Semantics-Aware Hidden Markov Model for Human Mobility

ABSTRACT

Understanding human mobility is increasingly important to many applications such as urban planning, traffic control and city management. Previous work mainly focuses on modelling spatial and temporal patterns of human mobility. However, the semantics of trajectory are ignored, thus failing to describe people's motivation behind mobility. In this paper, we propose a novel semantic-aware mobility model that captures human mobility motivation using large-scale semantic-rich spatial temporal data from location-based social networks. In our system, we first develop a multimodal embedding method to project user, location, time, and activity on the same embedding space in an unsupervised way while preserving original trajectory semantics. Then, we use hidden Markov model to learn latent states and transitions between them in the embedding space, which is the sum of location and activity embedding vector, to jointly consider spatial, temporal, and user motivations. In order to tackle the sparsity of individual mobility data, we further propose a von Mises-Fisher mixture clustering for user grouping so as to learn a reliable and fine-grained model for groups of users sharing mobility similarity. We evaluate our proposed method on two large-scale real-world datasets, where we validate the ability of our method to produce high-quality mobility models. We also conduct extensive experiments on the specific task of location and activity prediction. The results show that our model outperforms baseline mobility models with much higher prediction accuracy.

KEYWORDS

Graph Embedding, Hidden Markov Model, User Group, Human Mobility

1 INTRODUCTION

With the increasing popularity of personal mobile devices and location-based applications, large-scale trajectories of individuals are being recorded and accumulated at a faster rate than ever, which makes it possible to understand human mobility from a data-driven perspective. Modelling human mobility is widely regarded as one fundamental task for numerous applications: not only does it provide key insights for urban planning, traffic control, city management and government decision making, but also enables personalized activity recommendation and advertising.

As a result, there has been substantial previous work on human mobility modelling. The majority of previous work focuses on modelling the spatial and temporal patterns of human mobility. Human mobility is generally modelled as a stochastic process around fixed point [9] and various models for next location prediction [3, 8, 19–21] have been proposed. One shortcoming of these mobility models, however, is that they overlook the activity (often referred to as the semantics of trajectory) a person engages in at a location within a certain time. Thus, these models are not capable of explaining people's motivation behind mobility and unable to predict human activity. To tackle this problem, recently a few semantic-aware mobility models [27, 28, 30] have been proposed, which attempt to jointly model spatial, temporal and semantics aspects. Yet these models are still far from satisfactory as they do not properly distinguish motivation between users. For instance, people who appear at the same location for different motivation (e.g. a chef and a customer in the same restaurant) will be considered the same, while people visiting different locations for similar purposes (e.g. for entertainment, white collar *A* goes to gym after work while white collar *B* goes to the movie) are considered different. Therefore, the problem of semantic-aware mobility modelling remains very much an open question.

In this paper we propose a novel semantic-aware mobility model using large-scale semantic-rich spatial temporal data – from the location-based social networks Twitter, Foursquare and WeChat – which consist of user, location, time, and activity information. Specifically, the new proposed mobility model addresses the following two issues.

- The model is able to capture motivation underlying human mobility. For instance, it is able to identify that the movement of white collar *A* to gym and white collar *B* to a movie after work are similar in motivation because they both go for entertainment and relaxation. On the other hand, the model is also able to capture the difference between a customer and a chef appearing in the same restaurant, since they come to the restaurant for different purposes.
- The model is able to discover intrinsic states underlying human mobility as well as transition patterns among them. A state takes into account spatial, temporal and user motivation as a whole. For example, working in an office building at district *C* during the day is a possible state, and a user in this state having 80% chance to transit to the state of being in a restaurant at district *D* in the evening for food is a possible transition pattern.

Such semantics-aware mobility models are especially helpful and enables various applications. First of all, they are well-suited for next location and activity prediction [30], and thus benefit personalized recommendation and targeted advertising. Unlike existing work, our model jointly considers various aspects of human mobility, thus has the capacity to greatly enhance prediction accuracy. Secondly, it is potentially useful in revealing economic status of the city for decision makers since the model captures fine-grained routines and motivations in human mobility.

However, developing a semantics-aware mobility model is challenging due to three major reasons. (1) *Data integration*: It is difficult to integrate and represent spatial, temporal and semantic information as a whole since they belong to different spaces and have distinct representations. (2) *Model construction*: It is nontrivial to define latent states and identify transition patterns given the complexity and diversity of data. (3) *Data Sparsity*: It is challenging to construct a both reliable and fine-grained mobility model at the same time given the limited number of records for each individual user.

To tackle the above three challenges, we propose an embedding-based Hidden Markov Model (HMM) to capture patterns of human mobility. To address the data integration challenge, we propose a multimodal embedding method to project user, location, time and activity on the same embedding space based on co-occurrence frequency in an unsupervised way while preserving original semantics in the dataset. Through the embedding procedure, all users, locations, times and activities appearing in the original dataset are represented by a numeric vector of the same length, which can be directly compared using classical distance metric (e.g. cosine similarity). Then, we adopt HMM model in the embedding space to learn latent states and transitions between them for mobility modelling, where each latent state is the sum of location and activity embedding vector, so that spatial, temporal (temporal information affects the overall embedding and thus affects the HMM training process) and user motivations are jointly considered in the model. Moreover, to solve the problem of data sparsity, we propose a von Mises-Fisher mixture clustering on the user embedding vector for user grouping so as to learn a reliable and fine-grained model for groups of users sharing mobility similarity. We train a separate HMM model on each user group, and get an ensemble of high-quality HMM models. Finally, we project latent state of each user group back to original spatial, temporal and activity space to study human mobility patterns. Our contributions can be summarized as follows:

- We propose a novel mobility model which fully takes into account semantics in human mobility. It not only considers spatial and temporal aspects, but also the activity the user engages in as well as user motivation behind mobility. Furthermore, to the best of our knowledge, our model makes the first attempt to jointly consider these factors and their complex inner correlation in an unsupervised way.
- We first introduce the techniques of embedding into mobility modelling to propose a semantics-aware HMM model. After projecting user, location, time and activity on the same embedding space preserving original semantics, we train an ensemble of HMM models in the embedding space based on von Mises-Fisher mixture user grouping. We then project HMM latent state back to the usual space to analyze human mobility pattern. Through this latent-state-based modelling, we obtain high-quality group-level mobility model.
- We evaluate our proposed method on two large-scale real-world datasets. The results justify the ability of our method in producing high-quality mobility model. We also conduct extensive experiments on the specific task of location/activity prediction. We observe that our model outperforms baseline mobility models with much higher prediction accuracy. The accuracy increases by 12.8% and 3.22% on app collected dataset and foursquare dataset, respectively.

2 MOTIVATION AND MODEL OVERVIEW

In this section, we discuss the motivation in developing our mobility model. We first discuss the system design philosophy, and then provide an overview of our solution.

2.1 System Design Philosophy

We aim to propose a novel semantic-aware mobility model. Previous work mostly focuses on modelling spatial temporal patterns in trajectory regardless of semantics, i.e., clustering users by extracting features from geographical trajectory, learning Markov model in geographical space, etc. However, we believe that the semantic information such as POI attached to the trajectories is valuable since it reflects users' motivation behind mobility, making it possible to model human mobility in a deeper way. The key idea of our work, therefore, is to integrate semantic information with user trajectory data for mobility modelling so as to discover underlying and insightful human mobility patterns.

To capture semantics behind these different types of information in trajectory, we therefore propose a multimodal embedding method by constructing co-occurrence graph and conduct graph-embedding to project these information on the same latent space. By considering the co-occurrence relationship, the latent space remains the proximity of semantics among different types of information. The introduction of multimodal embedding further provides a natural solution to the challenges of data sparsity and model construction. On one hand, users engaging in similar activities and staying in nearby regions are closer to each other on the latent space. Thus, we can find user groups sharing mobility similarity using clustering methods in the latent space. We leverage the group mobility information to train a model for each group. On the other hand, semantics can be combined in latent space through vector addition. Thus we can train an HMM model whose observable state is the sum of location vector and POI type vector, which captures the intentions of user much more accurately. The projected data in the latent space reflects mobility pattern better than the original trajectory data, thus leads to better prediction performance. Furthermore, the design of this model makes it possible to predict POI type given location.

2.2 Model Overview

Based on the approach discussed above, we illustrate our proposed model overview in Fig. 1, which includes three major modules. The *representation learning* module constructs a heterogeneous graph and embeds personal, temporal, spatial and semantic information into a latent space. Based on the obtained latent space, the *Embedding-based User Grouping* module cluster users sharing similar mobility and life patterns and the *Group-level Hidden Markov Model* module learns human patterns with the embedded data. Now we first formally define the mobility modelling problem, and then introduce the system model with details of these three modules.

In our model, we discretize time into discrete time slot (t_1, t_2,....) and spatial space into finite set of area $\mathbf{L} = \{l_1, ...l_{N^L}\}$, where N^L is the total number of discretized area in L. For each user u in the set of all users $\mathbf{U}$, $y^u = (y^u_1, ...y^u_i ..., y^u_{N^u})$ denotes history trajectories of user u, where N^u denotes the number of sampling points in user u's trajectory. $y^u_i = (u, t^u_i, l^u_i)$ denotes the location l^u_i being visited by user u at time slot t^u_i and $l^u_i = (l_a, l_o)^u_i$. (Note that N^u of each user is probably different). Besides the trajectory data, our model also adopts semantic information set $\mathbf{H} = \{h_1, ...h_k, ..., h_{N^L}\}$, where $h_k = (l_k, P_k)$. $P_k = (p_1, ..., p_j ...p_{N^P})$ denotes the associated POIs in the area l_k, p_j denotes the POI type, and N^P is the number of POI

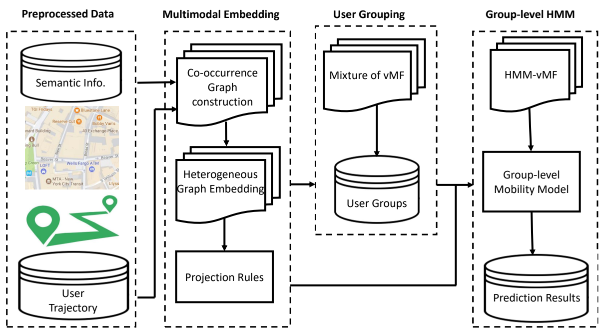


Figure 1: Overview of the proposed embedding based group-level human mobility model.

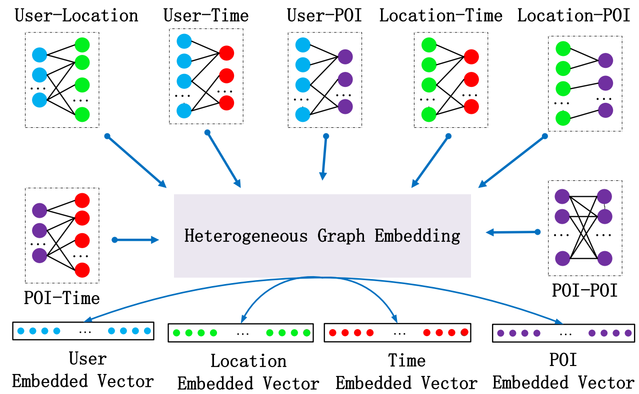


Figure 2: Illustration of the details of our representation learning model. The co-occurrence relationships construct 7 sub-graphs, which are jointly embedded with graph-based method.

types, i.e, the check-in POI in the history visited l_k. For evaluation, our model predicts the next state (location) $y^u_{N^u+1} = (t^u_{N^u+1}, l^u_{N^u+1})$ of each user u, based on the past trajectory $y^u = (y^u_1, ...y^u_i..., y^u_{N^u})$ and also predict semantic information set $\mathbf{H} = (h_1, ...h_k, ..., h_{N^L})$ given user and location.

Multimodal Embedding module constructs the structure of user (u), temporal (t), spatial ((l_o, l_a)) and user motivation/semantics (P) information. For high quality embedding, . When two units appear in the same record, *co-occurrence* happens. Based on the extracted *co-occurrence*, a heterogeneous graph is learned, which embeds the co-occurrence and neighboring relationships into one latent space. The graph encodes the human mobility intentions into vectors in that embedding space.

User Grouping module clusters users based on user embedding vectors in the latent space. Motivated by the effectiveness of cosine similarity in the embedding space[18, 29], we model each cluster of users as a von Mises-Fisher (vMF) distribution in the latent space. Naturally, we use mixture-of-vMFs model[1] to cluster users into groups in latent spaces for follow-up HMM training.

Group-level HMM module learns the transitions patterns in the latent space of a group of similar users. In the latent/embedding space, since the temporal and spatial proximity of human trajectory and intrinsic correlations between temporal, spatial and semantic information have been well captured, Hidden Markov is good enough for training and prediction. Similar to user grouping, each hidden state corresponds to a vMF distribution in the embedding space. For prediction, we first generate the combinations of locations and POI types and calculate the score of each combination. Based on this framework, we can sort the combinations by scores and get top k locations as prediction results. Moreover, we can predict the top k poi types when fixing the next location.

3 METHOD

In this section, we first discuss the multimodal embedding designed for the mobility model which captures the diversified semantics. Then we present embedding-based user-grouping and HMM, which learns a fine-grained semantic-aware mobility model in the embedding space.

3.1 Multimodal Embedding

The multimodal embedding module jointly maps all the user, spatial, temporal, and semantic information into the same low-dimensional space with their correlations preserved. While the semantic are natural POI types for embedding, the space and time are are continuous and there are no natural embedding units. To address this issue, we break the geographical space into equal-size regions and consider each region as a spatial unit ($500m * 500m$ grid). Similarly, we break one day into 24 hours distinguished by weekday and weekend and consider every hour as a basic temporal unit (totally 48 units). The embedding module first extracts the correlations between user, time, location and semantic information as co-occurrence relationships. Then, it embeds all the co-occurrence relationships into one latent space, which encodes the human mobility intentions into vectors in embedding space, as shown in Fig. 2.

3.1.1 Co-occurrence Relationship. The *co-occurrence* relationship describes the intrinsic correlations between different information: who (u), when (t), where ((l_a, l_o)) and what semantic (P). It happens when two different kinds of units appear in one record. This relationship reflects the intrinsic correlations between different information units. For example, the pre-processed data outputs a record with a temporal unit t, a spatial unit (l_o, l_a) and a POI list (p_1, p_2, p_{10}), and seven co-occurrence relationships exist: user-time($u - t$), user-location($u - (l_a, l_o)$), user-poi($u - p_1$, $t - p_2$) time-location ($t - (l_a, l_o)$), time-poi ($t - p_1$, $t - p_2$), location-poi ($(l_a, l_o) - p_1$, $(l_a, l_o) - p_2$), and poi-poi ($p_1 - p_2$).

3.1.2 Heterogeneous Graph Learning. The graph embedding encodes the different types of information into graph nodes and expresses their relationships with the edges and their weights. Different relationships are constructed with different graphs. The weights of these graph edges are normalized and combined together as a heterogeneous graph. In the graph, there exist three different node types corresponding to four unit (information) types (user, time, location and semantic). Each *co-occurrence* relationship constructs one edge, whose weight is set to be the normalized counts.

Besides the explicit relationships, the graph also keeps the implicit interactions among units. The implicit interaction means that two nodes are not directly connected but share a lot of common neighbors. In the embedding space, these nodes should be close in distance. Thus we first model each node's emission probability distribution based on their latent embedding. Then, we try to minimize the distance between the real observed distributions and this model distributions.

The likelihood of generated node j given node k is defined as

$$p(j|k) = \frac{\exp(-v_j^T \cdot u_k)}{\sum_{i \in U} \exp(-v_i^T \cdot u_k)}, \tag{1}$$

where u_k and v_j represent embedded vector of node k and j respectively. Note that for node j there are two different embedding vectors with different functions. v_j represents the vector when node j is the given node while u_j is the vector when node j acts as the emitted node. In addition, we define true distribution observation as:

$$\hat{p}(j|k) = \frac{w_{kj}}{d_k}, \tag{2}$$

where d_k is defined as $\sum_{l \in U} w_{kl}$ and w_{kj} represents the edge weight.

We also need to define the loss function for the sub-graph G_{UV} before we can minimize the distance between the embedding-based distributions and truly observed distributions:

$$L_{UV} = \sum_{j \in U} d_j d_{KL}(\hat{p}(\cdot|j)||p(\cdot|j)) + \sum_{k \in V} d_k d_{KL}(\hat{p}(\cdot|k)||p(\cdot|k)), \tag{3}$$

where $d_{KL}()$ is Kullback-Leibler divergence [13]. With three different nodes representing user(U), temporal (T), spatial (S) and semantic (H) information, the overall loss function can be obtained as:

$$L = L_{UT} + L_{US} + L_{UH} + L_{TS} + L_{TH} + L_{SH} + L_{HH}. \tag{4}$$

Due to high computational complexity of optimizing the loss function with large scale graph, stochastic gradient descent with negative sampling is adopted for quicker computation [18]. For an edge from node j to node k, the negative sampling method treats node k as a positive example while randomly selects N nodes, which are not connected to j as negative examples. As a result, we just need to minimize:

$$L' = -\log \sigma(u_j^T \cdot v_k) - \sum_{n=1}^{N} \log\sigma(-u_n^T \cdot v_k), \tag{5}$$

where $\sigma()$ represents the sigmod function [10].

3.2 Grouping-based HMM

3.2.1 User Grouping in the Embedding Space. After embedding different types of information into the embedding space, we obtain representation vectors for users, which maintains the semantic proximity in this latent space. Since cosine similarity is effective in the embedding space, we use vMF to model each cluster of users' vectors in the latent space. For a d-dimensional unit vector x that follows d-variate vMF distribution, its probability density function is given by

$$p(x|\mu, \kappa) = C_d(\kappa) exp(\kappa \mu^T x), \tag{6}$$

where the mean direction unit vector μ and the concentration parameter κ are two important parameters that describe vMF distribution. The normalization constant $C_d(\kappa)$ is given by

$$C_D(\kappa) = \frac{\kappa^{d/2-1}}{(2\pi)^{d/2} I_{d/2-1}(\kappa)}, \tag{7}$$

where $I_r(\cdot)$ means the modified Bessel function of the first kind and order r.

For clustering users into several groups that have similar mobility semantic patterns, naturally we use a mixture of vMF model to fit the embedded data of users. The probability of v_U in a k-vMF distribution is given by

$$p(v_U|\mu, \kappa) = \sum_{h=1}^{k} \alpha_h f_h(v_U|\mu_h, \kappa_h), \tag{8}$$

where α_h are the weights of each mixtures and sum to one.

We design an EM framework to maximize the probability of the whole k-vMF model. After we randomly set the initial value for each vMF, we repeat E-Step and M-Step until the parameters coverage. In E-step, we estimate the probability of each user U_i belonging to each group,

$$p(h|v_{U_i}, \mu, \kappa) = \frac{\alpha_h f_h(v_{U_i}|\mu_h, \kappa_h)}{\sum_{l=1}^{k} \alpha_l f_l(v_{U_i}|\mu_l, \kappa_l)}. \tag{9}$$

Also, we can adapt it into a formation of hard labels, i.e., assign each user to just one group.

$$p(h|v_{U_i}, \mu, \kappa) = \begin{cases} 1, & \text{if } h = argmax_{h'} p(h'|x_i, \mu, \kappa), \\ 0, & \text{otherwise.} \end{cases} \tag{10}$$

In our model, both the soft and hard assignments of hidden state is feasible, which is a tradeoff between efficiency and accuracy. In M-Step, we maximize the probability of the model by updating the parameter μ_h, κ_h for each cluster.

$$r_h = \sum_{i=1}^{n} x_i p(h|x_i, \mu, \kappa); \ \hat{\mu_h} = \frac{r_h}{||r_h||}; \ \hat{\kappa_h} = \frac{||r_h||d - ||r_h||^3}{1 - ||r_h||^2}. \tag{11}$$

When the difference of total probability of this model before and after one iteration is less than a threshold, this iterative process terminates. The output of this algorithm is the probability each user belonging to a certain group.

3.2.2 HMM-based model. Based on the representation vectors and the user groups, we utilize an HMM for each group of users to model the transitions among trajectories in the semantic latent space. As Fig. 3 shows, our HMM model chooses the embedding results from representation learning model as observations to model the sequence. Due to the semantics additive property, we choose the the sum of the vectors of location and POI type. The proximity of semantic vectors characterizing the activity of users should also be measured by the cosine similarity like the users' representation vectors. Thus, we utilize vMF distribution as the emission probabilty of each hidden state. Now, we first define related parameters in HMM and then discuss the prediction procedure.

Combining the past trajectory y^u of user u with semantic information $\mathbf{H}$, we redefine the trajectory expression as $\mathbf{T}^u = \{x_1, x_2, ..., x_N\}$, where $x_n = (l_a^u, l_o^u, P^u)$ and N is the length of trajectory. Without confusion, we ignore the superscript u below. With the method of

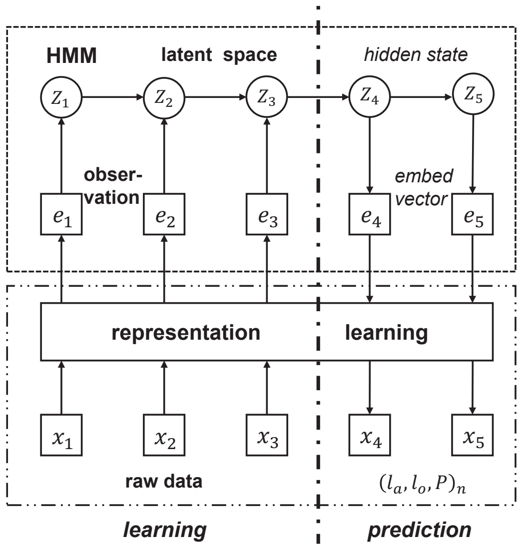


Figure 3: Illustration of the details of HMM-based prediction model in the latent space and its relationship with the physical locations.

representation learning, we embed the new trajectory $\mathbf{T}$ into one latent space $\mathbf{E} = \{e_1, e_2, ..., e_N\}$. Each observation $e_n = E_f((l_a^u, l_o^u)) + \sum_{p^u \in P^u} E_f(p^u)$, where E_f is the embedding function which projects different types of information into the embedding space. In order to generate the observation sequence $\mathbf{E}$, we assume that there are K hidden states $\mathbf{Z_K} = \{1, 2, ..., K\}$. Each observation x_n corresponds to one z_n, which belongs to $\mathbf{Z_K}$. Then, corresponding to the observation sequence $\mathbf{E}$, we have one hidden state sequence $\mathbf{Z} = \{z_0, z_1, ..., z_N\}$. To describe the overall model, we have:

- A K-dimensional vector $\mathbf{\Pi}$, where $\pi_k = p(z = k)$, which defines the initial value of hidden states;
- A matrix $\mathbf{A} = \{a_{ij}\} \in \mathbb{R}^{K \times K}$, which defines the transition probabilities between K hidden states with $a_{ij} = p(z_n = j | z_{n-1} = i)$;
- A set of vMFs $\mathbf{B} = \{vMF_i(e_n)\}$, where e_n is the emission from a hidden state $z_n = i$ into the embedding space.

Then, our model is parameterized by $\Phi = \{\mathbf{Z_K}, \mathbf{\Pi}, \mathbf{A}, \mathbf{B}\}$. The occurrence probability of an observation sequence $\mathbf{E} = \{e_1, e_2, ..., e_N\}$ with the state sequence $\mathbf{Z} = \{z_1, z_2, ..., z_N\}$ can be expressed as follows,

$$p(\mathbf{E}|\mathbf{Z}, \Phi) = \prod_{j=1}^{N} b_{i_j}(e_j). \tag{12}$$

The cumulative occurrence probability of observation sequence $\mathbf{E}$ is

$$p(\mathbf{E}|\Phi) = \sum_{\mathbf{Z}} p(\mathbf{E}|\mathbf{Z}, \Phi) \cdot p(\mathbf{Z}|\Phi) \tag{13}$$

$$= \sum_{\mathbf{Z}} \pi_{i_1} \cdot \prod_{j=1}^{N-1} b_{i_j}(e_j) a_{i_j i_{j+1}} \cdot b_{i_N}(e_N). \tag{14}$$

We design an Expection-Maximization (EM) based algorithm to estimate the parameters. After starting with initial parameters, the algorithm iterates between E-step and M-step to update parameters and finally converges to a stable state. In a particular iteration, E-step computes the likelihood function based on the parameters of previous iteration. Then, M-step updates the parameters by maximizing the likelihood function. Below, we introduce details about the parameter re-estimation process.

We first define two auxiliary probabilities $\xi_t(i,j) = p(z_{t+1} = j, z_t = i | \mathbf{E}, \Phi)$ and $\gamma_t(i) = p(z_t = i | \mathbf{E}, \Phi)$, where $t = 1, 2..., N$. To calculate efficiently, we exploit a forward-backward procedure[2] to calculate these two probabilities. The forward probability $\alpha_t(i)$ is

$$\alpha_t(i) = p(e_1, e_2, ..., e_t, z_t = i | \Phi). \tag{15}$$

Then $\alpha_{t+1}(j)$ can be calculated as follows,

$$\alpha_{t+1}(j) = \Big[\sum_{i=1}^{K} \alpha_t(i) a_{ij}\Big] b_j(e_{t+1}), \tag{16}$$

and initial values are $\alpha_t(i) = \pi_i b_i(e_1)$. The backward probability is $\beta_t(i) = p(e_{t+1}, e_{t+2}, ..., e_N | z_t = i, \Phi)$, which can be calculated as follows,

$$\beta_t(i) = \sum_{j=1}^{K} a_{ij} b_j(e_{t+1}) \beta_{t+1}(j), \tag{17}$$

with initial values $\beta_N(i) = 1$. Based on $\beta_t(j)$ and $\alpha_t(i)$, $\xi_t(i,j)$ can be calculated as

$$\xi_t(i,j) = \frac{\alpha_t(i) a_{ij} b_j(e_{t+1}) \beta_{t+1}(j)}{\sum_{m=1}^{K} \sum_{n=1}^{K} \alpha_t(m) a_{mn} b_n(e_{t+1}) \beta_{t+1}(n)}. \tag{18}$$

While $\gamma_t(i)$ can be calculated as

$$\gamma_t(i) = \frac{\alpha_t(i) \beta_t(i)}{\sum_{j=1}^{K} \alpha_t(j) \beta_t(j)}. \tag{19}$$

Based on $\xi_t(i,j)$ and $\gamma_t(i)$ with $\Phi^{(t)}$, the parameters of HMM can be updated by the following formulas,

$$\pi_i = \gamma_1(i); \; a_{ij} = \frac{\sum_{t=1}^{K-1} \xi_t(i,j)}{\sum_{t=1}^{K-1} \gamma_t(i)}; \tag{20}$$

$$r_i = \frac{\sum_{t=1}^{K} \gamma_t(i) e_t}{\sum_{t=1}^{K} \gamma_t(i)}; \hat{\mu_i} = \frac{r_i}{\|r_i\|}; \hat{\kappa_i} = \frac{\|r_i\| d - \|r_i\|^3}{1 - \|r_i\|^2}. \tag{21}$$

When the training process terminates, we obtain the HMM based semantics-aware mobility model. In order to leverage the model for location/activity prediction, we construct a set of length-2 sequences (e_N, e_{N+1}) where $e_N = E_f((l_{a,N}^u, l_{o,N}^u)) + E_f(P_N^u)$ and $e_{N+1} = E_f((l_{a,N+1}^u, l_{o,N+1}^u)) + E_f(P_{N+1}^u)$ for each location-POI type combination in candidates set. Then we calculate the probability of generating such a sequence from the model as the score S of the sequences given by

$$S(\{(l_{a,N+1}^u, l_{o,N+1}^u), P_{N+1}^u\}) = p(\{e_N, e_{N+1}\} | \pi, A, B)$$
$$= \sum_{m=1}^{K} \sum_{n=1}^{K} \pi_1 a_{12} C_d(\kappa_m) exp(\kappa_m \mu_m^T e_N) C_d(\kappa_n) exp(\kappa_n \mu_n^T e_{N+1})). \tag{22}$$

Based on the definition of score on location-POI type combination sequence, we further define the score of next location, as well as the score of next POI type when the location is given in prediction tasks. To predict the next location, we set the score of next location as the highest score of the location-POI type combination sequence that consists of this location. We can also predict the type of POI the user will visit given his/her next location. To predict the POI type, we just need to remain the combination of which the location lies in the given region and remove others. Then we can get the top-K list of POI types as the prediction results. Note that if one record contains more than one POI type, we will remove it from testing set. We will quantitatively evaluate our mobility model on these two tasks.

4 EVALUATION

In this section, we evaluate our system with two real-life datasets. The evaluation mainly contains four parts:

- Presenting case study and the corresponding insightful results to validate the ability of our model in discovering mobility patterns.
- Comparing our method with baseline methods including previous works and variants of our model.
- Illustrating the effect of main parameters in our model such as the dimension of embedding, the number of groups and the number of hidden states.
- Exploring the performance of our model on users with different mobility patterns including the number of different places visited and the trajectory's entropy.

4.1 Experimental Settings

4.1.1 Dataset. We use the following two different datasets to evaluate the performance of our system.

App Collected Dataset: It was collected by a popular localization platform. When users use related Apps, such as WeChat (the most popular online instant messenger in China), their location information will be uploaded to the servers and is collected by this platform. Thus, these records strongly imply users' behavior patterns. Overall, the utilized dataset is collected from $1,000$ anonymous users, who are active during September 17 to October 31, 2016 in Beijing. There are about 4 million records in our dataset, which is large-scale and guarantees the credibility of our study. Each record consists of the following fields: the anonymized ID of the user, the time of the record (accurate to second), the location information in the format of GPS coordinates and the associated POI types in this location. The records are very dense, so we filter out the pass-by points and extract about 40 thousand stays from the raw data by setting temporal and spatial thresholds [12]. It contains 19 types of POIs, including: food, enterprise, organization, shopping, life service, entertainment, fitness, vehicle, medical care, accommodation, tourist attraction, culture venue, school, finance, address name, infrastructure, estate, indoor and ancillary facility and others. Thus, this dataset is highly suitable for semantics-aware mobility modelling as mining semantic information embedded contribute to improved accuracy of mobility prediction.

Check-in location-based service application: This publicly available dataset comes from foursquare, a location-based service application. The data includes 45 thousands records of $1,000$ active users from February 25, 2010 to January 19, 2011. Each record consists of the following fields: the user ID, the time stamp(accurate to second), location, and the type of poi that the user check in. The poi types include travel, shop, professional, college, residence, outdoors, food, arts. Compared with the first dataset, the record in this dataset is sparser but the time span is longer.

An important difference between these two datasets is that the app collected dataset is passively recorded in the background while the foursquare dataset is recorded by users' active check-in. As a result, there are many records at home or working places in the first dataset which reflect users' real life patterns while there are only a few such records in the foursquare datasets since people tend to actively check in at explorative places such as shopping malls or restaurants in their spare time rather than home or working places.

4.1.2 Baselines. We compare our model with the following five state-of-the-art solutions as follows.

GeoGaussHMM is proposed in existing works[5][15]. It trains one HMM for all users' trajectory, where each hidden state generates location by a Gaussian distribution.

EmbedGaussHMM trains one HMM where each hidden state generates vectors representing locations in the latent space obtained by graph-embedding.

EmbedVmfHMM-I replaces the Gaussian distribution by vMF distribution in the last model so as to adapt to the cosine distance metric in the semantic latent space.

EmbedVmfHMM-II is an improved version of *EmbedVmfHMM-I*. Compared to *EmbedVmfHMM-I*, the hidden states generates vector which represents location vector plus POI type vector in the latent space.

Embed-Gmove constructs several HMMs and assigns user to each HMM by a soft label proportional to the probability of drawing the trajectory from the HMM, which is proposed in [30]. The model adopts the same HMM structure as *EmbedVmfHMM-II*.

EmbedVmfHMMGroup is our proposed model. It performs a mixture of vMF which clusters users in the latent space, and trains one model using previous method for each group of users.

GeoGaussHMM is a classical method for trajectory prediction, while the others are the variants of our methods. The last method is the best version of our model, which takes into account semantics, user grouping and the property of the embedding space.

We compare the performance of the *EmbedGaussHMM* method and *EmbedVmfHMM-I* method to show which kind of distribution is more appropriate for the emission probability function in the embedding space. We compare the performance of the *EmbedVmfHMM-I* and *EmbedVmfHMM-II* to show which is the better rule to project the raw record to the embedding space: location's vector or the sum of location's vector and the POI type's vector. We compare the *EmbedVmfHMM-II* and *EmbedVmfHMMGroup* to show the advantage of user grouping in constructing human mobility model. To show the advantage of our method for user grouping, we compare *EmbedVmfHMMGroup* with *Embed-Gmove*[30], which cluster users based on the transitions.

For evaluation, we design another task: predicting POI type given location, which is similar to the answer "If the user goes there later, what will he/she do?". For comparison, we compare our model with *EmbedVmfHMM-II* and another baseline:

RandomVisit randomly selects the POI types in this region and gets the top k list of POI types.

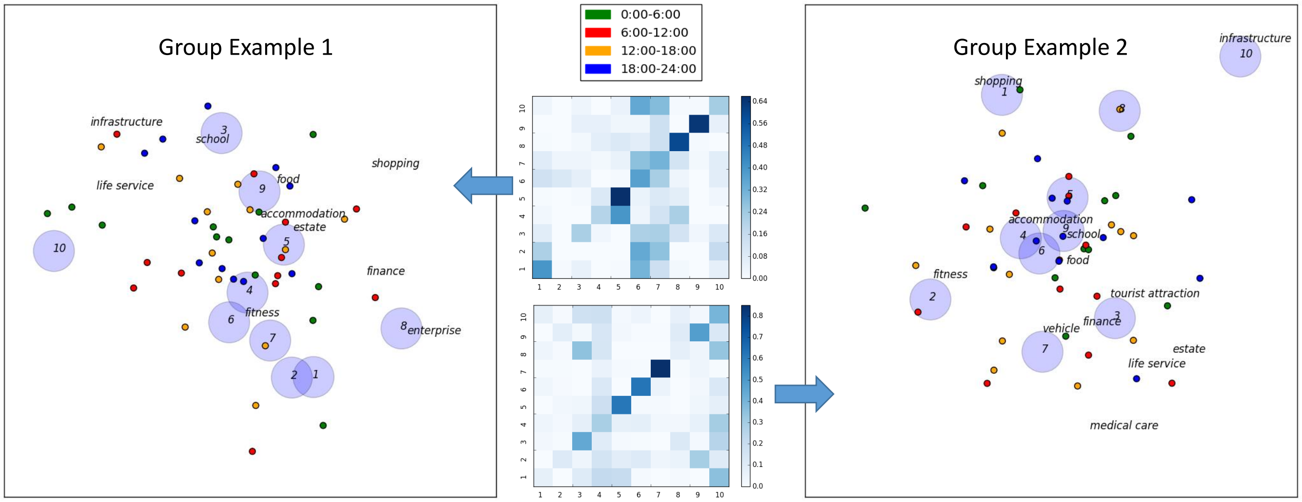


Figure 4: Two examples of the user groups. We map the embedding vectors of time, POI types and the central vector of each hidden state on a 2D plane with t-SNE[6]. The two heat maps in the middle represent HMM transition probability matrix for the two illustrated groups.

4.1.3 Evaluation Setup. We partition each dataset into training set and testing set. For the localization dataset, the first 36 days are regarded as training set while the remaining 10 days are the testing set. As the original records have several continuous records at the same place over time while pass-by records have no apparent meaning, we use the extracted stays as input data of the models. For the foursquare dataset, the records before October 1, 2010 (about seven months) are the training sets and the others (about two months) are the testing set. We use two tasks for the quantitative evaluation of our model. Our system includes three important parameters: the number of dimensions in embedding space E, the number of hidden states K and the number of user groups G. For performance comparison, we set $E = 50$, $G = 10$ and $K = 10$ for app collected dataset and $E = 200$, $G = 20$, $K = 10$ for the check-in dataset, which is obtained by careful tuning of parameters. To clearly demonstrate the parameter effect, we plot the performance curve by varying one parameter while fixing others.

4.2 Experimental Results

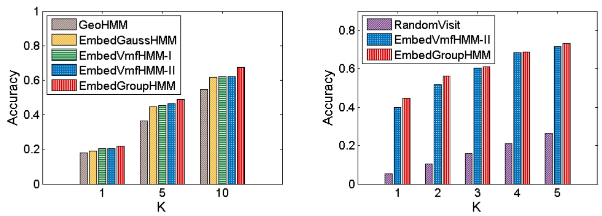
4.2.1 Case Study. After running our model on the two large-scale datasets, we obtained G mobility patterns corresponding to G groups of users. We select two examples from the app collected dataset to illustrate the physical meaning of the patterns discovered by our model. One merit of our model is that different types of information are comparable in the embedding space. Therefore, we can infer the semantics of hidden states through nearby information on the embedding space. To clearly demonstrate the mobility pattern with semantics, we map different types of information on a 2D plane with the proximity remained by t-SNE[6]. Also, we show the transition probability matrix by heat map. The depth of color represents the probability of transiting from vertical index hidden state to the horizontal index hidden state.

For the group example 1 in Fig. 4, we can infer that this group probably represents sport-lovers. We observe that the hidden states 6 and 7 means the activity of doing sports because they are near POI type 'fitness' (which contains gym, basketball court, natatorium, etc.) while the two temporal points during 12:00-18:00 refers to the most frequent time when people do sports. We also observe that this group of people transit to the state of doing sports from multiple hidden states. Furthermore, the state 5 often goes back to itself and is close to POI type 'estate' which strongly implies home.

For the group example 2 in Fig. 4, we can infer that this group represents tourists. First, there are hidden states near POI types 'tourism attraction', 'shopping', 'accommodation' but no hidden state near 'estate'. Furthermore, the hidden state 10 locates near POI type 'infrastructure' which consists of airports, train stations, bus stop, etc. Also, there are many hidden states that transfer to hidden state 10 which is coherent with the character of transportation.

4.2.2 Performance Comparison. To demonstrate the effectiveness and robustness of our proposed model, we test it on two real large-scale datasets: localization service and foursquare. We not only compare its performance with previous human mobility model, but also the variants of our model so as to show how much improvement of performance each part of our model contributes to.

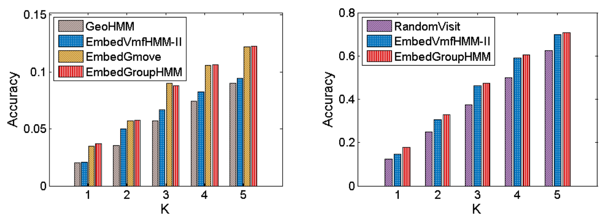
For localization service dataset, we show the accuracy of top K results in Fig.5. From Fig.5(a), we see that the performance of our model significantly improves by 4.1%, 12.5% and 12.8% for top 1, 5, 10 prediction, compared with previous mobility model based on mere temporal and spatial information. By comparing the performance of *GeoGaussHMM* and *EmbedGaussHMM*, we can observe that the multimodal embedding module, which maps semantic information into the latent space can improve the accuracy of prediction. Therefore, the effectiveness of using graph-embedding to capture semantic



(a) Top K Locations.

(b) Top K Poi Types.

Figure 5: App Collected Dataset: Prediction Accuracy of Top K.



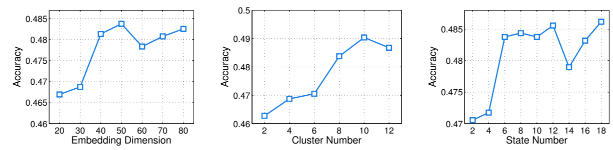
(a) Top K Locations.

(b) Top K Poi Types.

Figure 6: Check-in Dataset: Prediction Accuracy of Top K.

information to increase prediction accuracy is demonstrated. Then we validate that vMF is a better choice than Gaussian distribution as we observe that *EmbedVmfHMM-I* outperforms *EmbedGaussHMM*. Meanwhile, the time cost of training *EmbedVmfHMM-I* is much less than *EmbedGaussHMM*, because *EmbedGaussHMM* involves the calculation of inverse matrix whose complexity is $O(D^3)$. Then we explore how the projection methods affect the performance. Instead of the simple way to take the embedding vector of location as the observable state of HMM used in *EmbedGaussHMM* and *EmbedVmfHMM-I*, our model adopts the sum of location and POI type vector in the embedding space as basic unit. The experiment results verify that using the sum vector as observable state result in better performance. Finally, compared with *EmbedVmfHMM-II*, *EmbedVmfHMMGroup* group users into G groups and train one HMM for each group. We can observe a significant improvement in accuracy by user-grouping. From Fig.5(b), we demonstrate that after user grouping, the POI prediction becomes more accurate, as our model is capable of distinguishing people going to the same place under different motivations. Note that the first three baselines cannot predict POI type and thus not included in Fig.5(b).

For the foursquare dataset, we show the accuracy of top K results in Fig.6. From Fig.6(a), we can observe that our model significantly outperforms other baselines, which demonstrates that our model scales to different types of dataset. To show the merit of our user grouping method, we compare our method to the state-of-art user grouping method *Embed-Gmove* which is designed for sparse user-active-record data. We show that for most K value, our model outperforms this approach. On the other hand, due to the iterative structure of *Embed-Gmove*, the efficiency of our model is at least ten times better. Moreover, our framework is able to predict POI type more accurately compared with *Embed-Gmove*, as shown in Fig.6(b),.



(a) Effect of D.

(b) Effect of G.

(c) Effect of K.

Figure 7: Performance varies with the number of embedding dimensions, user groups and hidden states.

4.2.3 Parameter Effect. To understand the roles of system parameters in our proposed mobility model, we vary these parameters to plot the performance curve of our model. There are three important parameters in our model, which come from the three modules: embedding dimension D, number of user groups G and number of hidden state K. We use the accuracy of top 5 locations prediction as the main performance index based on which we tune other parameters. We show the experiment details as follows.

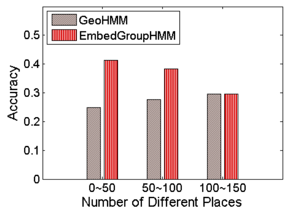
The first parameter D is the dimension of the embedding space in the *Multimodal Embedding* module. From Fig. 7 (a), we can observe that the accuracy is low when D is less than 40, which is coherent with our expectation: when the dimension of the embedding space is too low, it is not expressive enough to represent complex relationship from multiple types of information. When D reaches 50, the performance remains stable with a little fluctuation, which implies the dimension is now adequate to distinguish the different types of information. In sum, D decides the quality we embed the semantic information into our model.

The parameter G is the number of user groups from the *User Grouping* module. From Fig. 7 (b), we can observe that our model obtains the best performance when $G = 10$. The performance gets worse when the $G < 10$ or $G > 10$. When the number of user groups is too small, people with quite different mobility patterns are fused in one group which severely decreases the model performance. When the number of groups is too great, on the other hand, the model will suffer from data sparsity issue, as there are not enough data to train an HMM for each group. Thus, the optimal value of G, which helps the model attains the best performance, implies the actual number of user groups with similar mobility patterns.

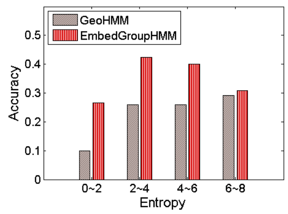
The parameter K is the number of hidden states from the *Group-level HMM* module. From Fig. 7 (c), we can observe that when $K < 6$, the performance is apparently lower than when $K >= 6$. This is because many different mobility behaviours are not properly distinguished when represented by a few hidden states (when K is too small). On the other hand, the training for each hidden state of HMM model is not sufficient when K is much greater than the actual number of mobility behaviours.

4.2.4 Prediction Discussion. Besides performance comparison and parameter experiment, we explore how user's attributes influence the accuracy of location prediction through our model. We select two attributes of user mobility: the number of different places the user visited, and the entropy of the trajectory which measures the irregularity of user's movement. The entropy S_{en} is given by,

$$S_{en} = \sum_{i=1}^{n} -P_i * log(P_i), \tag{23}$$



(a) Unique Locations Visited.



(b) Mobility Entropy.

Figure 8: Accuracy of Location predictions for users with different attributes.

where P_i is the proportion of frequency user visits $i-th$ place while n is the total number of different places the user has been to.

We divide all the users into three groups by the number of visited different places and into four groups by the entropy of the trajectory and separately calculate the accuracy for each group. From Fig. 8 (a), we can observe that the prediction accuracy of our model decreases while *GeoGaussHMM* increases as the number of different places the user has visited get larger. Therefore, we show that our model mainly improves the performance for users who visit fewer locations. From Fig. 8 (b), we can observe that the prediction accuracy of our model decreases as the entropy grows except when entropy is very small (which implies the case when the number of records of the user is very limited). This result is coherent to our common sense: the more irregular the trajectory, the harder it is to make prediction.

5 RELATED WORK

We summarize the closely related works from three aspects: trajectory-based mobility model, semantic-aware mobility model, and embedding-based spatial temporal knowledge discovery.

Trajectory-based mobility model: Extensive studies have been dedicated to model human mobility via large-scale trajectory data recorded by GPS, cellular towers and location-based service. Gonzalez et al. [9] study mobile cellular accessing trace and discover that human trajectories show a high degree of temporal and spatial regularity. Lu et al. [14] discover that the theoretical maximum predictability of human mobility is as high as 88%. Various works [8, 19–21] focus on mobility modelling for next location prediction. Baumann et al. [3] compare the performance of 18 prediction algorithms and present a model with high overall prediction accuracy which meanwhile reliably predicts transitions. So as to solve data sparsity problem in location prediction, Jeong et al. [11] propose a cluster-aided model which exploits past trajectories collected from all users while Mcinerney et al. [16] develop a Bayesian model of mobility in populations. One limitation of all these trajectory-based mobility models, however, is that this group of models do not properly capture semantics behind human mobility since they only take into account spatial and temporal information in trajectory data. Therefore they fail to provide insights as why people move from one location to another. In contrast, we develop a mobility model which jointly considers spatial, temporal and user motivation in trajectory data as a whole to understand human mobility.

Semantics-aware mobility model: Recently, several semantic-aware mobility models have been proposed [22] for spatial temporal data. The most relevant works are those on modelling semantic-rich location data from geo-tagged social media (GeoSM) as twitter and foursquare. Ye et al. [27] propose a mobility model to predict user activity at next step. Yuan et al. [28] propose a who+when+where+what model to jointly model user spatial-temporal topics. Zhang et al. [30] develop a group-level mobility model named GMove for GeoSM data, which includes a sampling-based keyword augmentation. Different from them, we incorporate representation learning method with Hidden Markov Model and propose a novel semantic-aware mobility model, which learns inner semantics embedded in human mobility in an unsupervised way instead of manually combining spatial, temporal and topic features. Our model thus achieves better performance than previous works.

Embedding-based spatial temporal knowledge discovery: Embedding, or representation learning is a category of unsupervised learning method that aims to extract effective and low-dimensional features from complicated and high-dimensional data [4, 17, 23, 25]. Recently representation learning methods have been used for spatial temporal data mining and knowledge discovery. Yao et al. [7] designed a recurrent neural network to capture the physical features of trajectories to detect trajectories that are similar in speed and acceleration patterns. Inspired by PTE [24], Zhang et al. [29] dynamically model the semantic meaning of spatial-temporal points based on their co-occurrence with the texts in social media's check-ins through constructing a spatial-temporal-textual network. For applications in location-based POI recommendation, graph-based representation learning method [26] and word2vec-inspired model [31] have been presented. Different from previous works, in this paper we first introduce representation learning method in mobility modelling and propose a semantic-aware model, which contributes to our understanding of the interplay between spatial, temporal and semantic aspects in human mobility and achieves better prediction performance.

6 CONCLUSION AND FUTURE WORK

In this paper, we proposed a semantics-aware Hidden Markov Model for human mobility modeling using large-scale semantic-rich spatial-temporal dataset. Distinct from existing studies, we took into account location, time, activity and user motivation behind human mobility as a whole. We first conducted multimodal embedding to jointly map these information into the same low-dimensional space with their correlations preserved. Then we used hidden Markov model to learn latent states and transitions between them in the embedding space. We also designed a vMF mixture model for clustering users so as to tackle data sparsity problem. We have evaluated our model on two datasets for the location and POI prediction, and it outperforms baseline methods significantly.

In the future, we plan to adapt our framework to take more semantic information (e.g. app usage and tweets) into account to better describe the user activity patterns. Moreover, the contradiction between highly-diversified human mobility pattern and sparsity of data, motivates us to further propose a model to train the HMM where users has their own transition probability matrix but sharing group-level hidden state parameters.

REFERENCES

[1] Arindam Banerjee, Inderjit S Dhillon, Joydeep Ghosh, and Suvrit Sra. 2005. Clustering on the unit hypersphere using von Mises-Fisher distributions. *Journal of Machine Learning Research* 6, Sep (2005), 1345–1382.

[2] Leonard E Baum, Ted Petrie, George Soules, and Norman Weiss. 1970. A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains. *The annals of mathematical statistics* 41, 1 (1970), 164–171.

[3] Paul Baumann, Wilhelm Kleiminger, and Silvia Santini. 2013. The influence of temporal and spatial features on the performance of next-place prediction algorithms. In *Proceedings of the 2013 ACM international joint conference on Pervasive and ubiquitous computing*. ACM, 449–458.

[4] Yoshua Bengio, Aaron Courville, and Pascal Vincent. 2013. Representation Learning: A Review and New Perspectives. *IEEE Transactions on Pattern Analysis Machine Intelligence* 35, 8 (2013), 1798.

[5] Budhaditya Deb and Prithwish Basu. 2015. Discovering latent semantic structure in human mobility traces. In *European Conference on Wireless Sensor Networks*. Springer, 84–103.

[6] Laurens Van Der Maaten and Geoffrey E Hinton. 2008. Visualizing Data using t-SNE. *Journal of Machine Learning Research* 9 (2008), 2579–2605.

[7] Yao Di, Zhang Chao, Zhu Zhihua, Huang Jianhui, and Bi Jingping. 2017. Trajectory Clustering via Deep Representation Learning. In *IJCNN*.

[8] Huiji Gao, Jiliang Tang, and Huan Liu. 2012. Mobile location prediction in spatio-temporal context. In *Nokia mobile data challenge workshop*, Vol. 41. 1–4.

[9] Marta C Gonzalez, Cesar A Hidalgo, and Albert-Laszlo Barabasi. 2008. Understanding individual human mobility patterns. *Nature* 453, 7196 (2008), 779–782.

[10] Jun Han and Claudio Moraga. 1995. The influence of the sigmoid function parameters on the speed of backpropagation learning. *From Natural to Artificial Neural Computation* (1995), 195–201.

[11] Jaeseong Jeong, Mathieu Leconte, and Alexandre Proutiere. 2016. Cluster-aided mobility predictions. In *Computer Communications, IEEE INFOCOM 2016-The 35th Annual IEEE International Conference on*. IEEE, 1–9.

[12] S. Jiang, Y. Yang, S Gupta, D Veneziano, S Athavale, and M. C. Gonzlez. 2016. The TimeGeo modeling framework for urban motility without travel surveys. *Proceedings of the National Academy of Sciences of the United States of America* 113, 37 (2016), E5370.

[13] Solomon Kullback and Richard A Leibler. 1951. On information and sufficiency. *The annals of mathematical statistics* 22, 1 (1951), 79–86.

[14] Xin Lu, Erik Wetter, Nita Bharti, Andrew J Tatem, and Linus Bengtsson. 2013. Approaching the limit of predictability in human mobility. *Scientific reports* 3 (2013).

[15] Wesley Mathew, Ruben Raposo, and Bruno Martins. 2012. Predicting future locations with hidden Markov models. In *Proceedings of the 2012 ACM conference on ubiquitous computing*. ACM, 911–918.

[16] James McInerney, Jiangchuan Zheng, Alex Rogers, and Nicholas R Jennings. 2013. Modelling heterogeneous location habits in human populations for location prediction under data sparsity. In *Proceedings of the 2013 ACM international joint conference on Pervasive and ubiquitous computing*. ACM, 469–478.

[17] Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. Efficient Estimation of Word Representations in Vector Space. *Computer Science* (2013).

[18] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. Distributed Representations of Words and Phrases and their Compositionality. *Advances in Neural Information Processing Systems* 26 (2013), 3111–3119.

[19] Anna Monreale, Fabio Pinelli, Roberto Trasarti, and Fosca Giannotti. 2009. Wherenext: a location predictor on trajectory pattern mining. In *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 637–646.

[20] Mikołaj Morzy. 2007. Mining frequent trajectories of moving objects for location prediction. *Machine Learning and Data Mining in Pattern Recognition* (2007), 667–680.

[21] Anastasios Noulas, Salvatore Scellato, Neal Lathia, and Cecilia Mascolo. 2012. Mining user mobility features for next place prediction in location-based services. In *Data mining (ICDM), 2012 IEEE 12th international conference on*. IEEE, 1038–1043.

[22] Christine Parent, Stefano Spaccapietra, Chiara Renso, Gennady Andrienko, Natalia Andrienko, Vania Bogorny, Maria Luisa Damiani, Aris Gkoulalas-Divanis, Jose Macedo, and Nikos Pelekis. 2013. Semantic trajectories modeling and analysis. *Acm Computing Surveys* 45, 4 (2013), 42.

[23] Jeffrey Pennington, Richard Socher, and Christopher Manning. 2014. Glove: Global Vectors for Word Representation. In *Conference on Empirical Methods in Natural Language Processing*. 1532–1543.

[24] Jian Tang, Meng Qu, and Qiaozhu Mei. 2015. Pte: Predictive text embedding through large-scale heterogeneous text networks. In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, 1165–1174.

[25] Jian Tang, Meng Qu, Mingzhe Wang, Ming Zhang, Jun Yan, and Qiaozhu Mei. 2015. Line: Large-scale information network embedding. In *Proceedings of the 24th International Conference on World Wide Web*. International World Wide Web Conferences Steering Committee, 1067–1077.

[26] Min Xie, Hongzhi Yin, Hao Wang, Fanjiang Xu, Weitong Chen, and Sen Wang. 2016. Learning graph-based poi embedding for location-based recommendation. In *Proceedings of the 25th ACM International on Conference on Information and Knowledge Management*. ACM, 15–24.

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

[28] Quan Yuan, Gao Cong, Zongyang Ma, Aixin Sun, and Nadia Magnenat Thalmann. 2013. Who, where, when and what: discover spatio-temporal topics for twitter users. In *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 605–613.

[29] Chao Zhang, Keyang Zhang, Quan Yuan, Haoruo Peng, Yu Zheng, Tim Hanratty, Shaowen Wang, and Jiawei Han. 2017. Regions, periods, activities: Uncovering urban dynamics via cross-modal representation learning. In *Proceedings of the 26th International Conference on World Wide Web*. International World Wide Web Conferences Steering Committee, 361–370.

[30] Chao Zhang, Keyang Zhang, Quan Yuan, Luming Zhang, Tim Hanratty, and Jiawei Han. 2016. Gmove: Group-level mobility modeling using geo-tagged social media. In *KDD: proceedings. International Conference on Knowledge Discovery & Data Mining*, Vol. 2016. NIH Public Access, 1305.

[31] Shenglin Zhao, Tong Zhao, Irwin King, and Michael R Lyu. 2017. Geo-Teaser: Geo-Temporal Sequential Embedding Rank for Point-of-interest Recommendation. In *Proceedings of the 26th International Conference on World Wide Web Companion*. International World Wide Web Conferences Steering Committee, 153–162.