Author/s:
Shi, H;Li, Y;Cao, H;Zhou, X;Zhang, C;Kostakos, V

Title:
Semantics-Aware Hidden Markov Model for Human Mobility

Date:
2021-03-01

Citation:
Shi, H., Li, Y., Cao, H., Zhou, X., Zhang, C. & Kostakos, V. (2021). Semantics-Aware Hidden Markov Model for Human Mobility. IEEE Transactions on Knowledge and Data Engineering, 33 (3), pp.1183-1194. https://doi.org/10.1109/TKDE.2019.2937296.

Persistent Link:
https://hdl.handle.net/11343/272718

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TKDE.2019.2937296, IEEE Transactions on Knowledge and Data Engineering

Semantics-Aware Hidden Markov Model for Human Mobility

Hongzhi Shi, Yong Li, *Senior Member, IEEE*
Hancheng Cao, Xiangxin Zhou, Chao Zhang, Vassilis Kostakos,

Abstract—Understanding human mobility benefits numerous applications such as urban planning, traffic control and city management. Previous work mainly focuses on modeling spatial and temporal patterns of human mobility. However, the semantics of trajectory are ignored, thus failing to model people's motivation behind mobility. In this paper, we propose a novel semantics-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 location 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 prediction. The results show that our model outperforms state-of-the-art mobility models with higher prediction accuracy and much higher efficiency.

Index Terms—User grouping, human mobility modeling, multimodal embedding, hidden Markov model.

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 regarded as one of the fundamental tasks 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 [1], [2], [3], [4], [5], [6], [7], [8], [9], [10]. The majority of them focus on modelling the spatial and temporal patterns. Human mobility is generally modelled as a stochastic process around fixed point [1] and various models for next location prediction [2], [3], [4], [5], [6] have been proposed. The main shortcoming of these mobility models, however, is that they overlook the activity (often referred to as the semantics of trajectory [11], [12], [13]) a person engages in at a location within a certain time, i.e., they are not capable of explaining people's motivation behind mobility. For instance, people who appear at nearby locations with different intents (e. g. a person going to office for work and a person going to the movie for entertainment in the same neighborhood) will be considered the same, while people visiting different locations for similar purposes (e. g. white-collar A goes to supermarket S_1 after work in region R_1 while white-collar B goes to supermarket S_2 after work in region R_2) are considered different.

To tackle this problem, recently a few semantics-aware mobility models [7], [8], [9] have been proposed, which attempt to jointly model spatial, temporal and semantic aspects. However, they manually combine spatial, temporal and topic features to take semantics into account, which still fail to properly distinguish motivation between users. Therefore, the problem of semantics-aware mobility modelling remains very much an open question.

Instead, in this paper, we aim at learning inner semantics embedded in human mobility in an unsupervised way to consider all the factors as a whole. We propose a novel semantics-aware mobility model using large-scale semantics-rich spatial-temporal data – from the location-based social networks such as 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 supermarket S_1 in region R_1 after work and white-collar B to supermarket S_2 in region R_2 after work are similar in motivation because they both go for shopping.

- *H. Shi, Y. Li and X. Zhou are with Beijing National Research Center for Information Science and Technology (BNRist), Department of Electronic Engineering, Tsinghua University, Beijing 100084, China (E-mail: shz17@mails.tsinghua.edu.cn, xx-zhou16@mails.tsinghua.edu.cn, liyong07@tsinghua.edu.cn.). H. Cao is with department of computer science, Stanford University (E-mail: hanchcao@cs.stanford.edu), this work was done when H. Cao was with Beijing National Research Center for Information Science and Technology (BNRist), Department of Electronic Engineering, Tsinghua University, Beijing 100084, China. C. Zhang is with School of Computational Science and Engineering, Georgia Institute of Technology (E-mail: czhang82@illinois.edu). V. Kostakos is with School of Computing and Information Systems, The University of Melbourne (E-mail: vassilis.easychair@kostakos.org).*

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TKDE.2019.2937296, IEEE Transactions on Knowledge and Data Engineering

On the other hand, the model is also able to capture the difference between a person going to office and another going for a movie in nearby locations, since they move 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 enable various applications. First of all, they are well-suited for next location prediction [8], 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 the 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 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 this 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 location 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 obtain an ensemble of high-quality HMM models. Finally, we project the 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 with 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. We train an ensemble of HMMs 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 prediction. We observe that our model outperforms baselines with higher prediction accuracy and incurs lower training cost.

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

Previous works mostly focus on modelling spatial-temporal patterns in trajectory regardless of semantics, i.e., clustering users by extracting features from the geographical trajectory, learning Markov model in geographical space, etc. However, we aim to propose a novel semantics-aware mobility model. In order to extract semantics, the type of 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. Generally, the type of POI with explicit semantics, such as shopping, school, tourist attraction, etc., that the user visited indicates the motivation why the user went to the location. The key idea of our work, therefore, is to integrate semantic information, which can be learnt from POI, 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 in the

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TKDE.2019.2937296, IEEE Transactions on Knowledge and Data Engineering

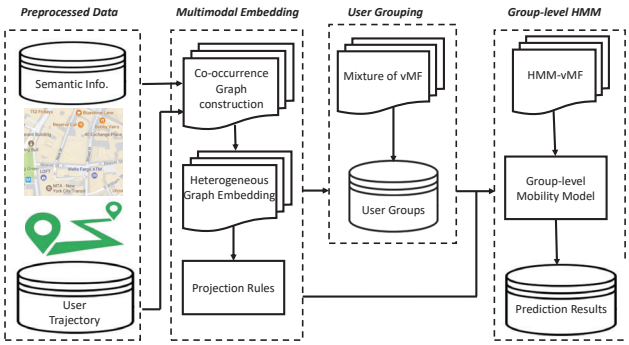


Fig. 1. Overview of the proposed embedding based group-level human mobility model.

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 are preserved in the latent embedding space. Thus we can train an HMM model whose observable states are the embedding vectors representing locations, 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.

2.2 Model Overview

Based on the approach discussed above, we introduce 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 clusters users sharing similar mobility and life patterns and the *Group-level Hidden Markov Model* module learns human mobility 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.

Multimodal Embedding module builds the structure of user (u), temporal (t), spatial ((l_o, l_a)) and user motivation/semantics (P) information. 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 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 [14], [15], we model each cluster of users as a von Mises-Fisher (vMF) distribution in the latent space. Naturally, we use mixture-of-vMFs model [16] 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 model 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 calculate the scores of locations in the candidate and obtain the top K results.

TABLE 1
The introduction of basic notations.

t_i	the i-th time slot
$\mathbf{L}$	set of areas
l^u	a location of user u
l_a	latitude
l_o	longitude
$\mathbf{U}$	the set of all users u
y^u	records of user u
$\mathbf{H}$	semantic information set
P	the set of POIs
p	a POI

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 $\mathbf{L}$. For each user u in the set of all users $\mathbf{U}$, $y^u = (y_1^u, ..., y_i^u, ..., y_{N^u}^u)$ denotes history trajectories of user u, where N^u denotes the number of sampling points in user u's trajectory. $y_i^u = (u, t_i^u, l_i^u)$ denotes the location l_i^u being visited by user u at time slot t_i^u and $l_i^u = (l_a, l_o)_i^u$. (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 types, i.e, the check-in POI in the history visited l_k. For evaluation, our model predicts the next state (location) $y_{N^u+1}^u = (t_{N^u+1}^u, l_{N^u+1}^u)$ of each user u, based on the past trajectory $y^u = (y_1^u, ..., y_i^u, ..., y_{N^u}^u)$. The basic notations are reported in Table 1.

3 Method

In this section, we first design a multimodal embedding module to capture the diversified semantics, and then present the user-grouping based HMM, which learns fine-grained semantics-aware mobility behaviors in the embedding space.

3.1 Multimodal Embedding

The designed multimodal embedding module jointly maps the user, time, location, and semantic information into the same low-dimensional space with their correlations preserved. While the semantics are natural POI types P for embedding, space and time 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 l ($500m * 500m$ grid). Similarly, we break one day into 24 hours distinguished by weekday and weekend and consider every hour as a basic temporal unit t (totally 48 units). Based on this division, the embedding module extracts the correlations between user, time, location and POI type as co-occurrence relationships,

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TKDE.2019.2937296, IEEE Transactions on Knowledge and Data Engineering

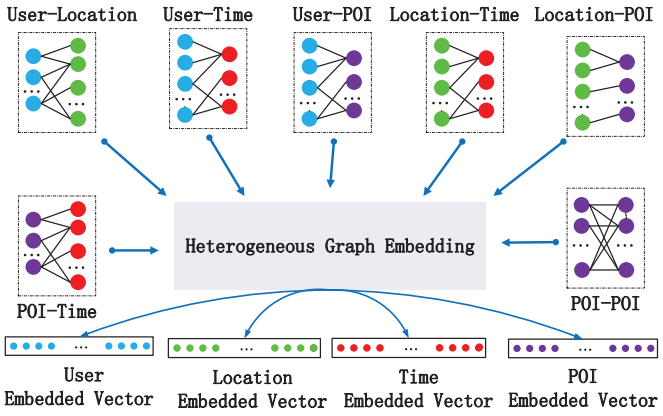


Fig. 2. Illustration of the details of our representation learning model. The co-occurrence relationships construct 7 sub-graphs, which are jointly embedded with graph-based method.

and then embeds all the co-occurrence relationships into one latent space to encode the human mobility intentions into vectors, as shown in Fig. 2.

3.1.1 Co-occurrence Relationship

The *co-occurrence* relationship describes the times of co-occurrences between different information. In our data, each record is composed of user, time, location and POI type, and the co-occurrence happens when two different kinds of units appear in one record. This relationship reflects the intrinsic correlations between different information units. Since the graph is a very natural data structure that describes the relationship between different units, we represent the co-occurrence relationship through constructing a heterogeneous graph. Specifically, the nodes of the graph include who (u), when (t), where ((l_a, l_o)) and why (P). The edge weights of the graph indicate the number of co-occurrences between two units.

3.1.2 Heterogeneous Graph Learning

Based on the *co-occurrence* relationship, we express their relationships with the edges and weights. In the graph, there exist four different node types corresponding to four unit (information) types (user, time, location and POI type). Each *co-occurrence* relationship constructs one edge, whose weight is set to be the counts. Besides the explicit relationships, the graph also keeps implicit interactions among units. The implicit interaction means that two nodes are not directly connected but share a lot of common neighbors. The more same neighbors two nodes share, the more semantically close they are. Correspondingly we will embed them closer in the semantic latent space to pave the way for the following models to capture the semantics. Based on this principle, we conduct the heterogeneous graph learning to project them into a common semantic space. Note that, in the embedding space, these nodes should be close in the cosine distance metric. 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 these model distributions.

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

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

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.

In order to minimize the distance between the embedding-based distributions and truly observed distributions, we define the loss function for the sub-graph G_{UV} as follows,

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

where $d_{KL}()$ is Kullback-Leibler divergence [17]. With four different nodes representing user(U), temporal (T), spatial (S) and POI type (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 adapted for computational efficiency [14]. 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 need to minimize an adapted loss function as

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

where $\sigma(\cdot)$ represents the sigmod function [18].

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 the latent space. Cosine distance is more effective than Euclidean distance for measuring the semantic proximity in the embedding space, i.e., only the directions of the embedding vectors matter, which is demonstrated in [14], [15]. Also, there are some semantic models that use von Mises-Fisher (vMF) distribution in word embedding space [19], [20] and multimodal embedding space [21].

Thus, we normalize all the embedding vectors to vectors with lengths of 1, i.e., projecting them into a $(d-1)$-dimensional spherical space. For such vectors on a unit sphere, we use vMF to model each cluster of users' vectors

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TKDE.2019.2937296, IEEE Transactions on Knowledge and Data Engineering

in the latent space. For a d-dimensional unit vector x that follows d-variate vMF distribution, its probability density function is given by

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

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. Note that, $C_d(\kappa)$ is obtained by normalization on the $(d-1)$-dimensional sphere instead of the whole d-dimensional space.

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

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

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 paremeters. We first update α_h given by,

$$\alpha_h = \sum_{i=1}^{n} p(h|x_i,\mu,\kappa)/N, \tag{11}$$

We then calculate r_h as follows,

$$r_h = \frac{\sum_{i=1}^{n} x_i p(h|x_i,\mu,\kappa)}{\sum_{i=1}^{n} p(h|x_i,\mu,\kappa)}, \tag{12}$$

at last, we update the parameter μ_h and κ_h based on r_h for each cluster given by,

$$\mu_h = \frac{r_h}{\|r_h\|}; \kappa_h = \frac{\|r_h\| d - \|r_h\|^3}{1 - \|r_h\|^2}. \tag{13}$$

When the difference of total probability of this model before and after one iteration is less than a threshold, this iterative process terminates.

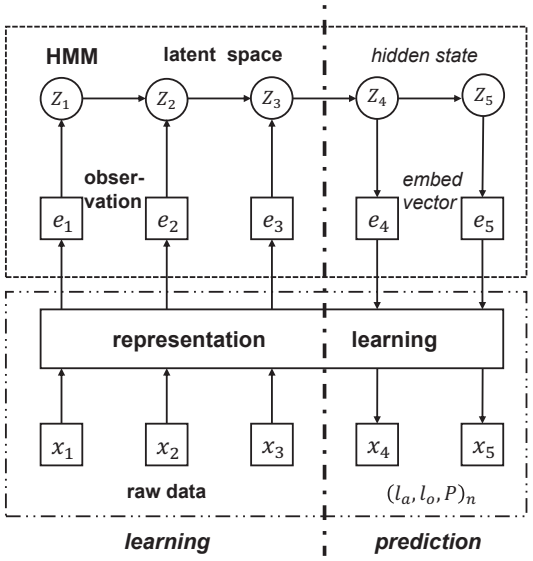


Fig. 3. Illustration of the details of HMM-based prediction model in the latent space and its relationship with the physical locations.

3.2.2 HMM-based model

Based on the representation vectors and the user groups, we design an HMM for each group of users to model the transitions among trajectories in the semantic latent space. It chooses the embedding vectors representing locations as observations to model the sequence as shown in Fig. 3. The proximity of semantic vectors characterizing the activity of users should also be measured by the cosine similarity like the users' representation vectors [14], [15]. Thus, we utilize vMF distribution as the emission probability of each hidden state.

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 transition 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 $\mathbf{B} = \{p_{vMF_i}(e_j)\}$, where e_j is the emission from a hidden state $z_j = i$ into the embedding space.

Then, our model is parameterized by $\Phi = \{\mathbf{\Pi}, \mathbf{A}, \mathbf{B}\}$. 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,

$$p(\mathbf{E}|\mathbf{Z},\Phi) = \prod_{i=1}^{N} p_{vMF_{z_i}}(e_i). \tag{14}$$

The cumulative occurrence probability of observation sequence $\mathbf{E}$ is expressed as follows,

$$p(\mathbf{E}|\Phi) = \sum_{\mathbf{Z}} p(\mathbf{E}|\mathbf{Z},\Phi) \cdot p(\mathbf{Z}|\Phi) \tag{15}$$

$$= \sum_{\mathbf{Z}} \pi_{i_1} \cdot \prod_{j=1}^{N-1} p_{vMF_i}(e_j) a_{i_j i_{j+1}} \cdot p_{vMF_i}(e_N). \tag{16}$$

The main difference between our proposed HMM and the tradition HMM is that we set the emission function of

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TKDE.2019.2937296, IEEE Transactions on Knowledge and Data Engineering

HMM as the vMF and set the observation as the embedding vectors representing locations instead of the locations' coordinates. By this way, we take rich semantics into account.

We adapt the Baum-Welch algorithm, an Expectation-Maximization (EM) procedure for HMM, to estimate the parameters in the embedding space. We first define two auxiliary probabilities:

$$\xi_t(i,j) = p(z_{t+1} = j, z_t = i|\mathbf{E}, \Phi), \tag{17}$$
$$\gamma_t(i) = p(z_t = i|\mathbf{E}, \Phi), \tag{18}$$

where $t = 1, 2..., N$. To calculate efficiently, we exploit a forward-backward procedure [22] to calculate these two probabilities. The forward probability $\alpha_t(i)$ is

$$\alpha_t(i) = p(e_1, e_2, ..., e_t, z_t = i|\Phi). \tag{19}$$

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

$$\alpha_{t+1}(j) = [\sum_{i=1}^{K} \alpha_t(i)a_{ij}]b_j(e_{t+1}), \tag{20}$$

and initial values are

$$\alpha_t(i) = \pi_i b_i(e_1). \tag{21}$$

The backward probability is

$$\beta_t(i) = p(e_{t+1}, e_{t+2}, ..., e_N|z_t = i, \Phi), \tag{22}$$

which can be calculated as follows,

$$\beta_t(i) = \sum_{j=1}^{K} a_{ij}b_j(e_{t+1})\beta_{t+1}(j), \tag{23}$$

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

$$\xi_t(i,j) = \frac{\alpha_t(i)a_{ij}b_j(e_{t+1})\beta_{t+1}(j)}{\sum_{m=1}^{K}\sum_{n=1}^{K}\alpha_t(m)a_{mn}b_n(e_{t+1})\beta_{t+1}(n)}. \tag{24}$$

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

$$\gamma_t(i) = \frac{\alpha_t(i)\beta_t(i)}{\sum_{j=1}^{K}\alpha_t(j)\beta_t(j)}. \tag{25}$$

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

$$\pi_i = \gamma_1(i);\ a_{ij} = \frac{\sum_{t=1}^{K-1}\xi_t(i,j)}{\sum_{t=1}^{K-1}\gamma_t(i)}; \tag{26}$$
$$r_i = \frac{\sum_{t=1}^{K}\gamma_t(i)e_t}{\sum_{t=1}^{K}\gamma_t(i)}; \mu_i = \frac{r_i}{\|r_i\|}; \kappa_i = \frac{\|r_i\|d - \|r_i\|^3}{1 - \|r_i\|^2}. \tag{27}$$

When the training process terminates, we obtain the HMM based semantics-aware mobility model.

In order to leverage the model for next location prediction, we construct a set of length-2 sequences (e_N, e_{N+1}) where $e_N = E_f(l_N^u)$ and $e_{N+1} = E_f(l_{N+1}^u)$ for the locations in candidates set. Then, we calculate the probability of generating such a sequence from the trained model as the score S of the sequences given by

$$S(l_{N+1}^u) \tag{28}$$
$$= p(e_N, e_{N+1}|\Phi) \tag{29}$$
$$= \sum_{m=1}^{K}\sum_{n=1}^{K}\pi p_{vMF_m}(e_N)a_{12}p_{vMF_n}(e_{N+1}) \tag{30}$$
$$= \sum_{m=1}^{K}\sum_{n=1}^{K}\pi_1 a_{12}C_d(\kappa_m)e^{\kappa_m\mu_m^T e_N}C_d(\kappa_n)e^{\kappa_n\mu_n^T e_{N+1}}. \tag{31}$$

where Φ is the set of parameters of HMM. Thus, we define the score of next location as the probability of generating such a sequence from the our trained model and generate list of locations with top K scores.

4 Evaluation

In this section, we evaluate our proposed model through next location prediction on two real-world large-scale datasets. We first introduce the experimental settings including datasets, baseline algorithms, parameters and hardware. Then, we evaluate our model in the following four parts:

- Presenting case study and the corresponding insightful results to validate the ability of our model in discovering semantic mobility patterns.
- Demonstrating the effectiveness and efficiency of our method compared with baselines, including state-of-the-art 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 real-world 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. Overall, the utilized dataset is collected from $7,000$ anonymous users, who are active during Sept. 17th to Oct. 31st, 2016 in Beijing. Each record consists of the following fields: the anonymized user, time, location of GPS coordinates, and the associated POI with type information.

Check-in Dataset: This publicly available dataset comes from Foursquare, a location-based service application. It includes $187,568$ records of $5,630$ active users from Feb. 25th, 2010 to Jan. 19th, 2011 in New York. Each record consists of the following fields: the user ID, the time stamp, location, and the type of POI that the user check in. The POI types include travel, shop, professional, college, residence, outdoors, food, arts.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TKDE.2019.2937296, IEEE Transactions on Knowledge and Data Engineering

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 solutions including the state-of-the-art methods.

Law [24] models the human mobility as a Lévy flight with long-tailed distributions.

GeoHMM [25] [26] trains one HMM for all users' trajectory, where each hidden state generates locations by a Gaussian distribution, which is a classical method for trajectory prediction.

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

EmbedVmfHMM 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 and improve the efficiency.

Gmove [8] is the state-of-the-art mobility model. It constructs several HMMs and assigns users to each HMM by a soft label proportional to the probability of drawing the trajectory from the HMM. For comparing the user grouping part, we set each HMM structure as *EmbedVmfHMM*.

On the other hand, our model, denoted by ***EmbedGroupHMM***, performs a mixture of vMF which clusters users in the latent space, and trains one model using *EmbedVmfHMM* for each group of users.

We compare the performance of the *EmbedGaussHMM* method and *EmbedVmfHMM* method to show which kind of distribution is more appropriate for the emission probability function in the embedding space. To show the advantage of our method for user grouping in embedding space, we compare *EmbedGroupHMM* with *Gmove* [8], which clusters users based on the transitions.

4.1.3 Evaluation Setup

We partition each dataset into the training set and testing set. For the app 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 check-in 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 semantics-aware location prediction as the task for evaluation. Formally, given a trajectory $y^u = (y_1^u, ...y_i^u..., y_{N^u}^u)$ as ground truth, we use y_i^u to predict y_{i+1}^u. First, we use y_{i+1}^u to generate a set of candidates. Specifically, we select the locations in the dataset that are less than 3km from the true location as candidate sets. Then we calculate the score of the combination of y_i^u and each candidate in the sets. The score is calculated by (28), which is the probability of generating such a sequence (y_i^u, y_{i+1}^u) by the mobility model with the learned parameters. At last, we sort all the candidates in descending order of score and calculate the accuracy of top K. The higher the accuracy is, the better the mobility model is.

In terms of the next place prediction enabled by our model, it 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 $E = 50$, $G = 20$, $K = 10$ for the check-in dataset by parameter tuning.

Note that, the core strength of our model is semantics-aware mobility modeling rather than trajectory prediction. Some techniques are specifically optimized for trajectory prediction, however, our model is more about semantic mobility pattern discovery, which is unsupervised and interpretable. So, we didn't add them into the baselines. Also, we don't claim that our method achieves the best performance for trajectory prediction compared to the state-of-the-art location prediction techniques.

To clearly demonstrate their effect, we evaluate the performance by varying one parameter while fixing others. We implemented our method (except the embedding part which is adapted by LINE [27] implemented in C++) and the baseline methods in JAVA and conducted all the experiments on a computer with 4.0 GHz Intel Core i7 CPU and 64GB memory.

4.2 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 projected into a common embedding space, which facilitates the comparison of semantic proximity. Therefore, we can infer the semantics of hidden states by finding the nearby information units in 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 [23]. 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 mean 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.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TKDE.2019.2937296, IEEE Transactions on Knowledge and Data Engineering

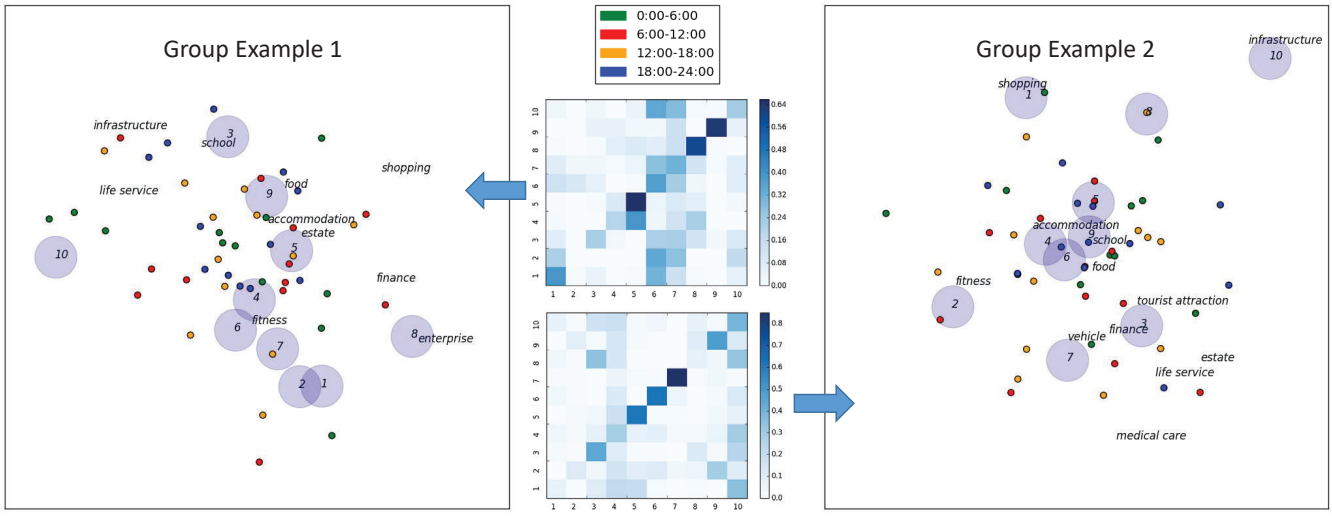
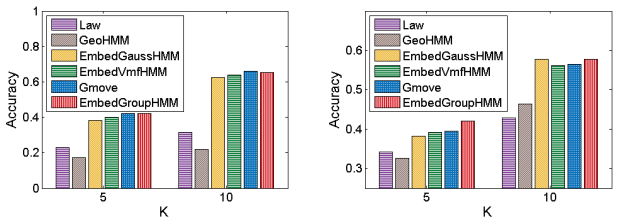


Fig. 4. Two examples of the user groups. We map the embedding vectors of time, POI types and the central vector of each hidden state on a 2D plane with t-SNE [23]. The two heat maps in the middle represent HMM transition probability matrixes.

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 the hidden state 10 which is coherent with the character of transportation. To avoid confusion, note that the POI type 'vehicle' includes petrol station, auto shop, etc.



(a) App Collected Dataset. (b) Check-in Dataset.

Fig. 5. Prediction Accuracy of Top K.

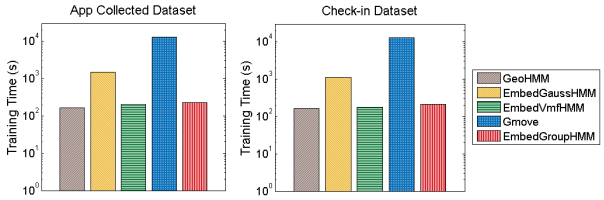


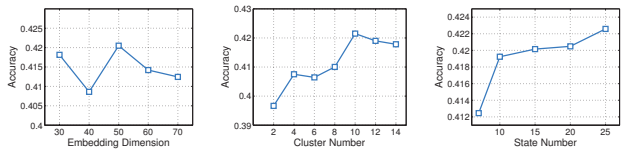
Fig. 6. Training time on two datasets.

4.3 Performance Comparison

To demonstrate the effectiveness and efficiency of our proposed model, we test it on two real large-scale datasets: app collected dataset and Foursquare. For both datasets, we show the accuracy of top 5 and top 10 results in Fig.5. For both datasets, the methods that take semantics into account by embedding significantly improve the performance for prediction. Also, the *EmbedVmfHMM* is better than *EmbedGaussHMM* while the speed is much faster shown in Fig. 6, which shows the superiority of vMF. Comparing our method with the *Gmove* [8], we can observe that our user grouping method achieves similar results while the speed is more than 80 times faster shown in Fig. 6. Finally, compared with *EmbedVmfHMM*, *EmbedGroupHMM* group users into G groups and train one HMM for each group. We can observe a significant improvement in accuracy by user-grouping.

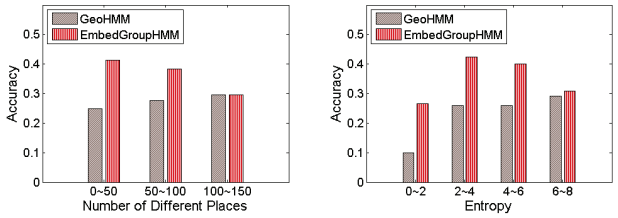
The efficiency of our proposed model is stable on both datasets. The time complexity of embedding part of our method is $O(DN|E|)$, where the D is the dimension of the embedding space, the N is the number of negative sampling and the $|E|$ is the number of edges in the graph. This part typically costs couples of minutes, which is adapted by LINE demonstrated that can scale for large datasets in [27]. In Fig. 6, we report the training time of our method and baselines (not counting embedding part) on both datasets with a logarithmic y-axis. *LAW* does not need to train the model, so we don't report it. We find that *EmbedVmfHMM* is more than 7 times faster than *EmbedGaussHMM*, because estimating parameters of vMF is faster than Gaussian distribution. Also, *Gmove* groups the users like our proposed method, but our method is more than 80 times faster than *Gmove*. Because our method only needs to clustering the users in embedding space by one time, while user grouping in *Gmove* is an iterative process. Typically, the user grouping and HMM in our algorithm typically costs couples of minutes. Due to the nature of HMM training, the time complexity is quadratic in K. Compared to training one HMM for all users, training one HMM for each group can

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TKDE.2019.2937296, IEEE Transactions on Knowledge and Data Engineering



(a) Effect of D. (b) Effect of G. (c) Effect of K.

Fig. 7. Effects of parameters.



(a) Unique Locations Visited. (b) Mobility Entropy.

Fig. 8. Accuracy of Location predictions for users with different attributes.

help decrease K to reduce time complexity. Also, the HMM training processes for different groups are independent, so it can be easily parallelized. Our method costs the similar time with the classical method *GeoHMM*. Overall, compared to the state-of-the-art baseline methods, our proposed method either achieves much better results or costs much less time.

4.4 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 respectively come from the three modules:

- The embedding dimension D in *Multimodal Embedding* module.
- The number of user groups G in *User Grouping* module.
- The number of hidden state K in *Group-level HMM* module.

We use the accuracy of top 5 locations prediction as the main performance indicator to tune the parameters. To save space, we only report the process of tuning parameters on app collected dataset. We show the experiment details as follows.

From Fig. 7 (a), we can observe that the accuracy is highest when $D = 50$. D decides the quality we embed the semantic information into our model. From Fig. 7 (b), we can observe that our model obtains the best performance when $G = 10$. 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. From Fig. 7 (c), we can observe that when $K < 10$, the performance is apparently lower than when $K >= 10$. This is because many different mobility behaviors are not properly distinguished when represented by a few hidden states (when K is too small).

4.5 Performance on Different User Groups

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 [28] of the trajectory which measures the irregularity of user's movement. The entropy S_{en} is given by,

$$S_{en} = \sum_{i=1}^{n} -P_i * log(P_i), \tag{32}$$

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 predictions.

4.6 Discussions

There are strong relations between users, time, space and POI types. On one hand, traditional methods cannot effectively capture all these factors for human mobility modeling. On the other hand, although deep learning can lead to a good performance, it often suffers from its poor interpretability, which is important for mobility modeling. Our model uses graph embedding method, a representation learning method, which successively takes all factors into account including times, locations, users, POI types to improve mobility modeling. Note that the POI types can also be replaced by words in tweeter, app usage information, or anything that can bring semantics to the trajectory.

Our method can automatically extract features from data including semantics. Based on the constructed semantics-rich latent space, we improve the quality of both user grouping and mobility modeling. For both user grouping and mobility modeling, the vectors near the centers of hidden states, which represent different kinds of information, can imply the semantics of them, so we can understand the results better. The mobility patterns of several user groups learned by our model can help us understand human mobility better and benefit personalized location-based services.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TKDE.2019.2937296, IEEE Transactions on Knowledge and Data Engineering

In this paper, we only extract the co-occurrence relationship of different kinds of items. We also can add more complex relationship such as the proximity between times and the transmission relationship between locations. Also, we can use the graph embedding method to bring semantics to improve other traditional trajectory tasks, such as local events detection and urban function discovery.

5 Applications and Discussion

In this section, we discuss the wide range of applicability of our proposed system and introduce how to adapt it to different scenarios or different data from three aspects.

5.1 Adaption to Different Semantic Fields

Our proposed system can use semantic fields in the trajectory records to improve the prediction accuracy. Such semantic fields include not only POI types but also various fields. For example, for the twitter data, each record consists of a time stamp, a location and a message. Each message can be regarded as a set of words, which can bring semantics like the POI in our data. We can construct a heterogeneous graph with nodes representing time stamps, locations and words. Even the mobile app usage attached on the trajectory may provide semantics, which may improve prediction accuracy.

If the data have more than one semantic fields, we can also construct a larger heterogeneous graph that contains more types of nodes, such as time units, locations, POIs and words in messages. Overall, our system is a uniform framework for introducing semantics to the raw trajectories to improve prediction accuracy, which can be easily adapted to various trajectories dataset with different semantic information.

5.2 Temporal Density

For different types of trajectory data, the temporal density of one user's trajectory varies from several records per day to hundreds of records per day. There is a difference in preprocessing for dense and sparse trajectory data. Our dataset collected by a localization platform is relatively dense. There are many continues records at the same locations with very little time difference. Also, there are some records generated when the user is moving such as walking or driving a car. Such records are very detailed, however, we want to use meaningful locations. Thus, we extract several stays where the user mainly spent his time by setting time and spatial thresholds and filter out the pass-by points.

For the sparse dataset, such as check-in data, there may be only several records in one day for one user. However, there is a strong correlation when the time between the two records is small and there is a very weak correlation when the time is long. Thus, we need to divide one whole trajectory into several segments in which the time between two adjacent records is always less than a time threshold such as one hour. Thus, we can use the segments of trajectories for HMM learning. For sparse data, there is another problem that the data may be not sufficient for training one HMM for each user. To tackle this problem, we need to cluster users with similar mobility into one group to make the data more sufficient for HMM learning.

5.3 Space Partition

There are different ways to partition the two-dimensional space composed of latitudes and longitudes for different demands, such as the grids or blocks. A data-driven way to partition the geological space is to cluster the locations by the algorithms of DBSCAN [29] or mean-shift [30] etc., which can automatically detect the densely located area. No matter how to partition the space, we will obtain regions as the nodes in the graph for embedding, where our framework is applicable.

Naturally, the spatial granularity will directly effect on the precision of prediction. Too coarse-grained regions significantly limit the applications of prediction, while too fine-grained regions lead to a numerous number of spatial units, which decrease the co-occurrence edges linked to it, i.e., make the graph sparse. In general, we partition the space according to the prediction demand.

6 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 [1], [2], [25], [28], [31], [32], [33]. Gonzalez et al. [1] study mobile cellular accessing trace and discover that human trajectories show a high degree of temporal and spatial regularity. Lu et al. [31] discover that the theoretical maximum predictability of human mobility is as high as 88%. Various works [2], [3], [4], [5] focus on mobility modelling for next location prediction. Baumann et al. [6] 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. [34] propose a cluster-aided model which exploits past trajectories collected from all users while Mcinerney et al. [35] develop a Bayesian model of mobility in populations. Wang et al. [36] propose a method for location prediction which takes both the mobility regularity and social conformity into account. Liu et al. [37] extend RNN into spatial-temporal recurrent neural networks (ST-RNN) for next place prediction by temporal and spatial information. Feng et al. [38] use RNN with attention model for location prediction, which can capture the multi-order properties in trajectories. 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,

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TKDE.2019.2937296, IEEE Transactions on Knowledge and Data Engineering

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 [39], [40] 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. [7] propose a mobility model to predict user activity at next step. Yuan et al. [9] propose a who+when+where+what model to jointly model user spatial-temporal topics. Zhang et al. [8] 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 [27], [41], [42], [43]. Recently representation learning methods have been used for spatial-temporal data mining and knowledge discovery. Yao et al. [44] 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 [45], Zhang et al. [15] 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. Yan et al. [46] adapt skip-gram model [42] for learning the representations of place types, and Cao et al. [11] propose representation learning based framework to embed trajectory semantics for living pattern recognition in population. Zhang et al. [21] propose a embedding-based method for online local event detection. For applications in location-based POI recommendation, graph-based representation learning method [47] and word2vec-inspired model [48] have been presented. Furthermore, Qian et al. [49] conduct knowledge graph embedding to capture the semantics. Yin et al. [50] aim to tackle the problem of the sparsity of user-POI matrix and cold-start issues for POI recommendation. Note that, location-based POI recommendation aims to recommend some POIs given user, time and location, however, mobility modeling focus on the transition patterns, which often predict the location given the previous location and the user. Yin et al. [51] exploit graph embedding for joint event partner recommendation. Chen et al. [52] propose a novel method of heterogenous information network embedding for link prediction. Zhao et al. [53] propose spatial-temporal latent ranking to model the impact of time for POI recommendation. Different from previous works, in this paper we first introduce representation learning method in mobility modeling 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.

This paper is an extended work comparing to its conference version [54], which adds detailed algorithms to train our proposed model, additional experiments of performance comparison on different user groups and discussions with deeper insights of our proposed model.

7 Conclusion and Future Work

In this paper, we proposed a semantics-aware hidden Markov model for human mobility modeling using large-scale semantics-rich spatial-temporal datasets. 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 designed hidden Markov model to learn latent states and transitions between them in the embedding space. We also proposed 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 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, we plan to leverage deep learning to automatically extract the semantics in the trajectories to improve user grouping and location prediction.

References

[1] M. C. Gonzalez, C. A. Hidalgo, and A.-L. Barabasi, "Understanding individual human mobility patterns," *Nature*, vol. 453, no. 7196, pp. 779–782, 2008.

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

[3] M. Morzy, "Mining frequent trajectories of moving objects for location prediction," *Machine Learning and Data Mining in Pattern Recognition*, pp. 667–680, 2007.

[4] A. Noulas, S. Scellato, N. Lathia, and C. Mascolo, "Mining user mobility features for next place prediction in location-based services," in *Data mining (ICDM), 2012 IEEE 12th international conference on*. IEEE, 2012, pp. 1038–1043.

[5] H. Gao, J. Tang, and H. Liu, "Mobile location prediction in spatio-temporal context," in *Nokia mobile data challenge workshop*, vol. 41, no. 2, 2012, pp. 1–4.

[6] P. Baumann, W. Kleiminger, and S. Santini, "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, 2013, pp. 449–458.

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

[8] C. Zhang, K. Zhang, Q. Yuan, L. Zhang, T. Hanratty, and J. Han, "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, 2016, p. 1305.

[9] Q. Yuan, G. Cong, Z. Ma, A. Sun, and N. M. Thalmann, "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, 2013, pp. 605–613.

[10] F. Xu, T. Xia, H. Cao, Y. Li, F. Sun, and F. Meng, "Detecting popular temporal modes in population-scale unlabelled trajectory data," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 2, no. 1, p. 46, 2018.

[11] H. Cao, F. Xu, J. Sankaranarayanan, Y. Li, and H. Samet, "Habit2vec: Trajectory semantic embedding for living pattern recognition in population," *IEEE Transactions on Mobile Computing*, 2019.

[12] H. Cao, J. Feng, Y. Li, and V. Kostakos, "Uniqueness in the city: Urban morphology and location privacy," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 2, no. 2, p. 62, 2018.

[13] H. Shi and Y. Li, "Discovering periodic patterns for large scale mobile traffic data: Method and applications," *IEEE Transactions on Mobile Computing*, vol. 17, no. 10, pp. 2266–2278, Oct 2018.

[14] T. Mikolov, I. Sutskever, K. Chen, G. Corrado, and J. Dean, "Distributed representations of words and phrases and their compositionality," *Advances in Neural Information Processing Systems*, vol. 26, pp. 3111–3119, 2013.

[15] C. Zhang, K. Zhang, Q. Yuan, H. Peng, Y. Zheng, T. Hanratty, S. Wang, and J. Han, "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, 2017, pp. 361–370.

[16] A. Banerjee, I. S. Dhillon, J. Ghosh, and S. Sra, "Clustering on the unit hypersphere using von mises-fisher distributions," *Journal of Machine Learning Research*, vol. 6, no. Sep, pp. 1345–1382, 2005.

[17] S. Kullback and R. A. Leibler, "On information and sufficiency," *The annals of mathematical statistics*, vol. 22, no. 1, pp. 79–86, 1951.

[18] J. Han and C. Moraga, "The influence of the sigmoid function parameters on the speed of backpropagation learning," *From Natural to Artificial Neural Computation*, pp. 195–201, 1995.

[19] J. Reisinger, A. Waters, B. Silverthorn, and R. J. Mooney, "Spherical topic models," in *Proceedings of the 27th international conference on machine learning (ICML-10)*, 2010, pp. 903–910.

[20] K. Batmanghelich, A. Saeedi, K. Narasimhan, and S. Gershman, "Nonparametric spherical topic modeling with word embeddings," *arXiv preprint arXiv:1604.00126*, 2016.

[21] C. Zhang, L. Liu, D. Lei, Q. Yuan, H. Zhuang, T. Hanratty, and J. Han, "Triovecevent: Embedding-based online local event detection in geo-tagged tweet streams," in *ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2017, pp. 595–604.

[22] L. E. Baum, T. Petrie, G. Soules, and N. Weiss, "A maximization technique occurring in the statistical analysis of probabilistic functions of markov chains," *The annals of mathematical statistics*, vol. 41, no. 1, pp. 164–171, 1970.

[23] L. V. Der Maaten and G. E. Hinton, "Visualizing data using t-sne," *Journal of Machine Learning Research*, vol. 9, pp. 2579–2605, 2008.

[24] D. Brockmann, L. Hufnagel, and T. Geisel, "The scaling laws of human travel," *Nature*, vol. 439, no. 7075, pp. 462–465, 2006.

[25] B. Deb and P. Basu, "Discovering latent semantic structure in human mobility traces," in *European Conference on Wireless Sensor Networks*. Springer, 2015, pp. 84–103.

[26] W. Mathew, R. Raposo, and B. Martins, "Predicting future locations with hidden markov models," in *Proceedings of the 2012 ACM conference on ubiquitous computing*. ACM, 2012, pp. 911–918.

[27] J. Tang, M. Qu, M. Wang, M. Zhang, J. Yan, and Q. Mei, "Line: Large-scale information network embedding," in *Proceedings of the 24th International Conference on World Wide Web*. International World Wide Web Conferences Steering Committee, 2015, pp. 1067–1077.

[28] J. Cranshaw, E. Toch, J. Hong, A. Kittur, and N. Sadeh, "Bridging the gap between physical location and online social networks," in *Proceedings of the 12th ACM international conference on Ubiquitous computing*. ACM, 2010, pp. 119–128.

[29] M. Ester, H.-P. Kriegel, J. Sander, X. Xu *et al.*, "A density-based algorithm for discovering clusters in large spatial databases with noise." in *Kdd*, vol. 96, no. 34, 1996, pp. 226–231.

[30] D. Comaniciu and P. Meer, "Mean shift: A robust approach toward feature space analysis," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 24, no. 5, pp. 603–619, 2002.

[31] X. Lu, E. Wetter, N. Bharti, A. J. Tatem, and L. Bengtsson, "Approaching the limit of predictability in human mobility," *Scientific reports*, vol. 3, 2013.

[32] F. Xu, Z. Tu, Y. Li, P. Zhang, X. Fu, and D. Jin, "Trajectory recovery from ash: User privacy is not preserved in aggregated mobility data," in *Proceedings of the 26th International Conference on World Wide Web*. International World Wide Web Conferences Steering Committee, 2017, pp. 1241–1250.

[33] S. Jiang, Y. Yang, S. Gupta, D. Veneziano, S. Athavale, and M. C. Gonzlez, "The timegeo modeling framework for urban motility without travel surveys." *Proceedings of the National Academy of Sciences of the United States of America*, vol. 113, no. 37, p. E5370, 2016.

[34] J. Jeong, M. Leconte, and A. Proutiere, "Cluster-aided mobility predictions," in *Computer Communications, IEEE INFOCOM 2016-The 35th Annual IEEE International Conference on*. IEEE, 2016, pp. 1–9.

[35] J. McInerney, J. Zheng, A. Rogers, and N. R. Jennings, "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, 2013, pp. 469–478.

[36] Y. Wang, N. J. Yuan, D. Lian, L. Xu, X. Xie, E. Chen, and Y. Rui, "Regularity and conformity: Location prediction using heterogeneous mobility data," pp. 1275–1284, 2015.

[37] Q. Liu, S. Wu, L. Wang, and T. Tan, "Predicting the next location: a recurrent model with spatial and temporal contexts," in *Thirtieth AAAI Conference on Artificial Intelligence*, 2016, pp. 194–200.

[38] J. Feng, Y. Li, C. Zhang, F. Sun, F. Meng, A. Guo, and D. Jin, "Deepmove: Predicting human mobility with attentional recurrent networks," in *Proceedings of the 2018 World Wide Web Conference on World Wide Web*. International World Wide Web Conferences Steering Committee, 2018, pp. 1459–1468.

[39] C. Parent, S. Spaccapietra, C. Renso, G. Andrienko, N. Andrienko, V. Bogorny, M. L. Damiani, A. Gkoulalas-Divanis, J. Macedo, and N. Pelekis, "Semantic trajectories modeling and analysis," *Acm Computing Surveys*, vol. 45, no. 4, p. 42, 2013.

[40] H. Cao, Z. Chen, F. Xu, Y. Li, and V. Kostakos, "Revisitation in urban space vs. online: A comparison across pois, websites, and smartphone apps," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 2, no. 4, p. 156, 2018.

[41] Y. Bengio, A. Courville, and P. Vincent, "Representation learning: A review and new perspectives," *IEEE Transactions on Pattern Analysis Machine Intelligence*, vol. 35, no. 8, p. 1798, 2013.

[42] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient estimation of word representations in vector space," *Computer Science*, 2013.

[43] J. Pennington, R. Socher, and C. Manning, "Glove: Global vectors for word representation," in *Conference on Empirical Methods in Natural Language Processing*, 2014, pp. 1532–1543.

[44] Y. Di, Z. Chao, Z. Zhihua, H. Jianhui, and B. Jingping, "Trajectory clustering via deep representation learning," in *IJCNN*, 2017.

[45] J. Tang, M. Qu, and Q. Mei, "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, 2015, pp. 1165–1174.

[46] B. Yan, K. Janowicz, G. Mai, and S. Gao, "From itdl to place2vec–reasoning about place type similarity and relatedness by learning embeddings from augmented spatial contexts," *Proceedings of SIGSPATIAL*, vol. 17, pp. 7–10, 2017.

[47] M. Xie, H. Yin, H. Wang, F. Xu, W. Chen, and S. Wang, "Learning graph-based poi embedding for location-based recommendation," in *Proceedings of the 25th ACM International on Conference on Information and Knowledge Management*. ACM, 2016, pp. 15–24.

[48] S. Zhao, T. Zhao, I. King, and M. R. Lyu, "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, 2017, pp. 153–162.

[49] T. Qian, B. Liu, Q. V. H. Nguyen, and H. Yin, "Spatiotemporal representation learning for translation-based poi recommendation," *ACM Transactions on Information Systems (TOIS)*, vol. 37, no. 2, p. 18, 2019.

[50] H. Yin, W. Wang, H. Wang, L. Chen, and X. Zhou, "Spatial-aware hierarchical collaborative deep learning for poi recommendation," *IEEE Transactions on Knowledge and Data Engineering*, vol. 29, no. 11, pp. 2537–2551, 2017.

[51] H. Yin, L. Zou, Q. V. H. Nguyen, Z. Huang, and X. Zhou, "Joint event-partner recommendation in event-based social networks," in *2018 IEEE 34th International Conference on Data Engineering (ICDE)*. IEEE, 2018, pp. 929–940.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TKDE.2019.2937296, IEEE Transactions on Knowledge and Data Engineering

[52] H. Chen, H. Yin, W. Wang, H. Wang, Q. V. H. Nguyen, and X. Li, "Pme: projected metric embedding on heterogeneous networks for link prediction," in *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. ACM, 2018, pp. 1177–1186.

[53] S. Zhao, T. Zhao, H. Yang, M. R. Lyu, and I. King, "Stellar: spatial-temporal latent ranking for successive point-of-interest recommendation," in *Thirtieth AAAI conference on artificial intelligence*, 2016.

[54] H. Shi, H. Cao, X. Zhou, Y. Li, C. Zhang, V. Kostakos, F. Sun, and F. Meng, "Semantics-aware hidden markov model for human mobility," in *Proceedings of the 2019 SIAM International Conference on Data Mining*. SIAM, 2019, pp. 774–782.

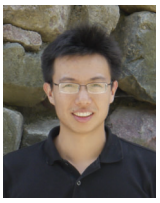
Hongzhi Shi received the B.S. degree in electronics and information engineering from Tsinghua University, Beijing, China, in 2017. He is currently a second-year Ph.D. student at the Department of Electronic Engineering of Tsinghua University, under the supervision of Prof. Yong Li. His research interests are in the areas of big data, including mobile traffic modeling, human mobility modeling and urban computing.

Yong Li (M'09-SM'16) received the B.S. degree in electronics and information engineering from Huazhong University of Science and Technology, Wuhan, China, in 2007 and the Ph.D. degree in electronic engineering from Tsinghua University, Beijing, China, in 2012. He is currently a Faculty Member of the Department of Electronic Engineering, Tsinghua University.

Dr. Li has served as General Chair, TPC Chair, TPC Member for several international workshops and conferences, and he is on the editorial board of three international journals. His papers have total citations more than 2300 (six papers exceed 100 citations, Google Scholar). Among them, eight are ESI Highly Cited Papers in Computer Science, and four receive conference Best Paper (run-up) Awards. He received IEEE 2016 ComSoc Asia-Pacific Outstanding Young Researchers and Young Talent Program of China Association for Science and Technology.

Hancheng Cao received his B.S. degree in Electronic Information Science and Technology from Tsinghua University, Beijing, China, in 2018, and he is currently pursuing Ph.D. degree in computer science department of Stanford University, CA. His research interests include data mining, ubiquitous computing and computational social science.

Xiangxin Zhou is currently an undergraduate majoring in electronics and information engineering in Tsinghua University, Beijing, China. His research interests are in the areas of big data, including human mobility modeling and urban computing.



Chao Zhang is an Assistant Professor at College of Computing, Georgia Institute of Technology. His research area is data mining and machine learning. He is particularly interested in developing label-efficient and robust learning techniques, with applications in text mining and spatiotemporal data mining. Chao has published more than 40 papers in top-tier conferences and journals, such as KDD, WWW, SIGIR, VLDB and TKDE. He is the recipient of the ECML/PKDD Best Student Paper Runner-up Award (2015) and the Chiang Chen Overseas Graduate Fellowship (2013). Before joining Georgia Tech, he obtained his Ph.D. degree in Computer Science from University of Illinois at Urbana-Champaign in 2018.

Vassilis Kostakos is a professor of human computer interaction in the School of Computing and Information Systems, University of Melbourne. He has held appointments with the University of Oulu, University of Madeira, and Carnegie Mellon University. He has been a fellow of the Academy of Finland Distinguished Professor Programme, and a Marie Curie Fellow. He conducts research on ubiquitous and pervasive computing, humancomputer interaction, and social and dynamic networks.