Downloaded 05/11/19 to 128.12.245.4. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

Semantics-Aware Hidden Markov Model for Human Mobility

Hongzhi Shi* Hancheng Cao † Xiangxin Zhou * Yong Li * Chao Zhang ‡
Vassilis Kostakos § Funing Sun ¶ Fanchao Meng ¶

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 semantics-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.

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]. 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, 8, 9, 10] 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 represented by POI type [11, 12]) 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 [5, 6, 13] 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 –

*Beijing National Research Center for Information Science and Technology (BNRist), Department of Electronic Engineering, Tsinghua University, Beijing, China. liyong07@tsinghua.edu.cn.

†Department of Computer Science, Stanford University. Work done at Tsinghua University.

‡School of Computational Science and Engineering, Georgia Institute of Technology.

§School of Computing and Information Systems, The University of Melbourne.

¶Tencent Inc.

Copyright © 2019 by SIAM
Unauthorized reproduction of this article is prohibited

Downloaded 05/11/19 to 128.12.245.4. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

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. 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 [6], 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 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 on each user group and obtain an ensemble of high-quality HMMs. 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. 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.

Copyright © 2019 by SIAM
Unauthorized reproduction of this article is prohibited

Downloaded 05/11/19 to 128.12.245.4. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

2 RELATED WORK

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, 14, 15]. 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. [14] discover that the theoretical maximum predictability of human mobility is as high as 88%. Various works [2, 3] focus on mobility modelling for next location prediction. Liu et al. [9] extend RNN into spatial-temporal recurrent neural networks (ST-RNN) for next place prediction by temporal and spatial information. Feng et al. [10] 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, temporal and user motivation in trajectory data as a whole to understand human mobility.

Semantics-aware mobility model: Recently, several semantics-aware mobility models have been proposed [16, 17] for spatial-temporal data. The most relevant works are those on modelling semantics-rich location data from geo-tagged social media (GeoSM) as twitter and foursquare. Ye et al. [5] propose a mobility model to predict user activity at next step. Yuan et al. [13] propose a who+when+where+what model to jointly model user spatial-temporal topics. Zhang et al. [6] 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 semantics-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 [18, 19]. Recently representation learning methods have been used for spatial-temporal data mining and knowledge discovery. Inspired by Zhang et al. [20] 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. [21] adapt skip-gram model [18] for learning the representations of place types. For applications in location-based POI recommendation, graph-based representation learning method [22] and word2vec-inspired model [23] have been presented. Zhang et al. [24] propose a embedding-based method for online local event detection. Different from previous works, in this paper we first introduce representation learning method in mobility modelling and propose a semantics-aware model, which contributes to our understanding of the interplay between spatial, temporal and semantic aspects in human mobility and achieves better prediction performance.

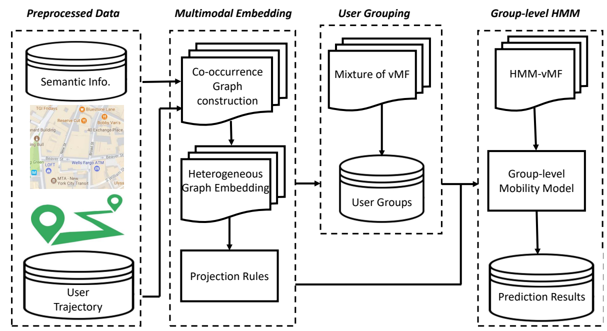


Figure 1: Overview of the proposed embedding based group-level human mobility model.

3 Model Overview

We introduce our proposed model overview in Fig. 1, which includes three major modules as follows.

Multimodal Embedding module builds the structure of user, temporal, spatial and user motivation/semantics 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[25, 20], we model each cluster of users as a von Mises-Fisher (vMF) distribution in the latent space. Naturally, we use mixture-of-vMFs model[26] to cluster users into groups in latent spaces for follow-up HMM training.

Copyright © 2019 by SIAM
Unauthorized reproduction of this article is prohibited

Downloaded 05/11/19 to 128.12.245.4. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

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.

4 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.

4.1 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, 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.

4.1.1 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.

4.1.2 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 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 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.

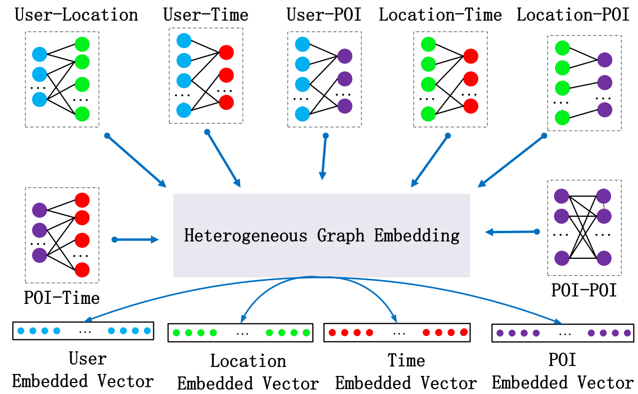


Figure 2: Illustration of the details of our representation learning model. The co-occurrence relationships construct 7 sub-graphs, which are jointly embedded with graph-based method.

The likelihood of generated node j given node k is defined as $p(j|k) = \frac{\exp(-v_j^T \cdot u_k)}{\sum_{i \in U} \exp(-v_i^T \cdot u_k)}$, where u_k and v_j represent embedding 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 $\hat{p}(j|k) = \frac{w_{kj}}{d_k}$, 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 subgraph G_{UV} as $L_{UV} = \sum_{j \in U} d_j d_{KL}(\hat{p}(\cdot|j)||p(\cdot|j)) + \sum_{k \in V} d_k d_{KL}(\hat{p}(\cdot|k)||p(\cdot|k))$, where $d_{KL}(\cdot)$ is Kullback-Leibler divergence [27]. With four different nodes representing user(U), temporal (T), spatial (S) and POI type (H) information, the overall loss function can be

Copyright © 2019 by SIAM
Unauthorized reproduction of this article is prohibited

obtained as

$$L = L_{UT} + L_{US} + L_{UH} + L_{TS} + L_{TH} + L_{SH} + L_{HH}. \tag{4.1}$$

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 [25]. 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

$$L' = -\log\sigma(u_j^T \cdot v_k) - \sum_{n=1}^{N} \log\sigma(-u_n^T \cdot v_k), \tag{4.2}$$

where $\sigma(\cdot)$ represents the sigmoid function [28].

4.2 Grouping-based HMM

4.2.1 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 [25, 20]. Also, there are some semantic models that use von Mises-Fisher (vMF) distribution in word embedding space [29, 30] and multimodal embedding space [24].

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

$$p(x|\mu,\kappa) = C_d(\kappa) exp(\kappa\mu^T x), \tag{4.3}$$

where the mean direction unit vector μ and the concentration parameter κ are two important parameters that describe vMF distribution. The normalization constant $C_d(\kappa) = \frac{\kappa^{d/2-1}}{(2\pi)^{d/2} I_{d/2-1}(\kappa)}$, 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. To estimate the parameters of vMF, we first calculate $r = \frac{\sum_{i=1}^{n} x_i}{n}$, then we can estimate the two parameters by $\hat{\mu} = \frac{r}{\|r\|}$ and $\hat{\kappa} = \frac{\|r\|d - \|r\|^3}{1-\|r\|^2}$.

In order to cluster users into several groups that have similar mobility semantic patterns, we use a mixture of vMF model to fit the embedding vectors of users.

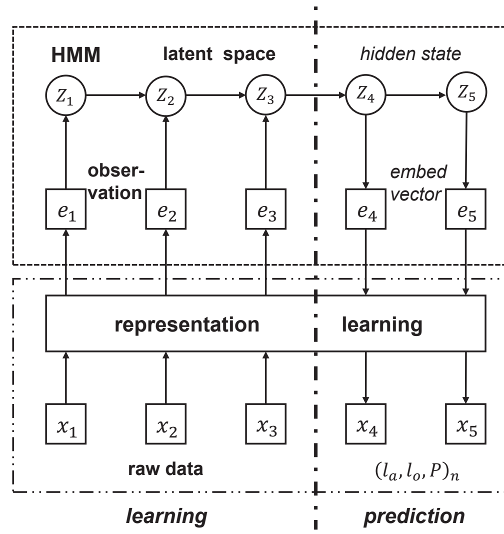


Figure 3: Illustration of the details of HMM-based prediction model in the latent space and its relationship with the physical locations.

The probability of v_U in a k-vMF distribution is given by,

$$p(v_U|\alpha,\mu,\kappa) = \sum_{h=1}^{k} \alpha_h f_h(v_U|\mu_h,\kappa_h), \tag{4.4}$$

where α_h is the weight of h-th mixture and sums to 1.

We design an EM frameworkto estimate $\alpha_h, \mu_h, \kappa_h$ for $h = 1, ..., k$ which maximize the probability of the whole k-vMF model. At last, we use the estimated parameters to obtain the probability each user belonging to a certain group given by,

$$p(h|v_{U_i},\mu,\kappa) = \frac{\alpha_h f_h(v_{U_i}|\mu_h,\kappa_h)}{\sum_{l=1}^{k} \alpha_l f_l(v_{U_i}|\mu_l,\kappa_l)}. \tag{4.5}$$

After obtaining the k user groups, we train one HMM for each of them.

4.2.2 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[25][20]. Thus, we utilize vMF distribution as the emission probability of each hidden state.

We adapt the Baum-Welch algorithm, an Expectation-Maximization (EM) procedure for HMM,

Downloaded 05/11/19 to 128.12.245.4. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

Copyright © 2019 by SIAM
Unauthorized reproduction of this article is prohibited

Downloaded 05/11/19 to 128.12.245.4. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

to estimate the parameters in the embedding space. The main difference between our proposed HMM and the tradition HMM is that we set the emission function of HMM as the vMF and set the observation as the embedding vectors representing locations instead of the locations' coordinates.

In order to leverage the model for next location prediction, we construct a set of length-2 sequences $(v_{l_n}, v_{l_{n+1}})$ for the locations in candidates set, where v_{l_n} is the embedding vector representing the current location l_n and $v_{l_{n+1}}$ is representing a location in the candidates. Then, we calculate the probability of generating such a sequence from the trained model as the score S of the sequences given by

$$S(l_{n+1}) = p(v_{l_n}, v_{l_{n+1}}|\Phi), \tag{4.6}$$

where Φ is the set of parameters of HMM. Thus, for all locations in the candidates, we define the scores as the probability of generating such a sequence from the our trained model and obtain a list of locations with top K scores.

5 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 three parts:

- Presenting case studies 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.

5.1 Experimental Settings

5.1.1 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.

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.

5.1.2 Baselines We compare our model with the following five solutions including the state-of-the-art methods.

Law [32] models the human mobility as a Lévy flight with long-tailed distributions.

GeoHMM [33] trains one HMM for all users' trajectory, where each hidden state generates locations by a Gaussian distribution.

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 [6] 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.

5.1.3 Evaluation Setup We partition each dataset into the training set and testing set. For the app collected 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, 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. 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 each candidate in the sets by (4.6). 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:

Copyright © 2019 by SIAM
Unauthorized reproduction of this article is prohibited

Downloaded 05/11/19 to 128.12.245.4. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

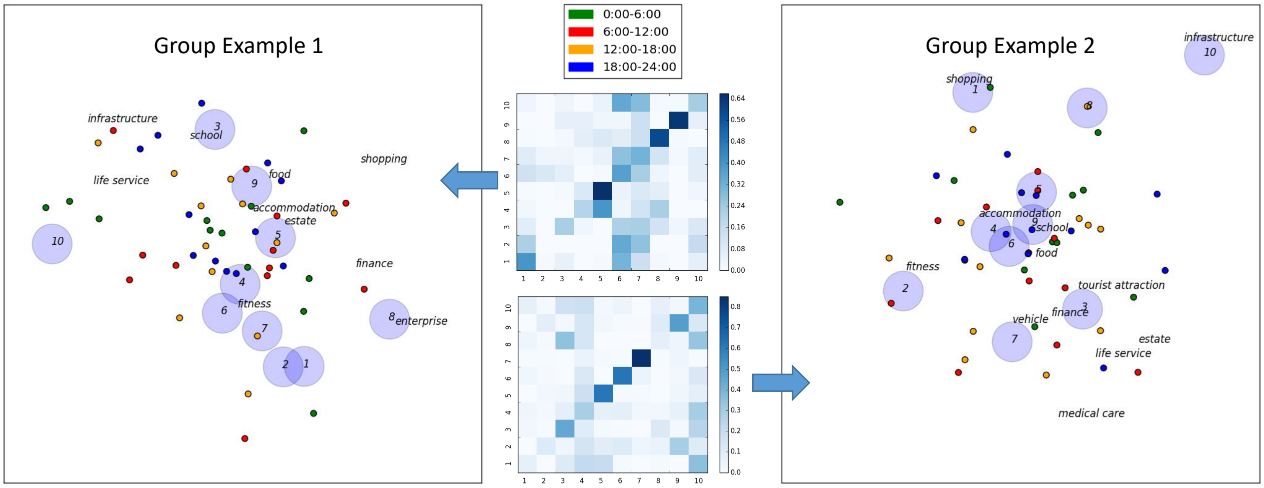


Figure 4: Two examples of the user groups. We map the embedding vectors of time, POI types and the central vector of each hidden state on a 2D plane with t-SNE[31]. The two heat maps in the middle represent HMM transition probability matrixes.

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.

We implemented our method (except the embedding part which is adapted by LINE[19] 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.

5.2 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[31]. 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.

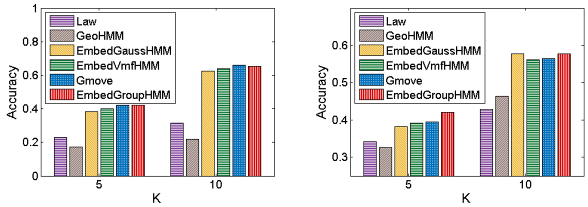
For the group example 1 in Fig.4, we can infer that this group probably represents sport-lovers. We observe that the hidden states 6 and 7 mean the activity of doing sports because they are near POI type ‘fitness’ (which contains gym, basketball court, natatorium, etc.) while the two temporal points 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.

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.

5.3 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. 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 comparing to the methods without considering semantics. 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* [6], we can observe

Copyright © 2019 by SIAM
Unauthorized reproduction of this article is prohibited

Downloaded 05/11/19 to 128.12.245.4. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php



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

Figure 5: Prediction Accuracy of Top K.

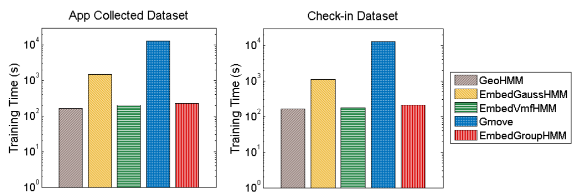
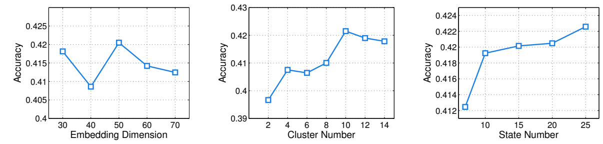


Figure 6: Training time on two datasets.

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 [19]. 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. Overall, compared to the state-of-the-art baseline methods, our proposed method either achieves much better results or costs much less time.

5.4 Parameter Effect To understand the roles of system parameters in our proposed mobility model, we vary these parameters to plot the performance curve of



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

Figure 7: Effects of parameters.

our model. There are three important parameters in our model, which respectively come from the three modules: (1) The embedding dimension D in *Multimodal Embedding* module. (2) The number of user groups G in *User Grouping* module. (3) 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.

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 $G = 10$. 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 $K \geq 10$. This is because many different mobility behaviors are not properly distinguished when represented by a few hidden states (K is too small).

6 Conclusion

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.

Acknowledgement

This work was supported in part by The National Key Research and Development Program of China under grant 2017YFE0112300, the National Nature Science Foundation of China under 61861136003, 61621091 and 61673237, Beijing National Research Center for Infor-

Copyright © 2019 by SIAM
Unauthorized reproduction of this article is prohibited

mation Science and Technology under 20031887521, and research fund of Tsinghua University - Tencent Joint Laboratory for Internet Innovation Technology.

References

[1] M. C. Gonzalez, C. A. Hidalgo, and A. L. Barabasi. Understanding individual human mobility patterns. *Nature*, 453(7196):779–782, 2008.

[2] A. Monreale, F. Pinelli, R. Trasarti, and F. Giannotti. Wherenext: a location predictor on trajectory pattern mining. In *KDD*, 2009.

[3] A. Noulas, S. Scellato, N. Lathia, and C. Mascolo. Mining user mobility features for next place prediction in location-based services. In *ICDM*, 2012.

[4] P. Baumann, W. Kleiminger, and S. Santini. The influence of temporal and spatial features on the performance of next-place prediction algorithms. In *Ubicomp*, 2013.

[5] J. Ye, Z. Zhu, and H. Cheng. What's your next move: User activity prediction in location-based social networks. In *SDM*, 2013.

[6] 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*, 2016.

[7] F. Xu, T. Xia, H. Cao, Y. Li, F. Sun, and Fanchao M. Detecting popular temporal modes in population-scale unlabelled trajectory data. *IMWUT*, 2018.

[8] 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. In *KDD*, 2015.

[9] Q. Liu, S. Wu, L. Wang, and T. Tan. Predicting the next location: a recurrent model with spatial and temporal contexts. In *AAAI*, 2016.

[10] J. Feng, Y. Li, C. Zhang, F. Sun, F. Meng, A. Guo, and D. Jin. Deepmove: Predicting human mobility with attentional recurrent networks. In *WWW*, 2018.

[11] H. Cao, J. Feng, Y. Li, and V. Kostakos. Uniqueness in the city: Urban morphology and location privacy. *IMWUT*, 2018.

[12] H. Shi and Y. Li. Discovering periodic patterns for large scale mobile traffic data: Method and applications. *IEEE Transactions on Mobile Computing*, 17(10):2266–2278, Oct 2018.

[13] 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 *KDD*, 2013.

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

[15] 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 *WWW*, 2017.

[16] 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*, 45(4):42, 2013.

[17] 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. *IMWUT*, 2018.

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

[19] J. Tang, M. Qu, M. Wang, M. Zhang, J. Yan, and Q. Mei. Line: Large-scale information network embedding. In *WWW*, 2015.

[20] 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 *WWW*, 2017.

[21] 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. *SIGSPATIAL*, 2017.

[22] M. Xie, H. Yin, H. Wang, F. Xu, W. Chen, and S. Wang. Learning graph-based poi embedding for location-based recommendation. In *CIKM*, 2016.

[23] S. Zhao, T. Zhao, I. King, and M. R. Lyu. Geo-teaser: Geo-temporal sequential embedding rank for point-of-interest recommendation. In *WWW*, 2017.

[24] 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 *KDD*, 2017.

[25] T. Mikolov, I. Sutskever, K. Chen, G. Corrado, and J. Dean. Distributed representations of words and phrases and their compositionality. *NIPS*, 2013.

[26] A. Banerjee, I. S. Dhillon, J. Ghosh, and S. Sra. Clustering on the unit hypersphere using von mises-fisher distributions. *JMLR*, 6(Sep):1345–1382, 2005.

[27] S. Kullback and R. A. Leibler. On information and sufficiency. *The annals of mathematical statistics*, 22(1):79–86, 1951.

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

[29] J. Reisinger, A. Waters, B. Silverthorn, and R. J. Mooney. Spherical topic models. In *ICML*, 2010.

[30] K. Batmanghelich, A. Saeedi, K. Narasimhan, and S. Gershman. Nonparametric spherical topic modeling with word embeddings. In *ACL*, 2016.

[31] L. V. Der Maaten and G. E. Hinton. Visualizing data using t-sne. *JMLR*, 9:2579–2605, 2008.

[32] D. Brockmann, L. Hufnagel, and T. Geisel. The scaling laws of human travel. *Nature*, 439(7075):462–465, 2006.

[33] W. Mathew, R. Raposo, and B. Martins. Predicting future locations with hidden markov models. In *Ubicomp*, 2012.

Downloaded 05/11/19 to 128.12.245.4. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

Copyright © 2019 by SIAM
Unauthorized reproduction of this article is prohibited