longrightarrow 4$ and $2 \longrightarrow 3$ in the other.

4.2 Access control dataset

We now report on our work in progress in applying our approach to a real dataset. We obtained an access control (AC) dataset consisting of card swipe logs collected by card readers in our facility. Employees have cards enabling them to access different areas of the facility, with some areas not subject to access control (freely accessible without needing to swipe). For each card swiped at a reader, the reader logs the time of swipe, card ID and whether access was granted. The card ID links to the owner's information in an employees database.

This type of access control dataset poses some interesting challenges and opportunities. One of the challenges is that a given card reader can control access to locations with a mix of very different uses like offices and conference rooms. This means that from a card reader reflects usage patterns of different types of rooms and can result in very heterogeneous spatial and temporal mobility patterns.

Another challenge is that in our setting, employees must swipe into an AC areas, but they do not swipe out. So if a person transitions from an AC area to a non-AC area, this move is not logged, and in building the graph representation, the move appears as a self-loop. This is equivalent to having partially observable data where the exact time a person left the AC area is unknown.

That being said, access control data, when collected, is cleaner and more accurate than location data inferred from wireless network access points, as is common in indoor human mobility literature. [2, 18, 26].

To conduct experiment, we selected a small subset of 5 card readers in our facility. Figure 6 shows preliminary results of inferring A_k from our AC dataset using the time of day as a binary feature (morning vs. afternoon). As can be seen, this feature is not appropriate for uncovering mobility profiles and the inferred parameters for both profiles are not very distinct.

Since the AC data is linked to the employee data that contains information like the employee's department, job title and office location, we see a lot of potential for using a much richer feature space that will allow the model to uncover and learn interesting mobility profiles.

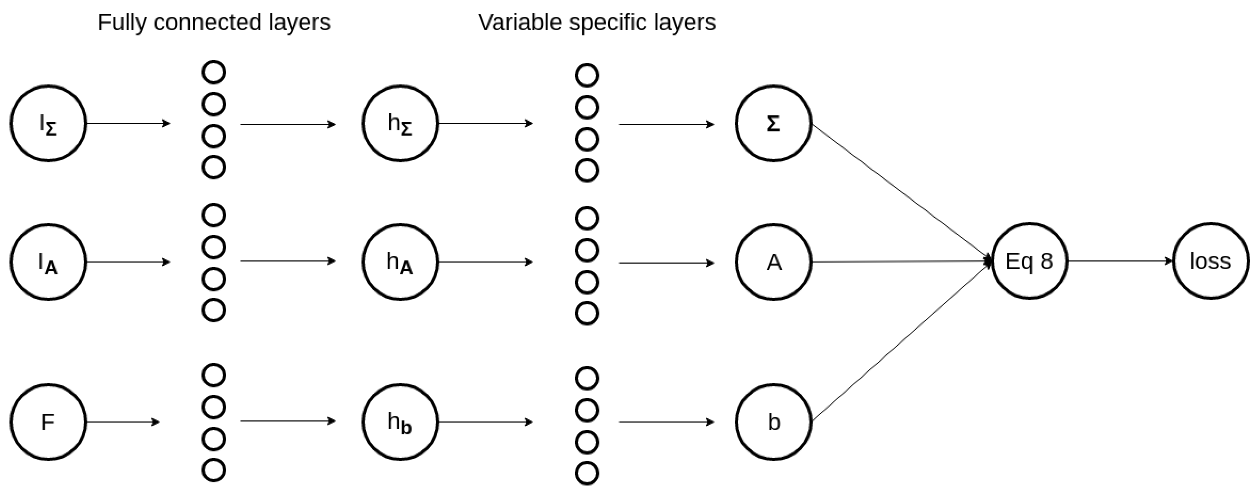

Figure 2: Neural network structure for learning the constrained latent variables A, Σ, b.

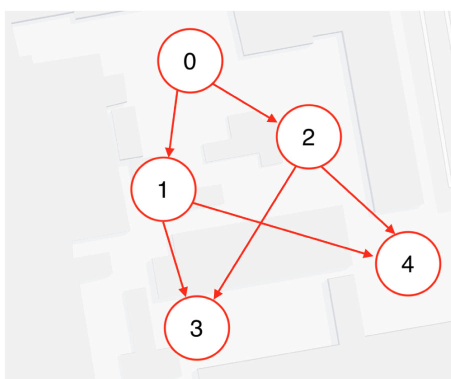

Figure 3: Synthetic building layout.

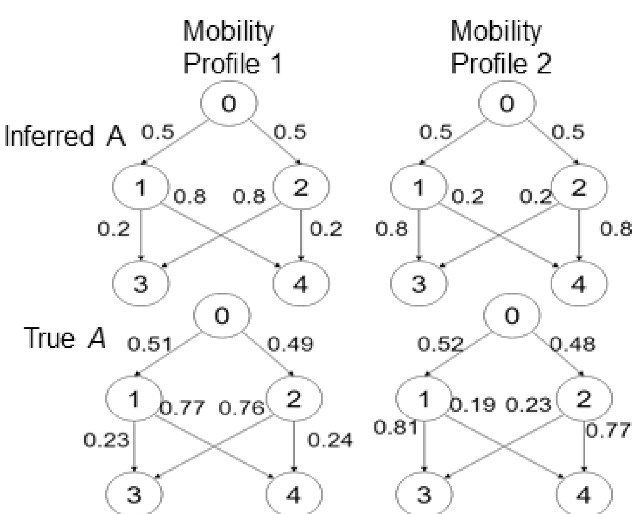

Figure 4: Inferred vs. true A_k for 2 mobility profiles using synthetic data.

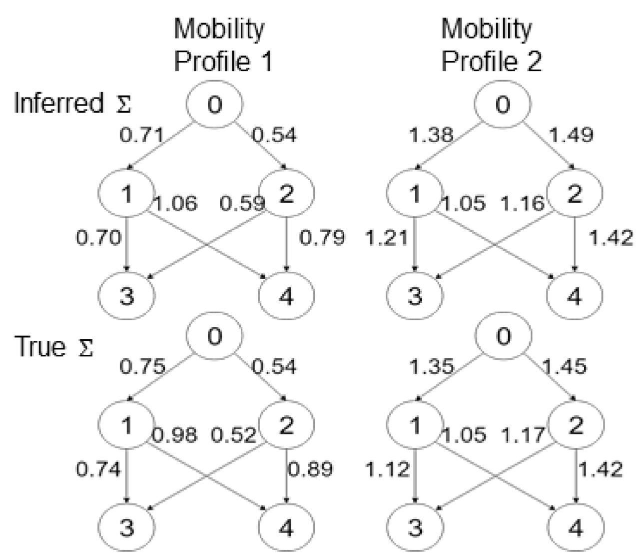

Figure 5: Inferred vs. true Σ_k for 2 mobility profiles using synthetic data.

We also see room for improvement in the model itself. We can add a *diversity term* to the objective function that encourages parameters of the different mobility profiles to be different in order to yield more distinct, and hopefully more interpretable profiles.

5 CONCLUSION

Being able to understand, model and predict human mobility in large buildings like corporate offices can have fundamental effects on the way these buildings are managed in terms of energy efficiency, emergency response, and other aspects. We propose a novel spatiotemporal model of human mobility from observed trajectories. Our approach posits the existence of different mobility profiles that reflect the heterogeneity in the way people move between locations

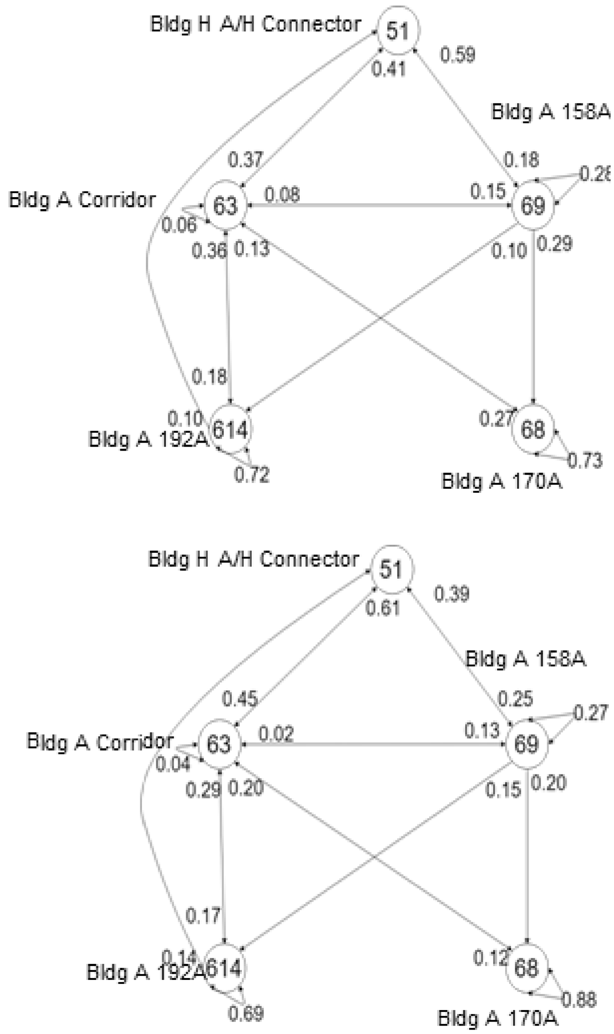

Figure 6: Inferred A_k for 2 mobility profiles using real access control data.

in a building. To learn the parameters of each mobility profile, we frame the model as a neural network and leverage the powerful existing machinery for neural networks. We demonstrate how our model and learning approach recover spatial and temporal mobility parameters from synthetic data. We also present our access control dataset from our corporate building, give sample initial results and outline the next steps of this work in progress.

6 ACKNOWLEDGMENT

This paper is based on work partially supported by the NSF (CAREER IIS-1750407), the NEH (HG-229283-15), ORNL (Task Order 4000143330) and a Facebook faculty gift.

REFERENCES

[1] Martín Abadi, Paul Barham, Jianmin Chen, Zhifeng Chen, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Geoffrey Irving, Michael Isard, et al. 2016. TensorFlow: A System for Large-Scale Machine Learning.. In *OSDI*, Vol. 16. 265–283.

[2] Fahad Alhasoun, May Alhazzani, Faisal Aleissa, Riyadh Alnasser, and Marta GonzÃąlez. 2017. City scale next place prediction from sparse data through similar strangers. In *Proceedings of Urban Computing Workshop.*

[3] William J Anderson. 2012. *Continuous-time Markov chains: An applications-oriented approach.* Springer Science & Business Media.

[4] Akinori Asahara, Kishiko Maruyama, Akiko Sato, and Kouichi Seto. 2011. Pedestrian-movement prediction based on mixed Markov-chain model. In *Proceedings of the 19th ACM SIGSPATIAL international conference on advances in geographic information systems.* ACM, 25–33.

[5] Magdalena Balazinska and Paul Castro. 2003. Characterizing mobility and network usage in a corporate wireless local-area network. In *Proceedings of the 1st international conference on Mobile systems, applications and services.* ACM, 303–316.

[6] Rodney J Baxter. 2016. *Exactly solved models in statistical mechanics.* Elsevier.

[7] Christopher M Bishop. 1994. Mixture density networks. (1994).

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

[9] Léon Bottou. 2010. Large-scale machine learning with stochastic gradient descent. In *Proceedings of COMPSTAT'2010.* Springer, 177–186.

[10] Simon Bourigault, Cedric Lagnier, Sylvain Lamprier, Ludovic Denoyer, and Patrick Gallinari. 2014. Learning social network embeddings for predicting information diffusion. In *Proceedings of the 7th ACM international conference on Web search and data mining.* ACM, 393–402.

[11] Edward Choi, Nan Du, Robert Chen, Le Song, and Jimeng Sun. 2015. Constructing disease network and temporal progression model via context-sensitive hawkes process. In *Data Mining (ICDM), 2015 IEEE International Conference on.* IEEE, 721–726.

[12] Manoranjan Dash, Kee Kiat Koo, Joao Bartolo Gomes, Shonali Priyadarsini Krishnaswamy, Daniel Rugeles, and Amy Shi-Nash. 2015. Next place prediction by understanding mobility patterns. In *Pervasive Computing and Communication Workshops (PerCom Workshops), 2015 IEEE International Conference on.* IEEE, 469–474.

[13] Manlio De Domenico, Antonio Lima, and Mirco Musolesi. 2013. Interdependence and predictability of human mobility and social interactions. *Pervasive and Mobile Computing* 9, 6 (2013), 798–807.

[14] Trinh Minh Tri Do and Daniel Gatica-Perez. 2012. Contextual conditional models for smartphone-based human mobility prediction. In *Proceedings of the 2012 ACM conference on ubiquitous computing.* ACM, 163–172.

[15] Carlos Duarte, Kevin Van Den Wymelenberg, and Craig Rieger. 2013. Revealing occupancy patterns in an office building through the use of occupancy sensor data. *Energy and Buildings* 67 (2013), 587–595.

[16] Varick L Erickson, Yiqing Lin, Ankur Kamthe, Rohini Brahme, Amit Surana, Alberto E Cerpa, Michael D Sohn, and Satish Narayanan. 2009. Energy efficient building environment control strategies using real-time occupancy measurements. In *Proceedings of the First ACM Workshop on Embedded Sensing Systems for Energy-Efficiency in Buildings.* ACM, 19–24.

[17] Manuel Gomez Rodriguez, Jure Leskovec, and Andreas Krause. 2010. Inferring networks of diffusion and influence. In *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining.* ACM, 1019–1028.

[18] Tristan Henderson, David Kotz, and Ilya Abyzov. 2008. The changing usage of a mature campus-wide wireless network. *Computer Networks* 52, 14 (2008), 2690–2712.

[19] Kurt Hornik, Maxwell Stinchcombe, and Halbert White. 1989. Multilayer feedforward networks are universal approximators. *Neural networks* 2, 5 (1989), 359–366.

[20] Jon Hutchins, Alexander Ihler, and Padhraic Smyth. 2007. Modeling count data from multiple sensors: a building occupancy model. In *Computational Advances in Multi-Sensor Adaptive Processing, 2007. CAMPSAP 2007. 2nd IEEE International Workshop on.* IEEE, 241–244.

[21] Yangqing Jia, Evan Shelhamer, Jeff Donahue, Sergey Karayev, Jonathan Long, Ross Girshick, Sergio Guadarrama, and Trevor Darrell. 2014. Caffe: Convolutional architecture for fast feature embedding. In *Proceedings of the 22nd ACM international conference on Multimedia.* ACM, 675–678.

[22] Michael I Jordan, Zoubin Ghahramani, Tommi S Jaakkola, and Lawrence K Saul. 1999. An introduction to variational methods for graphical models. *Machine learning* 37, 2 (1999), 183–233.

[23] Christian Koehler, Nikola Banovic, Ian Oakley, Jennifer Mankoff, and Anind K Dey. 2014. Indoor-ALPS: an adaptive indoor location prediction system. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing.* ACM, 171–181.

[24] Theodoros Lappas, Evimaria Terzi, Dimitrios Gunopulos, and Heikki Mannila. 2010. Finding effectors in social networks. In *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining.* ACM, 1059–1068.

[25] Yanchi Liu, Chuanren Liu, Xinjiang Lu, Mingfei Teng, Hengshu Zhu, and Hui Xiong. 2017. Point-of-Interest Demand Modeling with Human Mobility Patterns. In *KDD*. ACM, 947–955.
[26] Radu-Corneliu Marin, Ciprian Dobre, and Fatos Xhafa. 2012. Exploring predictability in mobile interaction. In *Emerging Intelligent Data and Web Technologies (EIDWT)*. IEEE, 133–139.
[27] Andriy Mnih and Karol Gregor. 2014. Neural variational inference and learning in belief networks. *arXiv preprint arXiv:1402.0030* (2014).
[28] Anna Monreale, Fabio Pinelli, Roberto Trasarti, and Fosca Giannotti. 2009. Wherenext: a location predictor on trajectory pattern mining. In *KDD*. ACM, 637–646.
[29] Mikołaj Morzy. 2007. Mining frequent trajectories of moving objects for location prediction. In *International Workshop on Machine Learning and Data Mining in Pattern Recognition*. Springer, 667–680.
[30] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. 2015. Faster r-cnn: Towards real-time object detection with region proposal networks. In *Advances in neural information processing systems*. 91–99.
[31] Manuel Gomez Rodriguez, David Balduzzi, and Bernhard Schölkopf. 2011. Uncovering the temporal dynamics of diffusion networks. *arXiv preprint arXiv:1105.0697* (2011).
[32] Ilya Sutskever, Oriol Vinyals, and Quoc V Le. 2014. Sequence to sequence learning with neural networks. In *Advances in neural information processing systems*. 3104–3112.
[33] Liaoruo Wang, Stefano Ermon, and John E Hopcroft. 2012. Feature-enhanced probabilistic models for diffusion network inference. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*. Springer, 499–514.
[34] Pengfei Wang, Yanjie Fu, Guannan Liu, Wenqing Hu, and Charu Aggarwal. 2017. Human Mobility Synchronization and Trip Purpose Detection with Mixture of Hawkes Processes. In *KDD*. ACM, 495–503.
[35] Duncan J Watts and Peter Sheridan Dodds. 2007. Influentials, networks, and public opinion formation. *Journal of consumer research* 34, 4 (2007), 441–458.